This is to certify that the
dissertation entitled

TRANSLATIONAL POWER DIFFERS BETWEEN BACTERIA PURSUING DIFFERENT ECOLOGICAL STRATEGIES

presented by

Les Dethlefsen

has been accepted towards fulfillment
of the requirements for the

Doctorate of Philosophy	in	Microbiology and Molecular Genetics & Ecology, Evolutionary Biology, and Behavior

Major Professor's Signature

4/27/04
Date

TRANSLATIONAL POWER DIFFERS BETWEEN BACTERIA PURSUING DIFFERENT ECOLOGICAL STRATEGIES

by

Les Dethlefsen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology and Molecular Genetics
Program in Ecology, Evolutionary Biology and Behavior

2004

ABSTRACT

TRANSLATIONAL POWER DIFFERS BETWEEN BACTERIA

PURSUING DIFFERENT ECOLOGICAL STRATEGIES

by

Les Dethlefsen

Translation, the polymerization of amino acids into protein, consumes more energy than any other process in the bacterial cell. The translational apparatus, including ribosomes, translation factors, tRNA molecules, tRNA synthetases, and other less abundant components, accounts for a substantial fraction of bacterial cell mass. Hence, selection for optimal performance of the translational apparatus is likely to be strong. Nonetheless, premature translational termination events, known as processivity errors, are not rare in laboratory-adapted *Escherichia coli,* the only organism for which data exist. Genetic, biochemical and physiological evidence, as well as models of translation, suggest the existence of an evolutionary tradeoff between translational power (the net rate of protein synthesis per mass invested in the translational apparatus) and translational yield (the net mass of protein synthesized per energy consumed). Microorganisms selected for fast maximal growth rates and a rapid response to resource abundance may favor high translational power to permit rapid protein synthesis and rapid growth, at the expense of more frequent processivity errors that waste energy by generating truncated polypeptides that are subsequently degraded. On the other hand, microorganisms adapted to exploit small fluctuations and limited availability of resources may favor high translational yield to minimize resource thresholds for growth and survival, at the expense of a reduced rate

of protein synthesis for a given investment in the translational apparatus. This hypothesis is tested using a collection of recent soil isolates containing bacteria of contrasting ecological strategies in each of several diverse phylogenetic groups, as well as one well-characterized representative of each ecological strategy. The specific growth rate, cell density and cell volume were measured for each of these 10 bacterial strains in batch culture in two media; a novel protocol was developed for measurement of the DNA, RNA and protein content of small samples of bacterial biomass. As predicted, translational power is higher in bacteria capable of a rapid growth response to abundant resources. Codon usage analysis can provide a comparison of the relative strength of selection for translational power between strains, so the relationship between codon bias and ecological strategy was examined in a large number of bacteria with fully sequenced genomes, using the number of copies of the ribosomal RNA operon per genome as an index of ecological strategy. The observed pattern of stronger translational selection in organisms with more copies of the rRNA operon is consistent with expectations based on macromolecular measurements. This pattern is better explained by a cost associated with translational power rather than the absence of a benefit of translational power among strains adapted to small resource fluctuations. Because codon bias directly affects translational power, we investigated whether variation in codon bias among organisms could explain the observed variation in translational power. However, the degree to which codon bias accelerates translation in *E. coli* is too small to explain the observed variation in translational power between strains. Instead, differences in translational power among microbes must be explained by differences in the performance of the translational apparatus itself.

This work is dedicated to my wife, Michelle, and my sons, Conrad and Arik.

ACKNOWLEDGMENTS

I am grateful for many sources of financial support that have allowed me to complete the work reported in this dissertation: a Michigan State University Distinguished Graduate Fellowship, the Center for Microbial Ecology at Michigan State University, a STAR Fellowship from the U.S. Environmental Protection Agency, a Center for Biological Modeling/Quantitative Biology Interdisciplinary Research Award, and the U.S. National Science Foundation in various guises. I am thankful as well for additional support provided by the Department of Microbiology and Molecular Genetics and the Graduate School at Michigan State University.

I give my sincere thanks to the many individuals who have enriched my education during graduate school, only a few of whom can be named here. Committee members, all of whom have made intellectual contributions to this work: Dr. John Breznak, Dr. Frank Dazzo, Dr. Michael Klug, Dr. Richard Lenski and Dr. Alan Tessier. Professors who deserve the status of honorary committee members: Dr. Julius Jackson and Dr. James Tiedje. All my colleagues, past and present, in the Schmidt and Breznak labs, particularly Dr. Bradley Stevenson and Dr. Joel Klappenbach. Many students and faculty colleagues in the Department of Microbiology and Molecular Genetics, the Center for Microbial Ecology and the Ecology, Evolutionary Biology and Behavior Program.

Finally, I express my deepest appreciation and sincere respect for my mentor, Dr. Thomas M. Schmidt.

TABLE OF CONTENTS

TABLE OF CONTENTS, continued

LIST OF TABLES

LIST OF FIGURES

KEY TO SYMBOLS AND ABBREVIATIONS

16S rRNA.....the small subunit ribosomal RNA gene of Bacteria and Archaea

χ^2.....the chi-square distribution of statistics

μ .....the specific growth rate of a culture in balanced exponential growth

A_{562}.....light absorbance at 562 nm, measure of color development in BCA assay

A-site.....the ribosomal site where an amino acyl-tRNA molecule enters the translational cycle

BCA .....bicinchoninic acid, key reagent in a colorimetric assay of protein content

BSA.....bovine serum albumin, a protein standard used for protein assays

CMEIAS .....Center for Microbial Ecology Image Analysis System, computer software to facilitate cell measurements from photomicrographs

CMR.....Comprehensive Microbial Resource, a website maintained by The Institute for Genomic Research

D.....the DNA content (by mass) of a cell or culture

DAPI .....4',6-diamidino-2-phenylindole, a DNA-binding fluorescent dye

EDTA.....ethylenediaminetetraacetic acid, a chelating agent

EF-G.....translational elongation factor G

EF-Tu .....translational elongation factor Tu

g.....the generation time of a culture in balanced exponential growth

g.....the acceleration due to gravity at the Earth's surface, used as a unit of centrifugal acceleration

g_p.....the proteome generation time, an estimate of the time required for replication of the proteome with different hypothesized degrees of codon bias

HE .....highly expressed genes

MM .....the macromolecular content (by mass) of a cell or culture, the sum of the DNA, RNA and protein content

N.....any one of the 4 possible nucleotides (in a nucleotide sequence)

Nc .....the effective number of codons, an index of codon bias comparing observed codon frequencies to uniform synonymous codon use

N'c .....a modification of Nc, an index of codon bias comparing observed to expected codon frequencies

ΔN'c .....the transformation 61–N'c, an index of the strength of translational selection when observed codon frequencies are based on highly expressed genes and expected codon frequencies are genome averages

OD_{420}.....optical density at 420 nm, a measure of culture turbidity used for determinations of growth rate

PMT .....photomultiplier tube voltage, a sensitivity adjustment for a fluorescence scanner

P .....the protein content (by mass) of a cell or culture

P-site .....the ribosomal site where peptidyl-tRNA is bound during the majority of the translational cycle, following translocation and prior to the subsequent peptidyl transfer event

R.....the RNA content (by mass) of a cell or culture

R.....a purine nucleotide, either adenine (A) or guanine (G)

rrn .....the ribosomal RNA operon

s_{bias}.....the translation rate benefit of codon bias, estimated as the fractional decrease in the time required to replicate the proteome given a certain level of codon bias, in comparison to the time that would be required with little or no codon bias

SDS .....sodium dodecyl sulfate, a detergent

TCA.....trichloroacetic acid, a reagent that precipitates proteins

TE.....Tris-EDTA, a buffer used during nucleic acid measurements

X.....an unspecified nucleotide (in a nucleotide sequence)

Y.....a pyrimidine nucleotide, either cytosine (C) or, in RNA, uracil (U)

Chapter 1. Introduction: Unexplained variability in translational performance between microbes

The scientific literature contains a number of explicit comparisons of the concentration of ribosomes or of RNA between a slowly growing microbe and *Escherichia coli* [1-4]. These reports note that the concentrations found in their respective strains are not unusual in absolute terms. The estimates of the number of ribosomes per volume or the mass of RNA per volume in the slowly growing strains are similar to values determined for *E. coli* [5]. However, each report emphasizes the unexpectedness of finding similar values in the slowly growing organism and in *E. coli* growing at a considerably faster rate. Since ribosome and RNA concentrations vary with growth rate in *E. coli*, the slow-growing organisms have more ribosomes or RNA than expected, if the relationship between ribosome content and growth rate established for *E. coli* [5] is extended to other species.

Stated more precisely, these observations violate the expectation that the amount of protein synthesized per ribosome per time interval is approximately constant across strains. If the expectation were correct, fewer ribosomes would be needed per mass of protein in a strain with a longer generation time. This expectation can be decomposed into two assumptions: that the translation rate is similar between strains, and that the fraction of active ribosomes is similar between strains. Implicit in these assumptions is the idea that the quantity of ribosomes is matched to the demand for protein synthesis, since the presence of superfluous ribosomes at a given level of protein synthesis must lead to either a decrease in the average rate of translation, or a decrease in the fraction of ribosomes that are active, or both. All 4 studies that compared the ribosome or RNA

content of a slowly growing microbe to *E. coli* went on to discuss the validity of these assumptions in light of their unexpected results; 3 of the reports conclude that slower translation in the slowly growing bacterial strain is likely [2-4].

Indeed, despite earlier arguments to the contrary [6,7], it is now generally accepted that the translation rate varies with growth rate even within *E. coli,* ranging from 12 to 21 amino acids polymerized per ribosome per second as the specific growth rate varies from 0.4 to1.7 hr^{-1}, assuming that the fraction of active ribosomes is constant [5]. Two explanations have been offered for this variation. Koch has suggested that during slow growth, the ribosome content of *E. coli* provides a capacity for protein synthesis that is, in fact, greater than the demand. He argues that maintaining 'surplus ribosomes' is beneficial during slow growth in sub-optimal conditions in a variable environment, because they permit a more rapid return to faster growth rates when conditions improve [8,9]. Consistent with this prediction, inactive ribosome dimers have been discovered in *E. coli,* and the fraction of ribosomes forming such dimers increases as growth rate declines in *E. coli* [10-12]. Such dimers could have been mistaken for translationally-active polysomes in earlier work that appeared to establish a constant fraction of active ribosomes across growth rates [13]. Failure to account for this phenomenon (as in reference 5) would generate an apparent decline in the translation rate per active ribosome at slower growth rates, by overestimating the number of active ribosomes.

In contrast, by considering the cost of supplying the ribosomes with substrates as well as the cost of the ribosomes in a growing cell, Ehrenberg and Kurland have argued that

genuinely slower translation is to be expected at slower growth rates, if organisms have been selected to maximize their growth rate in a range of growth conditions [14]. They find that the optimal allocation of resources involves an asymptotic approach to saturation of the ribosome with ternary complexes (i.e. the maximum translation rate) as the organism approaches its maximum growth rate [14]. A number of predictions of the Ehrenberg-Kurland model are supported by empirical data [15]. Although this explanation is quite distinct from the explanation offered by Koch, the two are not mutually exclusive. Neither explanation, however, suggests a benefit for inactive ribosomes or slow translation in organisms that are adapted to stable environments, or that are growing at their maximum rate. Under these conditions, the selective pressures identified by Koch and by Ehrenberg and Kurland would favor rapid translation and a high fraction of active ribosomes. In this context, in organisms growing at or near their own maximal rate but more slowly than *E. coli,* the existence of *E. coli*-like concentrations of ribosomes or RNA [1-4] remains unexplained.

To extend these findings, we have searched the scientific literature for data on the protein and RNA content of microbes during balanced, exponential growth at known rates. To provide a consistent and rigorous framework for comparison, we will compare the total rate of protein synthesis of a cell or of a culture, normalized to the biomass invested in the protein synthesis machinery [16]. This type of comparison, made explicitly or implicitly, is used in each of the studies discussed above that investigate translation in slowly-growing microbes [1-4]. We introduce a new term, 'translational power', to refer to the translational output per mass of the translational apparatus. However, before

presenting the results of our comparison, we must address a semantic issue: Why introduce a new term for a concept and a quantity that have already been named?

The concept and the semantics of translational power

We use the new term 'translational power' to describe the translational output per mass of the translational apparatus, precisely the same concept (and the same quantitative parameter, see below) that was originally defined as 'ribosome efficiency' [5-7,16]. In recent years, this concept has more commonly been called 'translational efficiency' [14,17], particularly in the context of explaining codon usage bias [15,18,19]. Although we are reluctant to depart from established terminology, we do so to avoid an inconsistency with the meaning of 'efficiency' as it is used in many other areas of science and in colloquial usage. In the physical sciences and in many areas of biology, efficiency refers to a comparison between the output and the input of a process, particularly the fluxes of energy and/or mass. For example, the efficiency of a machine is the work performed by the machine divided by the energy supplied to it [20]. Less than perfect efficiency occurs if not all the energy required to operate the machine is made available for useful work, e.g., energy is lost as heat. Similarly, ecologists define 'trophic transfer efficiency' as the fraction of energy contained in the biomass at one trophic level which is transferred to the biomass of the next trophic level [21]. These scientific meanings of 'efficiency' are consistent with the common notion that a process operating with little waste is highly efficient.

According to these conventions, calculations of efficiency make no direct reference to the *rate* at which a process occurs. Instead, physicists and engineers use the term 'power' to refer to the rate of energy consumption or the rate at which work is performed [22]. The semantic distinction between power (or rate) and efficiency is important, because many real and idealized physical systems can approach maximal efficiency only by occurring at infinitesimally slow rates; i.e. there is a tradeoff between power and efficiency. Several attempts to demonstrate that a power-efficiency tradeoff must be an important universal constraint for biological systems [23-25] have justifiably been criticized for the inappropriate use of thermodynamic concepts and other flaws [26-28]. Despite the failure of these universal arguments, numerous specific examples of tradeoffs that can be described in terms of power (ultimately affecting growth rates) and efficiency (of resource utilization) have been described [29]. The organisms involved include rodents [30-32], frogs [33], snails [34], insects [35], freshwater crustaceans [36-38], trees [39], plants [40-42], and phytoplankton [43], based on mechanisms such as foraging behavior, metabolic rates, resource allocation and resource affinity, as well as other unknown aspects of cell physiology or biochemistry. A number of comparisons of coexisting bacterial species have also provided evidence for power-efficiency tradeoffs [44-47], as have comparisons of engineered mutant strains [48]. However, the absence of apparent tradeoffs in some carefully designed studies of bacteria demonstrates that such tradeoffs are not inevitable [49-51]. Even if power-efficiency tradeoffs occur only in some contexts, there is value in maintaining a semantic distinction derived from both scientific and colloquial usage between power (implying a rapid rate) and efficiency (implying low waste).

However, the term 'translational efficiency' blurs this distinction because it refers to a rate, expressed in units of inverse time. It is a measure of the power of the protein synthesis subsystem to drive the self-replication of the protein-dominated autocatalytic system to which it belongs. (It is exactly analogous to the power:mass ratio recognized as a critical performance characteristic for engines that must move their own mass as one component of a vehicle.) Furthermore, we will suggest in this dissertation that obtaining a high translational output per mass invested in the translational apparatus results in a high frequency of wasteful errors, meaning that a high mass-normalized translational output trades off with the energetic efficiency of translation. For these reasons, we prefer the term 'translational power' as a description of the translational output per mass of the translational apparatus that is more consistent with the expected meanings of 'power' and 'efficiency' based on other areas of science and on colloquial usage.

However, there would be obvious drawbacks in attempting to redefine the established term 'translational efficiency' as our convenient term for referring to the energetic efficiency of protein synthesis. Hence, in keeping with the meaning of the term 'yield' in microbiology, we define 'translational yield' as the mass of protein produced per energy consumed by the translational apparatus.

Translational power measures both the translation rate and the active fraction of ribosomes

Translational power measures both the average translation rate and the fraction of active ribosomes, which we demonstrate as follows, using the approach of chapter 6 of

reference 52. The average translation rate of a cell or culture (also called the peptide chain growth rate [5,52,53]) is the rate of amino acid polymerization per active ribosome:

$$\text{translation rate} = \frac{\text{number of amino acids polymerized per unit time}}{\text{number of active ribosomes}}.$$

For a culture in balanced, exponential growth, the rate of increase of any culture component X is $dX/dt = \mu X$, where μ is the specific growth rate. Hence, the rate of protein synthesis (i.e. the translational output) in a culture growing at rate μ, containing a mass P of protein, is simply μP. Translational output expressed in units of mass can be converted to a rate expressed in numbers of amino acids by dividing by the average mass of an amino acid.

$$\text{number of amino acids polymerized per unit time} = \mu P/(\text{average mass of amino acid})$$

The number of ribosomes in a culture containing a mass R of RNA can be found by multiplying R by the fraction of RNA that is ribosomal, and then dividing by the mass of RNA in a ribosome. However, only a fraction of these ribosomes are active at any given time. Thus,

$$\#\text{ of active ribosomes} = \frac{R \times (\text{ribosomal fraction of RNA})}{(\text{mass of RNA per ribosome})} \times (\text{active fraction of ribosomes})$$

Substituting the two latter equations into the first equation yields:

$$\text{translation rate} = \frac{\mu P/(\text{average mass of amino acid})}{\frac{R \times (\text{ribosomal fraction of RNA})}{(\text{mass of RNA per ribosome})} \times (\text{active fraction of ribosomes})}.$$

After rearranging terms, we have:

$$(\text{translation rate}) \times (\text{active fraction of ribosomes}) = \frac{\mu P}{R} \times C,$$

where

$$C = \frac{(\text{mass of RNA per ribosome})/(\text{average mass of amino acid})}{(\text{ribosomal fraction of RNA})}.$$

The quantity $\mu P/R$ represents translational power [5,16], the translational output per biomass (measured as RNA) invested in the translational apparatus. It is clear that translational power reflects both the translation rate and the fraction of active ribosomes.

What of the term we have labeled *C*, implying a constant? The two quantities in the numerator, the mass of RNA in a ribosome and the average mass of an amino acid, are indeed constant or nearly constant, both within a strain at different growth rates, and across strains. However, despite the constant ribosomal fraction of RNA reported in reference 5, other data indicates that the rRNA fraction decreases from about 85% to

about 75% as growth rate declines in *E. coli* from 1.7 hr^{-1} to 0.28 hr^{-1} [54,55], a result which is expected on theoretical grounds [14,56]. This variation is not dramatic; it would reduce translational power by only 12%, if the translation rate and active fraction of ribosomes were unchanged. Data are also available from 2 of the 4 comparative studies discussed earlier. The rRNA fraction is reported as 84% for *Halobacterium cutirubrum* at specific growth rates of both 0.10 hr^{-1} and 0.05 hr^{-1}, after the authors made the deliberately generous assumption that messenger RNA comprises 5% of the total RNA [4]. The rRNA fraction is about 85% for *Rickettsia prowazekii* at a specific growth rate of ~0.07 hr^{-1}, after a correction is made for 2-3% messenger RNA [2]. These data do not suggest that variation in the ribosomal fraction of RNA will obscure the relationship between translation rate, the active fraction of ribosomes, and translational power, even for comparisons between strains that grow at very different rates.

Comparisons of translational power

Table 1.1 summarizes comparisons of translational power between *E. coli* and all other microbial species for which data could be found in the literature, including the comparative studies discussed earlier. *E. coli* is represented by the Bremer and Dennis data [5], which are typical of the data reported for *E. coli* in many other studies. Similarly, comparisons between *E. coli* and 2 closely related species of enteric bacteria, *Salmonella enterica* and *Enterobacter aerogenes,* are made using only a single representative study for the latter strains, chosen from among several published reports. For the remaining species, only a single published study was available for comparison, except for one species represented by two studies, both of which were included. For strains not grown

Table 1.1: Comparisons of translational power

Comparison organism	specific growth rate[a] (hr^{-1})	actual growth temperature[b] (correction)[c]	compared by[d]	*E. coli* comparison growth rate[e] (hr^{-1})	translational power[f]	Reference
Sphingopyxis alaskensis	0.29	30°C (1.80)	RC[g]	>1.73[h]	<17%[h]	3
sulfate reducing strain PT2[i]	0.40	23°C (3.34)	RNA	>1.73[h]	<17%[h]	57
Streptomyces coelicolor	0.54	30°C (1.80)	RNA	>1.73[h]	<21%[h]	1
Halobacterium cutirubrum	0.10	37°C	rRNA	0.49	22%	4
Rickettsia prowazekii	0.09	34°C (1.28)	RC[g]	0.37	24%	2
Synechococcus sp. 6301	0.16	39°C (0.85)	RNA	0.36	42%	58
Streptomyces hygroscopicus	0.58 – 0.90[j]	28°C (2.14)	RNA	0.82 – >1.73[h,j]	<42%[h] – 110%[j]	59
Megasphaera elsdenii	0.20	39°C (0.85)	RNA	0.46	44%	60
Bacillus cereus	0.61	34°C (1.28)	RNA	1.22	51%	61
Selenomonas ruminantium	0.30 – 0.43[j]	39°C (0.85)	RNA	0.35 – 0.88[j]	50% – 78%[j]	62,63
Salmonella enterica	1.66	37°C	rRNA	1.63	102%	16
Enterobacter aerogenes	0.94	35°C (1.18)	RNA	0.80	123% – 154%[j]	64
Lactococcus lactis	1.94	30°C (1.80)	RNA	0.51	391%	65

a-j. Footnotes on facing page.

Table 1.1 footnotes

a. The highest growth rates (temperature corrected for 37°C, see text) for which macromolecular data were available for the comparison organism; shown as the specific growth rate = ln(2)/(*generation time*).

b. Actual growth temperature of the comparison organism.

c. Factor by which actual growth rate has been multiplied to estimate the temperature corrected growth rate expected at 37°C, which is shown in column 2. Temperature correction factors based on data from reference 66.

d. Similar investment in the translational apparatus between the comparison organism and *E. coli* assessed as follows: RC, similar ribosome concentration; RNA, similar protein:RNA ratio; rRNA, similar protein:rRNA ratio. For all comparisons, *E. coli* data were taken from reference 5.

e. Growth rate at which *E. coli* matches comparison organism for investment in translational apparatus, assessed by the criteria shown in column 4.

f. Translational power of comparison organism, expressed as a percentage of the translational power of *E. coli.*

g. For consistency, the comparison of this organism to *E. coli* using ribosome concentration (RC) made in the original reference is not reported. Instead, ribosome concentration as a function of growth rate for *E. coli* was recalculated from the data of reference 5, assuming a cytoplasmic volume of 1.1 fl at a growth rate of 1.03 hr^{-1} and a constant density of dry cell mass across growth rates. The comparison of translational power for this organism assumes a similar protein concentration to *E. coli.*

h. Comparison organism makes a larger investment in translational apparatus than *E. coli* growing at the fastest rate for which data are available. Comparison is made conservatively to the highest *E. coli* growth rate.

i. Related to *Desulfovibrio vulgaris* by 16S rRNA gene sequence analysis.

j. Range of values shown corresponds to the maximum growth rates obtained for this organism in different culture conditions.

at 37°C, we assume that the growth rate, but not the macromolecular content, would be altered by growth in the same medium at a different temperature [67]. The growth rates reported for these strains were adjusted to the growth rates expected at 37°C using the linear range of the relationship reported by Farewell and Neidhardt [66]. (Although this relationship was generated with *E. coli,* the comparison is mathematically identical whether the temperature correction is applied to *E. coli* or the comparison strain.)

The comparisons in Table 1.1 are made at the fastest growth rate for which data are available for each of the organisms compared to *E. coli*, to reduce the influence of ribosomes that are either inactive or translating at less than their maximal rates. However, the *E. coli* data used in each comparison are taken from a growth rate such that the investment in the translational apparatus (assessed as ribosome concentration or the protein:RNA ratio) is similar in the two organisms (see discussion below). A comparison at similar investment levels reflects the expectation that the selective pressure to maximize translational output varies with the biomass of the apparatus [14,68]. If the comparisons had always been made to the fastest *E. coli* growth rate (i.e. where its translational power is highest), the disparity in translational power would be greater for most of the comparisons shown. It would appear that the expectation of a constant translational power is not supported by the published data, particularly for comparisons between microbes adapted to different ranges of growth rates. While translational power is higher in *E. coli* and other fast-growing organisms, it is lower in slow-growing organisms, ranging from less than 17% to 42% of the value for *E. coli.* Slowly growing microbes seem to translate more slowly, or have a larger fraction of inactive ribosomes, than microbes capable of rapid growth.

Selection on translational power

Why should some microbes have low translational power? One answer could be that translational power doesn't contribute to fitness for these microbes; indeed, Sharp and his colleagues have suggested that this is the case for several microbes with unusual patterns of codon use [69,70] (see also Chapter 3). However, one of the implications of the Ehrenberg-Kurland model of the growing cell is that all components of the cell can be considered as growth-limiting, in proportion to their mass [14]. Although Ehrenberg and Kurland obtained this result rather more rigorously via differential calculus, their logic can be revealed by an evolutionary thought experiment: If, in some specific environmental context, the metabolic flux through a particular biochemical pathway of a bacterium is the limiting factor for growth, there is strong selective pressure to increase the flux through that pathway. Even if the activity of the cell components responsible for that pathway cannot be improved, the flux through the pathway could be increased by increasing the abundance of those cell components. Since, by hypothesis, other metabolic pathways of the cell are *not* the limiting factors for growth, biomass could be diverted from those functions without reducing the growth rate. Obviously, such a process would lead eventually to the growth-rate maximized cell, with biomass allocated to various cell components so that the flux though all branches of metabolism is balanced, neither limiting to growth nor in excess of demand.

If the cell components responsible for the limiting pathway were initially a very small fraction of cell mass, only a slight reallocation of total biomass could increase the flux through that pathway dramatically. On the other hand, if the limiting pathway already

required an investment of the dominant fraction of cell mass, diverting all the 'excess' biomass from all other pathways to the dominant pathway would provide a relatively small fractional increase in the limiting flux. Similarly, interference with the components of a minor pathway in the growth rate-maximized cell (e.g., a mutation) can be accommodated with only a slight growth rate penalty, whereas interference of the same relative severity in the components of a dominant pathway has a larger effect on the growth rate. It is in this sense that all pathways in the growth rate-maximized cell contribute to growth rate limitation in proportion to their mass.

This analysis provides insight into a phrase used above, selection to improve the 'activity of the cell components'. In the context of maximizing growth, improved activity must be interpreted as mass-normalized activity. A cell is poorly served by obtaining a new enzyme with twice the activity of the old enzyme but triple its mass; the cell would do better to increase the expression of the old enzyme. Note that the concept of mass-normalized activity, when applied to the translational apparatus, is precisely translational power. The Ehrenberg-Kurland model of the cell identifies selection for rapid growth with selection for high translational power, a particular example of the selection for high mass-normalized activity, i.e. high power, in all cell components.

Clearly, this model is an idealization unlikely to be matched by actual bacteria, for several reasons, including spatial and temporal heterogeneity of the environment: the growth rate-maximized cell is a moving target as environmental conditions change. Cells will be selected to track the optimal phenotype for their current environment via

regulatory changes, but neither the evolutionary nor the physiological adaptation will be perfect or instantaneous. Even so, the most massive components of the cell will be under the strongest selective pressure, and the most central components of the cell, those essential under all environmental conditions, will be under the most continuous selective pressure. According to both these criteria, the allocation of biomass to the protein synthesis system is likely to be more constrained by natural selection than allocation towards other subsystems of the cell. The explanation that translational power may be low in a particular microbe because protein synthesis makes little contribution to its fitness is difficult to sustain if a substantial fraction of the cell's biomass is invested in the translational apparatus. At least for the slowly growing strains represented in Table 1.1, measurements of ribosome or RNA content suggest that this is the case.

Tradeoffs affecting translational power

A second potential explanation for low translational power could be the existence of evolutionary tradeoffs. If high translational power is incompatible with some other trait that also contributes to fitness, at least under some environmental conditions, the degree to which organisms display high translational power may depend on the conditions to which they are adapted. Koch's explanation of 'surplus ribosomes' in *E. coli* [8] invokes a tradeoff between high translational power during slow growth and the ability to rapidly shift up to a faster growth rate. There is evidence for a tradeoff involving not just the quantity, but the performance of ribosomes as well. Mikkola and Kurland investigated 65 natural isolates of *E. coli* from the ECOR collection [71], and found a wide range of growth rates during batch culture in either rich or minimal medium [72,73]. The fastest

growth rates among the natural isolates were comparable to the growth rates of several laboratory-adapted strains of *E. coli* in the same media; the slowest growth rate was nearly 3 times slower. The translation rates *in vitro* of ribosomes removed from these strains also varied widely, with the fastest translation rates comparable to the translation rate of ribosomes extracted from laboratory-adapted strains. The only variable in these comparisons of translation rate was the source of the ribosomes; all other components of the *in vitro* translation system were held constant. There was a linear relationship between the growth rate of a strain and the translation rate *in vitro* of the ribosomes taken from that strain. Seven of the natural isolates that spanned the observed range of growth rates and translation rates were used to inoculate glucose-limited chemostats. Remarkably, after evolving in this environment for only about 280 generations, the descendents of all 7 strains converged on the values typical of laboratory-adapted strains for both growth rates in batch culture and translation rates *in vitro* [73,74].

At the very least, it can be concluded that the selective pressures for rapid growth experienced in typical laboratory environments are not universally experienced by *E. coli* in its native habitat. However, one additional piece of evidence argues for the existence of an evolutionary tradeoff, with alternative phenotypes apparently favored in different natural settings, but only one of the phenotypes favored in the lab. The natural isolates characterized by slower growth and slower translation were better able to survive carbon starvation than either laboratory strains or fast-growing natural isolates, and selection for rapid growth and rapid translation led to the loss of this trait. Unfortunately, very little information is available on the genetic changes that occurred during adaptation of the

slowly growing natural *E. coli* strains to laboratory conditions. Both the rapidity of the evolutionary change and the trajectory of phenotypic changes over the 280 generations suggest that only 1-3 mutations are involved [73,74]. The linear relationship between the growth rates of the strains and the translation rates of isolated ribosomes under constant *in vitro* conditions, both before and after selection, strongly suggests that the ribosome itself is involved in the adaptation.

Ecological strategies of microbes

Alternative responses to an evolutionary tradeoff that are favorable under different circumstances, or that ensure fitness by different mechanisms in the same environment, can be described as ecological strategies. Low translational power with enhanced starvation survival, or high translational power with diminished starvation survival, are ecological strategies available to *E. coli*. We hypothesize that similar strategic choices may explain differences in translational power observed over much greater evolutionary distances, as represented by the comparisons between bacterial species shown in Table 1.1.

Our treatment of bacterial ecological strategies is informed by the tradition of life history analysis in ecology, particularly the categorization of traits as *r*-selected or K-selected (indicating that the trait enhances intrinsic growth rate or carrying capacity, respectively) [75-77]. Another source of insight has been the long-standing interest in the implications of resource availability for microbial physiology, summarized by some as contrasts between traits found or expected in oligotrophic and copiotrophic

microbes [78,79], although such descriptions are not without controversy [80,81]. While these pairs of contrasting terms are related, in that copiotrophic microbes may possess many r-selected traits and oligotrophic microbes many K-selected traits, these two pairs of contrasts are certainly not identical [80]. However, both of these distinctions capture some important aspects of the contrasting ecological strategies we believe may be responsible for differences of translational power between strains. The rapid growth and high translational power shown by laboratory adapted *E. coli* are r-selected traits beneficial during resource abundance. Starvation survival, which we interpret as reflecting better use of limited endogenous resources, would be a K-selected trait and advantageous in oligotrophic environments. However, the thrust of this work is not to find correlations between traits because they are selected under the same ecological conditions, but to explore the possibility that a tradeoff exists that could explain low translational power in many diverse strains. Without a compelling reason to assert that either resource availability or the degree of crowding is the primary determining factor for the ecological strategies we seek to investigate, we will avoid describing the hypothetical strategies in these more established, but more general terms.

A more recent and more specific description of alternative ecological strategies involves differences between microbes in the number of copies of the ribosomal RNA (*rrn*) operon per genome. A high *rrn* copy number, because it allows rapid ribosome synthesis, contributes both to a fast maximal growth rate and to a capacity for rapid acceleration of the growth rate when growth conditions improve [82]. Hence, it is believed to be an adaptation to niches characterized by episodes of resource abundance. A benefit

of low *rrn* copy number has not been demonstrated unequivocally, but there may be a benefit related to a reduced burden of unnecessary expression when even basal levels of *rrn* expression exceed the need for ribosome synthesis during slow growth [83]. It has been demonstrated that a particular *E. coli* strain with a near-complete deletion of one of the 7 *rrn* operons experiences a fitness penalty during batch culture that is not apparent during chemostat growth [84]. Hence, low *rrn* number may be an adaptation to stable niches characterized by slow growth and low nutrient availability [3,83]. In any case, it would be expected that microbes with lower *rrn* copy number would be less able to respond rapidly to a sudden increase in resource availability than microbes with higher *rrn* copy number. Klappenbach and his colleagues confirmed this expectation by comparing the *rrn* copy number of soil bacterial isolates that first formed visible colonies either within 1-2 days, or after more than 7 days [85].

We selected strains with known times of initial colony appearance and known numbers of *rrn* operons per genome from among the bacteria examined by Klappenbach [85,86] for the research on translational power reported in this dissertation. To refer succinctly to the known traits of the organisms, we use the term 'rapid responders' for bacteria that formed colonies early and have high *rrn* copy number, and the term 'slow responders' for bacteria with the opposite traits. These terms apply to the response of the strains to resource abundance in terms of population growth only; they do not imply different abilities to respond to the environment in terms of sensory signal transduction or resource transport. 'Rapid response' is a useful description not only of what we know about the characteristics of a subset of these strains, but also of the strategy we believe these strains

represent. It implies both a capacity for rapid acceleration of growth rate and a high maximal growth rate. Clearly, a rapid growth response can confer a fitness advantage in some conditions.

'Slow response', however, while equally descriptive of another subset of strains, is less apt as a description of an ecological strategy. We do not imagine that a dilatory growth response is itself contributing to the fitness of these organisms; rather, it is the unavoidable consequence of some other, unknown advantageous traits. On the strength of the *E. coli* experiments involving starvation survival [74] and the proposed benefit of low *rrn* copy number [83,85,86], we surmise that at least one such trait involves maximizing yield, the conversion of resources into biomass. Additional support for the existence such a tradeoff will be found in Chapter 3, where we propose mechanisms linking high translational power (a component of the rapid response strategy) to frequent errors in translation that would reduce translational yield. However, we have at present no direct evidence that our slowly responding strains are, in fact, better able to convert resources into biomass or protein than the rapidly responding strains. Hence, we will continue to use the term 'slow responders' to refer to a particular subset of the organisms in our collection. The contrasting ecological strategies discussed in this chapter are depicted in Figure 1.1.

Figure 1.1: Contrasting ecological strategies of microbes

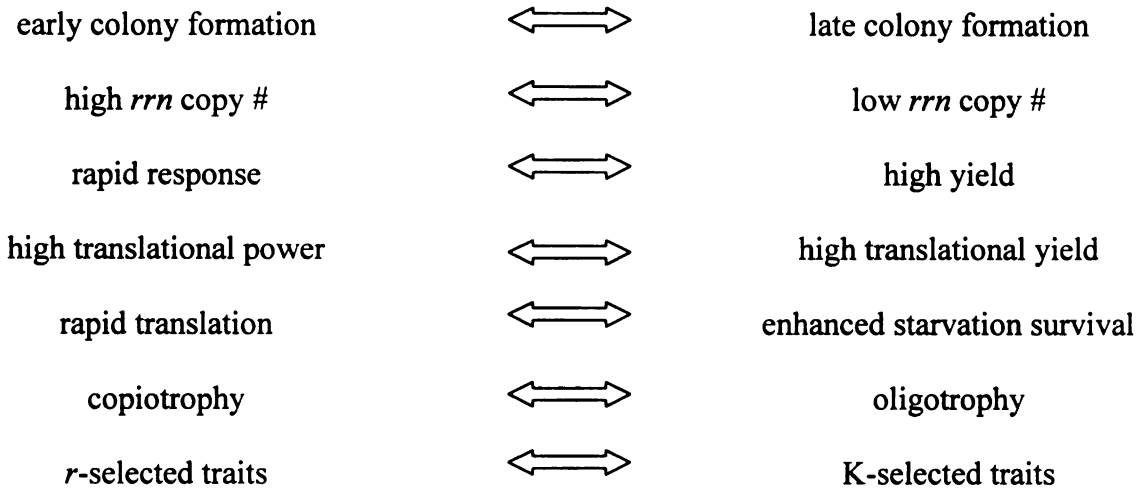


Figure 1.1. Examples of contrasting ecological strategies available to microbes. The pair of characteristics on each line have been suggested to be related by one or more evolutionary tradeoffs. The characteristics listed in each column are not necessarily synonymous or equivalent. However, an organism may often be described by several of the terms in a single column, either because of a mechanistic relationship between the traits or because the traits are favored under similar ecological circumstances. Of course, none of the contrasts implies merely a dichotomy, but rather a spectrum of possibilities. As an operational definition, the organisms we refer to as rapid responders or slow responders in this dissertation were classified as such on the basis of colony appearance time and *rrn* gene copy number.

Conditions favoring translational power or translational yield

Assuming the existence of a tradeoff between translational power and translational yield, we will describe conditions in which we expect each strategy to be favored. Dramatic fluctuations of resource availability occur when bacteria respond to episodes of resource abundance with exponential growth that continues until the resources are consumed; these conditions resemble serial batch cultivation of bacteria in the laboratory. We hypothesize that competition for abundant but ephemeral resources favors high translational power to permit fast growth. In energy-replete conditions, low translational yield would be deleterious only insofar as it limits the growth rate; the waste of resources *per se* would be irrelevant. The periods of starvation between episodes of resource abundance may favor high translational yield over high translational power; however, we anticipate a threshold effect. Unless starvation lasts long enough so that a subset of organisms experience mortality, or at least some impairment of their subsequent ability to grow, differences in translational yield may not influence relative fitness. We imagine that translational yield is only selected to be high enough to ensure survival through the longest starvation intervals that are typical for a habitat.

If the resource flux in a habitat is reasonably well matched over some period of time with the rate of bacterial mortality (e.g., from bactivory or viral lysis), conditions may more closely resemble continuous culture in the laboratory, with continuous growth at a more or less constant levels of resource availability and bacterial population density. Perhaps counter-intuitively, selection in such conditions favors the fastest possible growth at the ambient resource concentration, and hence high translational power over high

translational yield [87]. This remains true even if the resource concentration is low enough to constrain the growth rate of a strain far below its maximal value. As long as growth continues and the resources to support growth are drawn from a common pool, organisms that grow more quickly are superior competitors regardless of yield. Improvements in translational yield may improve fitness in these conditions, but only if they have the effect of increasing growth rates at the ambient resource concentrations. There is no fitness benefit in conserving resources at the cost of even a slight reduction in growth rate, if the resources spared from consumption are shared among all competitors.

We predict that conditions favoring high translational yield over high translational power can occur either when not all organisms are growing, or when resources are not shared. If the minimal resource levels that can support growth or viability are influenced by the energy demand for maintenance protein synthesis, higher translational yield would lower these resource thresholds. If resource levels in a habitat are such that some organisms can grow while other cannot, or some organisms can remain viable while others cannot, then selection will favor high translational yield. Selection for improved translational power would not be absent, but it would occur only among the organisms that are capable of growth. Even among growing strains, we expect high translational yield to be favored if resources are not shared. Imagine a microbial population in a habitat where resource levels only occasionally rise above the minimal level at which resource uptake occurs, and at most attain concentrations only slightly above the minimal level. The threshold of 'only slightly above the minimal level' is such that all available resources are taken up by the existing population before cells can grow sufficiently to increase their capacity for

uptake. The assimilated resources are endogenous, i.e. not shared. If growth is sustained primarily by drawing on endogenous resources that are periodically replenished, instead of by continuous uptake of resources from a common pool, selection will favor increasing the translational yield. The crucial distinction between these conditions and conditions favoring translational power is not low resource availability *per se,* but the fact that an organism and its progeny have no opportunity to obtain additional resources by growing more quickly. As long as this condition holds, translational yield will be favored.

The remainder of this dissertation examines the hypothesis that bacteria capable of a rapid growth response to abundant resources have high translational power, whereas bacteria with a slow growth response have low translational power. Chapter 2 describes our choice of bacterial strains to test this hypothesis, the methods developed to measure bacterial macromolecular content, and the macromolecular data itself, as well as related data on cell size and cell density. Chapter 3 compares translational power between rapidly and slowly responding bacteria, using the data of Chapter 2, and applies codon analysis to test the hypothesis that the strength of selection for translational power varies along the same ecological axis. Chapter 3 also discusses evidence that low translational yield is an unavoidable consequence of high translational power, providing a mechanistic basis for selection against high translational power under some environmental conditions. Because biased codon use can increase translational power, and because the degree of codon bias tends to be greater among rapidly responding bacteria, Chapter 4 investigates the possibility that differences in the degree of codon bias are a sufficient explanation for the observed differences in translational power among microbes.

Chapter 1 References

1. Shahab, N., F. Flett, S.G. Oliver and P.R. Butler, 1996. *Growth rate control of protein and nucleic acid content in Streptomyces coelicolor A3(2) and Escherichia coli B/r.* Microbiology-UK, **142**:1927-1935.

2. Pang, H.L. and H.H. Winkler, 1994. *The concentrations of stable RNA and ribosomes in Rickettsia prowazekii.* Molecular Microbiology, **12**(1):115-120.

3. Fegatella, F., J. Lim, S. Kjelleberg and R. Cavicchioli, 1998. *Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium Sphingomonas sp. strain RB2256.* Applied and Environmental Microbiology, **64**(11):4433-4438.

4. Chant, J., I. Hui, D. Dejongwong, L. Shimmin and P.P. Dennis, 1986. *The protein synthesizing machinery of the Archaebacterium Halobacterium cutirubrum - molecular characterization.* Systematic and Applied Microbiology, **7**(1):106-114.

5. Bremer, H. and P.P. Dennis, 1996. *Modulation of chemical composition and other parameters of the cell by growth rate*; pp. 1553-1569 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

6. Maaløe, O. and N.O. Kjeldgaard, 1966. *Control of Macromolecular Synthesis.* Microbial and Molecular Biology Series, ed. B.D. Davis. W.A. Benjamin, Inc., New York.

7. Maaløe, O., 1979. *Regulation of the protein-synthesizing machinery - ribosomes, tRNA, factors, and so on*; pp. 487-542 in *Gene Expression*; R.F. Goldberger, ed. Plenum Press, New York.

8. Koch, A.L. and C.S. Deppe, 1971. *In vivo assay of protein synthesizing capacity of Escherichia coli from slowly growing chemostat cultures.* Journal of Molecular Biology, **55**:549-562.

9. Koch, A.L., 1980. *Inefficiency of ribosomes functioning in Escherichia coli growing at moderate rates.* Journal of General Microbiology, **116**(1):165-171.

10. Wada, A., K. Igarashi, S. Yoshimura, S. Aimoto and A. Ishihama, 1995. *Ribosome modulation factor - stationary growth phase-specific inhibitor of ribosome functions from Escherichia coli.* Biochemical and Biophysical Research Communications, **214**(2):410-417.

11. Wada, A., R. Mikkola, C.G. Kurland and A. Ishihama, 2000. *Growth phase-coupled changes of the ribosome profile in natural isolates and laboratory strains of Escherichia coli.* Journal of Bacteriology, **182**(10):2893-2899.

12. Yamagishi, M., H. Matsushima, A. Wada, M. Sakagami, N. Fujita and A. Ishihama, 1993. *Regulation of the Escherichia coli rmf gene encoding the ribosome modulation factor - growth phase-dependent and growth rate-dependent control.* EMBO Journal, **12**(2):625-630.

13. Forchhammer, J. and L. Lindahl, 1971. *Growth rate of polypeptide chains as a function of cell growth rate in a mutant of Escherichia coli 15.* Journal of Molecular Biology, **55**(3):563-568.

14. Ehrenberg, M. and C.G. Kurland, 1984. *Costs of accuracy determined by a maximal growth rate constraint.* Quarterly Reviews of Biophysics, **17**(1):45-82.

15. Kurland, C.G., 1987. *Strategies for efficiency and accuracy in gene expression.* Trends in Biochemical Sciences, **12**(4):126-128.

16. Kjeldgaard, N.O. and C.G. Kurland, 1963. *The distribution of soluble and ribosomal RNA as a function of growth rate.* Journal of Molecular Biology, **6**:341-348.

17. Kurland, C.G., D. Hughes and M. Ehrenberg, 1996. *Limitations of translational accuracy*; pp. 979-1004 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

18. Andersson, S.G.E. and C.G. Kurland, 1990. *Codon preferences in free-living microorganisms.* Microbiological Reviews, **54**(2):198-210.

19. Sharp, P.M. and W.H. Li, 1987. *The Codon Adaptation Index - a measure of directional synonymous codon usage bias, and its potential applications.* Nucleic Acids Research, **15**(3):1281-1295.

20. Adkins, C.J., 1983. *Equilibrium Thermodynamics*, 3rd. ed. Cambridge University Press, Cambridge, U.K.

21. Begon, M., J.L. Harper and C.R. Townsend, 1996. *Ecology: Individuals, Populations and Communities*, 3rd. ed. Blackwell Science, Oxford.

22. French, A.P., 1971. *Newtonian Mechanics*. The M.I.T. Introductory Physics Series. W. W. Norton & Co., New York.

23. Odum, H.T. and R.C. Pinkerton, 1955. *Time's speed regulator: The optimum efficiency for maximum power output in physical and biological systems.* American Scientist, **43**:331-343.

24. Smith, C.C., 1976. *When and how much to reproduce - trade-off between power and efficiency.* American Zoologist, **16**(4):763-774.

25. Silvert, W., 1982. *The theory of power and efficiency in ecology.* Ecological Modelling, **15**(2):159-164.

26. Watt, W.B., 1986. *Power and efficiency as indexes of fitness in metabolic organization.* American Naturalist, **127**(5):629-653.

27. Mansson, B.A. and J.M. McGlade, 1993. *Ecology, thermodynamics and H.T. Odum's conjectures.* Oecologia, **93**(4):582-596.

28. Corning, P.A. and S.J. Kline, 1998. *Thermodynamics, information and life revisited. Part I: 'To be or entropy'.* Systems Research and Behavioral Science, **15**(4):273-295.

29. Arendt, J.D., 1997. *Adaptive intrinsic growth rates: An integration across taxa.* Quarterly Review of Biology, **72**(2):149-177.

30. Reichman, O.J. and D. Oberstein, 1977. *Selection of seed distribution types by Dipodomys merriami and Perognathus amplus.* Ecology, **58**(3):636-643.

31. Brown, J.S., B.P. Kotler and W.A. Mitchell, 1994. *Foraging theory, patch use, and the structure of a Negev desert granivore community.* Ecology, **75**(8):2286-2300.

32. Mueller, P. and J. Diamond, 2001. *Metabolic rate and environmental productivity: Well-provisioned animals evolved to run and idle fast.* Proceedings of the National Academy of Sciences of the United States of America, **98**(22):12550-12554.

33. Riha, V.F. and K.A. Berven, 1991. *An analysis of latitudinal variation in the larval development of the wood frog (Rana sylvatica).* Copeia, (1):209-221.

34. Schmitt, R.J., 1996. *Exploitation competition in mobile grazers: Trade-offs in use of a limited resource.* Ecology, **77**(2):408-425.

35. Gotthard, K., S. Nylin and C. Wiklund, 1994. *Adaptive variation in growth rate - life history costs and consequences in the speckled wood butterfly, Pararge aegeria.* Oecologia, **99**(3-4):281-289.

36. Reznick, D., L. Nunney and A. Tessier, 2000. *Big houses, big cars, superfleas and the costs of reproduction.* Trends in Ecology & Evolution, **15**(10):421-425.

37. Tessier, A.J., M.A. Leibold and J. Tsao, 2000. *A fundamental trade-off in resource exploitation by Daphnia and consequences to plankton communities.* Ecology, **81**(3):826-841.

38. Tessier, A.J. and P. Woodruff, 2002. *Trading off the ability to exploit rich versus poor food quality.* Ecology Letters, **5**(5):685-692.

39. Kobe, R.K., S.W. Pacala, J.A. Silander and C.D. Canham, 1995. *Juvenile tree survivorship as a component of shade tolerance.* Ecological Applications, **5**(2):517-532.

40. Chapin, F.S., 1980. *The mineral nutrition of wild plants.* Annual Review of Ecology and Systematics, **11**:233-260.

41. McGraw, J.B. and F.S. Chapin, 1989. *Competitive ability and adaptation to fertile and infertile soils in 2 Eriophorum species.* Ecology, **70**(3):736-749.

42. Chapin, F.S., K. Autumn and F. Pugnaire, 1993. *Evolution of suites of traits in response to environmental stress.* American Naturalist, **142**:S78-S92.

43. Sommer, U., 1985. *Comparison between steady state and non-steady state competition - experiments with natural phytoplankton.* Limnology and Oceanography, **30**(2):335-346.

44. Matin, A. and H. Veldkamp, 1978. *Physiological basis of selective advantage of a Spirillum sp. in a carbon-limited environment.* Journal of General Microbiology, **105**(4):187-197.

45. Jannasch, H.W., 1979. *Microbial ecology of aquatic low nutrient habitats*; pp. 243-260 in *Strategies of Microbial Life in Extreme Environments: report of the Dahlem Workshop*; M. Shilo, ed. Verlag Chemie, Weinhein.

46. Veldkamp, H., H. Van Gemerden, W. Harder and H.J. Laanbroek, 1984. *Competition among bacteria: an overview*; pp. 279-290 in *Current Perspectives in Microbial Ecology*; M.J. Klug and C.A. Reddy, eds. American Society for Microbiology, Washington, D. C.

47. Cavicchioli, R., M. Ostrowski, F. Fegatella, A. Goodchild and N. Guixa-Boixereu, 2003. *Life under nutrient limitation in oligotrophic marine environments: An eco/physiological perspective of Sphingopyxis alaskensis (formerly Sphingomonas alaskensis).* Microbial Ecology, **45**(3):203-217.

48. Helling, R.B., 2002. *Speed versus efficiency in microbial growth and the role of parallel pathways.* Journal of Bacteriology, **184**(4):1041-1045.

49. Velicer, G.J. and R.E. Lenski, 1999. *Evolutionary trade-offs under conditions of resource abundance and scarcity: Experiments with bacteria.* Ecology, **80**(4):1168-1179.

50. Velicer, G.J., T.M. Schmidt and R.E. Lenski, 1999. *Application of traditional and phylogenetically based comparative methods to test for a trade-off in bacterial growth rate at low versus high substrate concentration.* Microbial Ecology, **38**(3):191-200.

51. Vasi, F., M. Travisano and R.E. Lenski, 1994. *Long term experimental evolution in Escherichia coli. 2. Changes in life history traits during adaptation to a seasonal environment.* American Naturalist, **144**(3):432-456.

52. Ingraham, J.L., O. Maaløe and F.C. Neidhardt, 1983. *Growth of the Bacterial Cell.* Sinauer, Sunderland, Massachusetts.

53. Neidhardt, F.C., J.L. Ingraham and M. Schaechter, 1990. *Physiology of the Bacterial Cell.* Sinauer, Sunderland, Massachusetts.

54. Dong, H.J., L. Nilsson and C.G. Kurland, 1996. *Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates.* Journal of Molecular Biology, **260**(5):649-663.

55. Rosset, R., J. Julien and R. Monier, 1966. *Ribonucleic acid composition of bacteria as a function of growth rate.* Journal of Molecular Biology, **18**:308-320.

56. Berg, O.G. and C.G. Kurland, 1997. *Growth rate-optimised tRNA abundance and codon usage.* Journal of Molecular Biology, **270**(4):544-550.

57. Poulsen, L.K., G. Ballard and D.A. Stahl, 1993. *Use of ribosomal RNA fluorescence in situ hybridization for measuring the activity of single cells in young and established biofilms.* Applied and Environmental Microbiology, **59**(5):1354-1360.

58. Parrott, L.M. and J.H. Slater, 1980. *The DNA, RNA and protein composition of the Cyanobacterium Anacystis nidulans grown in light-limited and carbon dioxide-limited chemostats.* Archives of Microbiology, **127**(1):53-58.

59. Riesenberg, D. and F. Bergter, 1979. *Dependence of macromolecular composition and morphology of Streptomyces hygroscopicus on specific growth rate.* Zeitschrift fur Allgemeine Mikrobiologie, **19**(6):415-430.

60. Mink, R.W. and R.B. Hespell, 1981. *Survival of Megasphaera elsdenii during starvation.* Current Microbiology, **5**(1):51-56.

61. Boudreaux, D.P. and V.R. Srinivasan, 1981. *A continuous culture study of growth of Bacillus cereus T.* Journal of General Microbiology, **122**(1):129-136.

62. Mink, R.W. and R.B. Hespell, 1981. *Long term nutrient starvation of continuously cultured (glucose-limited) Selenomonas ruminantium.* Journal of Bacteriology, **148**(2):541-550.

63. Mink, R.W., J.A. Patterson and R.B. Hespell, 1982. *Changes in viability, cell composition, and enzyme levels during starvation of continuously cultured (ammonia-limited) Selenomonas ruminantium.* Applied and Environmental Microbiology, **44**(4):913-922.

64. Tempest, D.W. and J.W. Dicks, 1967. *Inter-relationships between potassium, magnesium, phosphorus and ribonucleic acid in the growth of Aerobacter aerogenes in a chemostat*; pp. 140-153 in *Microbial Physiology and Continuous Culture, Proceedings of the Third International Symposium*; E.O. Powell, C.G.T. Evans, R.E. Strange and D.W. Tempest, eds. Her Majesty's Stationary Office, London.

65. Beresford, T. and S. Condon, 1993. *Physiological and genetic regulation of ribosomal RNA synthesis in Lactococcus.* Journal of General Microbiology, **139**:2009-2017.

66. Farewell, A. and F.C. Neidhardt, 1998. *Effect of temperature on in vivo protein synthetic capacity in Escherichia coli.* Journal of Bacteriology, **180**(17):4704-4710.

67. Schaechter, M., O. Maaloe and N.O. Kjeldgaard, 1958. *Dependency on medium and temperature of cell size and chemical composition during balanced growth of Salmonella typhimurium.* Journal of General Microbiology, **19**(3):592-606.

68. Mikkola, R. and C.G. Kurland, 1988. *Media dependence of translational mutant phenotype.* FEMS Microbiology Letters, **56**(3):265-269.

69. Lafay, B., J.C. Atherton and P.M. Sharp, 2000. *Absence of translationally selected synonymous codon usage bias in Helicobacter pylori.* Microbiology-UK, **146**:851-860.

70. Lafay, B., A.T. Lloyd, M.J. McLean, K.M. Devine, P.M. Sharp and K.H. Wolfe, 1999. *Proteome composition and codon usage in spirochaetes: Species-specific and DNA strand-specific mutational biases.* Nucleic Acids Research, **27**(7):1642-1649.

71. Ochman, H. and R.K. Selander, 1984. *Standard reference strains of Escherichia coli from natural populations.* Journal of Bacteriology, **157**(2):690-693.

72. Mikkola, R. and C.G. Kurland, 1991. *Is there a unique ribosome phenotype for naturally-occurring Escherichia coli?* Biochimie, **73**(7-8):1061-1066.

73. Kurland, C.G. and R. Mikkola, 1993. *The impact of nutritional state on the microevolution of ribosomes*; pp. 225-237 in *Starvation in Bacteria*; S. Kjelleberg, ed. Plenum Press, New York.

74. Mikkola, R. and C.G. Kurland, 1992. *Selection of laboratory wild-type phenotype from natural isolates of Escherichia coli in chemostats.* Molecular Biology and Evolution, **9**(3):394-402.

75. MacArthur, R.H. and E.O. Wilson, 1967. *The Theory of Island Biogeography.* Monographs in Population Biology. Princeton University Press, Princeton, New Jersey.

76. Pianka, E.R., 1970. *r-selection and K-selection.* American Naturalist, **104**(940):592-597.

77. Andrews, J.H. and R.F. Harris, 1986. *r-selection and K-selection and microbial ecology.* Advances in Microbial Ecology, **9**:99-147.

78. Hirsch, P., M. Bernhard, S.S. Cohen, J.C. Ensign, H.W. Jannasch, A.L. Koch, K.C. Marshall, A. Matin, J.S. Poindexter, S.C. Rittenberg, D.C. Smith and H. Veldkamp, 1979. *Life under conditions of low nutrient concentrations group report*; pp. 357-372 in *Strategies of Microbial Life in Extreme Environments: report of the Dahlem Workshop*; M. Shilo, ed. Verlag Chemie, Weinhein.

79. Poindexter, J.S., 1981. *Oligotrophy - fast and famine existence.* Advances in Microbial Ecology, **5**:63-89.

80. Koch, A.L., 1979. *Microbial growth in low concentrations of nutrients*; pp. 261-279 in *Strategies of Microbial Life in Extreme Environments: report of the Dahlem Workshop*; M. Shilo, ed. Verlag Chemie, Weinhein.

81. Morita, R.Y., 1997. *Bacteria in Oligotrophic Environments.* Chapman & Hall, New York.

82. Condon, C., D. Liveris, C. Squires, I. Schwartz and C.L. Squires, 1995. *Ribosomal RNA operon multiplicity in Escherichia coli and the physiological implications of rrn inactivation.* Journal of Bacteriology, **177**(14):4152-4156.

83. Stevenson, B.S. and T.M. Schmidt, 1998. *Growth rate-dependent accumulation of RNA from plasmid-borne rRNA operons in Escherichia coli.* Journal of Bacteriology, **180**(7):1970-1972.

84. Stevenson, B.S., 2000. *Life history implications of ribosomal RNA gene copy number in Escherichia coli*, Ph.D. Dissertation, Microbiology Department, Michigan State University. East Lansing.

85. Klappenbach, J.A., 2001. *Ecological and evolutionary implications of ribosomal RNA gene copy number in heterotrophic soil bacteria*, Ph.D. Dissertation, Department of Microbiology and Molecular Genetics, Michigan State University. East Lansing.

86. Klappenbach, J.A., J.M. Dunbar and T.M. Schmidt, 2000. *rRNA operon copy number reflects ecological strategies of bacteria.* Applied and Environmental Microbiology, **66**(4):1328-1333.

87. Hansen, S.R. and S.P. Hubbell, 1980. *Single nutrient microbial competition: Qualitative agreement between experimental and theoretically forecast outcomes.* Science, **207**:1491-1493.

Chapter 2. Macromolecular composition of bacterial soil isolates representing contrasting ecological strategies

Chapter 2 Abstract

The growth rate, culture density, cell size and macromolecular composition were assessed in batch culture in two different media for each of 10 bacterial strains. The 8 recent soil isolates and 2 well-characterized strains were chosen to represent either of 2 contrasting ecological strategies. The 'rapid responder' strategy was defined operationally by the ability to form a visible colony within 2 days during initial isolation and by possession of 4 or more copies of the ribosomal RNA operon per genome. 'Slow responders' required a week or more to form a visible colony during initial isolation and possessed 1 or 2 ribosomal RNA operons per genome. Each ecological strategy was represented by members of 4 diverse phylogenetic groups commonly found in soil. Some trends in the data were consistent with expectations, such as the rank order of DNA, RNA and protein mass within strains, and a positive correlation of both cell size and the RNA:DNA ratio with growth rate across strains. However, the variability between strains is such that trends observed across strains are often of limited value in predicting the characteristics of individual strains. Overall, the macromolecular composition of rapidly responding and slowly responding bacteria are similar, although their growth rates in the same medium are considerably different.

Chapter 2 Introduction

Knowledge of the macromolecular composition of bacteria is fundamental for understanding bacterial physiology and ecology. For example, the investigations of the Copenhagen School of microbiology into the DNA, RNA and protein content of enteric bacteria during stable and transient states of growth in a variety of culture conditions [1,2] led to fundamental insights about the control and coordination of bacterial growth [3,4]. These insights provide a system-wide view of the behavior of the self-replicating, homeostatic metabolic network that is a bacterial cell [5,6]. An explanation of this behavior remains the goal of ongoing efforts to develop highly detailed models of bacterial cell growth based on genomic, proteomic and metabolomic data [7,8]. At the opposite end of the biological size scale, bacterial macromolecular content influences global cycles of carbon, nitrogen, phosphorus and other elements [9-11], because macromolecular content is linked to both the elemental composition and the growth rate of bacteria [12].

An increasing awareness of the scope of microbial diversity reveals that the majority of previous studies of macromolecular content have focused on a small group of closely related bacteria that share similar ecological characteristics. *Escherichia coli* and other enteric bacteria have received the most attention; relatively few studies have investigated microbes from the many other phylogenetic lineages that exist. Although some studies in the latter group have noted contrasts with the patterns observed in *E. coli* [13-16], we have not found any published report comparing the DNA, RNA and protein content of different bacterial species chosen specifically to represent either evolutionary or ecological diversity.

Because the fundamental roles of the 3 major macromolecules are similar in all known cellular life, with DNA as an information storage molecule, RNA providing the central components of the protein synthetic machinery, and protein as the predominant functional molecule, we expect similarities in the macromolecular composition of microbes regardless of their phylogenetic affiliation. Nonetheless, comparisons of macromolecular content across a diverse group of microbes are important for investigating the generality of the patterns found among the enteric bacteria.

Chapter 2 Materials and Methods

Unless specifically mentioned, Sigma was the source for all chemicals.

Strains and media

The soil isolates used in this study were all obtained from long term plating experiments that followed essentially the same procedure [17-19], designed to permit the isolation of strains which vary widely in the time required for the formation of a visible colony on solid media. Briefly, dilutions of a soil suspension were spread onto a dilute nutrient solution solidified with 1.5% agar; plates were incubated for at least 10 days at room temperature with adequate humidity to prevent dehydration. Plates were examined daily for new colonies; colonies were marked uniquely by the day of appearance to allow subsequent recovery of isolates with known colony appearance times. Strains S21027/HF3 and S24542/HS5 were a gift from T. Hattori, derived from rice paddy soil in Japan, using a 1/100 dilution of Nutrient Broth (Difco) as the nutrient solution [17]. The remaining soil isolates were a gift from J. Klappenbach, derived from agricultural soil at the Long Term Ecological Research site at the Kellogg Biological Station, Michigan, USA [18,19]. The nutrient solution used for isolation of strains EC2, EC4, EC5 and LC9 was a 1/100 dilution of Nutrient Broth (Difco); for strains PX3.14 and PX3.15 the nutrient solution was 5 mM succinate in a basal salts solution. These 8 strains were chosen from a larger collection of soil isolates obtained in the same fashion, with the goal of obtaining related pairs of soil bacteria from a number of different phylogenetic groups. Strains that formed spores or that did not grow with dispersed turbidity in liquid culture were deliberately avoided to facilitate growth rate measurements.

The two members of each related pair of strains were selected to represent contrasting ecological strategies that we characterize as a rapid growth response or a slow growth response to resource abundance. The 'rapid responders' were isolated from colonies that were first visible 1-2 days after plating, and contained 4 or more copies of the ribosomal RNA (*rrn*) operon per genome USA [18,19]. The 'slow responders' were isolated from colonies that were first visible 7 days or more after plating, and contained 1 or 2 *rrn* copies per genome [18,19]. While the time interval between plating and the formation of visible colonies decreased for both groups during subsequent cultivation in the lab, the rapid responders continued to form colonies more quickly than the slow responders on a variety of solid media. In addition to these 8 soil isolates, we investigated two well-characterized strains representing the contrasting ecological strategies. *E. coli* REL607, a derivative of *E. coli* B/r, was a gift from R. Lenski [20] and represented a rapidly responding strain. *Sphingopyxis alaskensis* RB2256 (originally *Sphingomonas*), a marine ultramicrobacterium representing a slowly responding strain, was a gift from R. Cavicchioli and M. Ostrowski [21,22]. This strain was isolated by dilution culture techniques as described by Schut and colleagues [23,24]. For all strains, freezer stocks in R2BV (see below) with 10% glycerol were prepared and stored at -80°C soon after the isolation of soil strains or immediately after receiving strains into the lab, to minimize adaptation to the lab environment.

Each of these 10 strains was grown in batch culture in two media, to make cell measurements and to harvest biomass during balanced exponential growth at two

different growth rates. The first medium for all strains was R2BV, containing per liter: glucose, 0.5 g; proteose peptone, 0.5 g; yeast extract, 0.5 g; casamino acids, 0.5 g; soluble starch, 0.5 g; KH_2PO_4, 0.3 g; sodium pyruvate, 0.3 g; $MgSO_4$, 0.05 g; and 1 ml vitamin solution. For most strains, the second medium was PVY/10, containing per liter: Na_2HPO_4, 0.71 g; KH_2PO_4, 0.68 g; $(NH_4)_2SO_4$, 0.66 g; $MgSO_4$, 0.12 g; proteose peptone, 0.5 g; yeast extract, 0.05 g; and 1 ml each vitamin solution and trace element solution. *E. coli*, EC2, HS5 and LC9 did not grow adequately in PVY/10, so the second medium for these strains was R2B-GCS, which is identical to R2BV with the omission of glucose, casamino acids, soluble starch, and vitamins. The vitamin solution contains, per liter: 200 mg of i-inositol, and 100 mg of each of the following: biotin, choline-Cl, folic acid, lipoic acid, nicotinamide, pantothenate, para-aminobenzoic acid, pyridoxal HCl, riboflavin, and thiamine HCl. The trace elements solution contains, per liter: concentrated (11.6 N) HCl, 6.03 ml; $FeSO_4 \cdot 7H_2O$, 2.085 g; $ZnSO_4$, 143.6 mg; $MnCl_2 \cdot 4H_2O$, 89.1 mg; H_3BO_3 6.2 mg; $CoCl \cdot 6H_2O$, 190.3 mg; $CuCl_4 \cdot 2H_2O$, 1.7 mg; $NiCl_2 \cdot 6H_2O$, 23.8 mg; $Na_2MoO_4 \cdot 2H_2O$, 48.4 mg.

Growth and harvest

Cells were grown aerobically at 25°C in 7-10 ml medium in 18 mm diameter tubes shaking at 200 rpm (*E. coli*, HF3, EC2, EC4, EC5 and PX3.15), or in 45 ml medium in baffled nephelometer flasks of 250 ml capacity shaking at 100 rpm (*S. alaskensis,* PX3.14, HS5 and LC9). Growth was monitored by turbidity in a Spectronic 20D+ spectrophotometer (Milton Roy) at 420 nm; this short wavelength was chosen to maximize sensitivity at low cell density. Biomass for macromolecular analysis was

obtained by inoculating a pair of tubes from freezer stock; during exponential growth, one of these cultures provided a 1% or smaller inoculum for 4 additional tubes containing the same medium. If the exponential growth rates in these tubes were consistent with each other and with prior experience, one of these tubes provided a 1% or smaller inoculum for a set of culture vessels (4 tubes or 3 flasks) containing the same medium, used to harvest bacterial biomass for macromolecular analysis. For strain HF3 growing in PVY/10, one additional transfer was made prior to harvest when the target optical density for harvest was missed. All culture vessels inoculated at the same time needed to show essentially the same growth rate, with at least 4 turbidity measurements obtained during exponential growth, and a final optical density near the target value, in order to be used for macromolecular analysis. Specific growth rates were calculated as the slope of a linear regression of the natural logarithm of OD_{420} with time for each culture vessel, using the final 4 optical density measurements prior to harvest. All growth rates are reported as the specific growth rate $\mu = \ln(2)/g$, where g is the generation time.

The target density for harvest of biomass was half the optical density at which departure from exponential growth was first detectable during preliminary growth experiments, or at $OD_{420}= 0.2$, whichever was lower, so that cultures were at least one generation time away from a detectable response to changing culture conditions. At harvest, for all strains except PX3.15, 3-4 replicate aliquots (analytical replicates) containing 2-10 ml of culture were obtained from each replicate culture vessel (experimental replicates) of a particular strain-medium combination. For strain PX3.15 in each growth medium, 8 tubes (experimental replicates) containing 10 ml culture each contributed a single 9.5 ml

aliquot for macromolecular analysis. Culture aliquots for determination of DNA, RNA and protein content were centrifuged at 10,000*g* and 4°C for 10-30 minutes, medium was decanted, and cell pellets were stored at –80°C until analysis. Also at the time of harvest, separate culture aliquots were fixed in formaldehyde for cell enumeration, and photomicrographs of cells were taken for biovolume determination.

Overview of macromolecular analysis

A single run through the macromolecular analysis protocol involved 1-2 days work with 8 cell pellets; with the exception of strain PX3.15, analysis of pellets from different culture replicates, media types and strains were dispersed over many days of analysis to control for any day-to-day variation in measurements. The initial cultures of strain PX3.15 in both media provided inadequate biomass for analysis. Larger biomass samples from subsequent cultures of this strain were analyzed after the analysis of other strains was completed, with material from replicate cultures in both media analyzed over two separate runs.

For a single run of the macromolecular analysis protocol, 8 cell pellets were thawed in a lysis buffer and sonicated to disrupt cells. Sonicated material from each pellet was subsampled for DNA, RNA and protein measurements; each macromolecule was measured in multiple, independent dilution series of sonicated cell material in a 96-well plate format. Thus, there was analytical replication at each step of analysis for each molecule measured from a single cell pellet, as well as analytical replication of cell pellets derived from a single culture vessel (except for strain PX3.15, for which each cell

pellet represented a distinct culture vessel). This approach allowed individual outlying values to be recognized and discarded.

Cell lysis

Sonication was the method chosen for cell lysis (in combination with freeze-thaw treatment for gram positive strains) to permit rapid, complete lysis without loss of material or addition of reagents that could interfere with subsequent analysis. To minimize nuclease and other enzymatic activity, bacterial cell material was held continuously on ice or in ice water baths after removal from storage at -80°C until shortly before fluorescence measurements for nucleic acid determination or until the start of trichloroacetic acid (TCA) extraction for protein determination. Initial fluorescence values were not affected by an overnight room temperature incubation of the prepared 96-well plates, suggesting that nuclease activity subsequent to warming of the material did not affect nucleic acid measurements (data not shown).

Cell pellets were generally thawed in 3 ml of ice-cold nuclease-free lysis buffer (10mM Tris, 1mM ethylenediaminetetraacetic acid (EDTA), 2% (v/v) ethanol, 0.05% (v/v) Igepal CA-630 detergent, 200 mg/l sodium deoxycholate, pH 7.0). However, some strain-medium combinations with cell pellets of low macromolecular content were thawed in only 1.5 ml lysis buffer. Cell pellets from gram positive strains EC5 and PX3.15 were subjected to 5 cycles of freeze-thaw between a dry ice-ethanol bath and a 95°C water bath prior to the addition of lysis buffer. Cell material and lysis buffer were transferred to 15 ml capacity conical bottom glass centrifuge tubes (Kimble) for sonication; the conical shape maximizes the insertion depth of the sonication probe to reduce foaming,

and glass better conducts heat away from sonicated material to the surrounding ice water bath. Sonication used a 1/8″ tip tapered probe in a Sonifier model 450 (Branson) at half-maximum power with a 25% duty cycle for 90 seconds. This duration of sonication was 2-3 times the length of time after which no further increase in fluorescent signal was observed in preliminary experiments; sonication could continue for 8-10 minutes at these settings with no loss of fluorescent signal (data not shown). Sonicated cell material was investigated for the efficiency of cell lysis both by plating and by microscopy, which indicated that neither viable nor intact cells remained after sonication.

Nucleic acid determination

The proprietary fluorescent dyes PicoGreen and RiboGreen (Molecular Probes, [25,26]) were chosen for nucleic acid determination because of their sensitivity, wide dynamic range, and ease of use in comparison to traditional colorimetric methods such as the diphenylamine reaction for DNA and the orcinol reaction for RNA. Except as noted below, measurements followed the procedure suggested by the manufacturer (with the high range dye concentration for RiboGreen), using 150 µl diluted sample and 150 µl working dye solution per well in clear-bottom, black 96-well microtiter plates (Costar). Control experiments with reagents and cell material from all strain-medium combinations at the same concentrations used for nucleic acid measurements indicated that in the absence of the fluorescent dyes, fluorescence was negligible. Similarly, we found no inhibition of the expected fluorescence from known concentrations of DNA and RNA by cell material of any strain at the concentrations use for nucleic acid determination. A nonlinear fluorescence response at the highest concentrations of cell material from strain EC4 during early measurements may have indicated the presence of an inhibitor of

RiboGreen fluorescence. Additional dilution of EC4 cell material restored a linear relationship between the concentration of cell material and RiboGreen fluorescence response.

Fluorescence was measured from images obtained on a Storm 860 fluorescence scanner (Molecular Dynamics) using blue laser excitation with 100 micron resolution and 1000 volts PMT; edges of the microtiter plates were removed prior to sample loading so the flat well bottoms rested directly on the glass optical bed of the scanner. Fluorescent plate images were analyzed using ImageQuant 5.0 software (Molecular Dynamics). Fluorescence per well was calculated by summing pixel intensity within uniformly sized circles applied to the image of each well. The circles were sized slightly smaller than the wells; positions of the circles were adjusted manually as necessary to minimize the influence of dust or tiny bubbles which were evident in the images of some wells. Use of a fluorescent plate reader with the ability to stimulate PicoGreen and RiboGreen fluorescence at their optimal excitation wavelength would improve sensitivity compared to a fluorescent scanner, but would not provide the ability to recognize and sometimes compensate for artifacts in the wells.

Immediately after sonication, cell material was vortexed and triplicate aliquots were diluted 1:10 (or more for some strain-medium combinations) into 1.00 ml TE buffer (10 mM Tris, 1mM EDTA) for nucleic acid assays. For strains which required an initial dilution greater than 1:10 (and hence a smaller aliquot of sonicated cell material added to 1.00 ml TE, see below), a compensatory volume of the lysis buffer was added as well, so

the final volume and chemical environment of the triplicate initial dilutions was identical for all strains and media, equivalent to a 1:10 dilution of the lysis buffer into 1.00 ml TE. Each of the triplicate initial dilutions from a single cell pellet was assayed in a 3-well dilution series for each of the two fluorescent dyes. The dilution series comprised no additional dilution, an additional 2:3 dilution, and an additional 1:3 dilution subsequent to the initial 1:10 dilution. The diluent for these wells was a 1:10 dilution of lysis buffer in TE, so that the chemical environment provided by the buffer remained unchanged in all sample wells; only sonicated cell material was diluted.

Estimates of the DNA concentration in sample wells were derived from PicoGreen fluorescence, compared to a linear standard curve of PicoGreen fluorescence from a dilution series of λ phage DNA. PicoGreen is advertised as a DNA-specific dye; in a solution containing an equal mass of single-stranded RNA and double stranded DNA, the RNA-induced fluorescence is less than 5% of DNA-induced fluorescence when the ratio of PicoGreen to total nucleic acids is high [25]. However, as the ratio of PicoGreen to total nucleic acids decreases in a solution containing a 1:1 mass ratio of DNA and RNA, RNA-induced fluorescence can approach 20% of DNA fluorescence [25]. Thus, for sonicated cell material containing as much or more RNA than DNA, the cell material must be sufficiently dilute to ensure that PicoGreen fluorescence can be interpreted as a DNA signal. For this study, PicoGreen fluorescence from the dilution series of sonicated cell material was examined to ensure that different dilutions gave approximately the same estimate of DNA concentration in the original sample. A trend of higher DNA estimates with increasing sample concentration in all 3 dilution series from the independent initial

dilutions was taken as evidence that RNA-induced fluorescence was making an increased contribution to total PicoGreen fluorescence in the more concentrated samples. In this case, DNA estimates were derived only from the fluorescence of more dilute wells, and future measurements of cell pellets from that strain-medium combination used a higher initial dilution.

Estimates of RNA concentration in sample wells were derived by a mathematical subtraction procedure using both RiboGreen and PicoGreen fluorescence data, since RiboGreen fluorescence is induced by both RNA and DNA [26]. The DNA contribution to the RiboGreen fluorescence of a sample well was estimated from the PicoGreen-derived estimate of sample DNA content and a linear standard curve of RiboGreen fluorescence from a λ DNA dilution series. This estimated DNA-stimulated RiboGreen fluorescence was subtracted from the total RiboGreen fluorescence of sample wells; residual fluorescence was assumed to be due to RNA. The RNA-stimulated RiboGreen fluorescence was converted to an estimated concentration of RNA using a linear standard curve of RiboGreen fluorescence from a dilution series of pure *E. coli* rRNA.

Neither the protocol suggested by the manufacturer for DNase digestion prior to RNA measurement with RiboGreen, nor a potential alternative of estimating RNA content by RNase-induced fluorescence loss were followed, both because a preliminary test indicated that the digestion products of either nuclease inhibited RiboGreen fluorescence somewhat (data not shown), and because the subtraction protocol we implemented involved fewer manipulations at the bench. However, estimates of sample RNA

concentration from the subtraction procedure are expected to be less precise than estimates of DNA concentration with our approach, since errors of DNA measurement affect estimates of RNA concentration as well.

All nucleic acid standard curves used with PicoGreen and RiboGreen were triplicate 5 point, 3-fold dilution series beginning at 1000 ng/ml (rRNA) or 500 ng/ml (λ DNA), plus 5 sonicated reagent blank wells and 5 diluent blank wells. RNA and DNA standard solutions were prepared at 10x concentrations (10 μg/ml and 5 μg/ml respectively) in lysis buffer, sonicated, and diluted 1:10 in TE, following the same procedure as for samples, prior to making 3-fold dilutions to the working standard concentrations and storing single-use aliquots of standards at -80°C. The sonicated reagent blank was lysis buffer, sonicated as for samples and diluted 1:10 in TE. Both the PicoGreen and RiboGreen fluorescence at a given sample dilution were required to be within the range of fluorescence observed in the nucleic acid standards, otherwise wells at that dilution were excluded from calculations. If the fluorescence was too high from the most concentrated sample wells, subsequent measurements of cell pellets from that strain-medium combination used a higher initial dilution of sonicated cell material. If fluorescence was too low from the least concentrated sample wells, multiple cell pellets from a single culture vessel of that strain and medium were combined during subsequent analyses if possible, otherwise single pellets were resuspended in only 1.5 ml lysis buffer instead of 3.0 ml to attain a higher concentration of cell material.

Nucleic acid measurements of individual wells deviating by more than 3 studentized residuals from the mean of replicate wells were discarded; if 2 of the 3 wells from a single initial dilution were excluded on this basis, the 3rd well from that initial dilution was excluded also. The intercept of a linear regression of all remaining fluorescence values from one cell pellet vs. sample concentration was compared to the mean fluorescence of blank wells. If the intercept differed by more than 10% from the measured blank and the omission of the most concentrated sample wells both improved the fit of the linear regression and brought the intercept within 10% of the measured blank, it was assumed that the total nucleic acid content of the most concentrated wells exceeded the linear range of the assay. The nucleic acid estimate derived from lower concentration wells was retained, and future cell pellets from that strain-medium combination were subjected to a higher initial dilution.

Protein determination

The bicinchoninic acid (BCA) assay was chosen for protein determination since it is more sensitive and suffers less from protein-protein variability than other colorimetric protein assays [27]; commercially available fluorescent protein dyes are highly sensitive to the chemical environment of the assay and not necessarily protein-specific. The MicroBCA assay kit (Pierce Chemical) was chosen for convenience; procedures followed the manufacturer's recommendations for microtiter plate measurements except as noted below. Hot TCA extraction of protein [28] from sonicated cell material was performed both to reduce interstrain variability in the chemical environment of the protein assay and to remove protein from the Tris-based buffer used for cell lysis.

Following sonication of all cell pellets during a single run through the macromolecular analysis protocol, duplicate 1.00 ml aliquots of sonicated cell material from each pellet were added to 500 ml of 30% TCA (w/v) in locking 1.7 ml capacity microcentrifuge tubes and immediately incubated at 80°C for 30 minutes. However, for cell pellets originally diluted into 1.5 ml lysis buffer instead of 3.0 ml, only a single 1.0 ml TCA extraction was performed. Tubes were then transferred to an ice water bath for 30 minutes, and centrifuged at 10,000*g* at 4°C for 30 minutes to pellet the precipitated proteins. TCA was decanted, tubes were spun briefly, and remaining TCA was removed by aspiration without disturbing the protein pellet. Protein pellets were either analyzed immediately or frozen at -20°C overnight prior to analysis.

Protein pellets were resuspended by adding 50 µl of alkaline SDS (5% (v/v) sodium dodecyl sulfate in 0.1 N NaOH) and shaking for 1 hour, followed by addition of 1.00 ml of saline (0.9% (w/v) NaCl). The protein solution derived from each independent TCA extraction was analyzed in 6 wells of an untreated, clear, flat-bottomed microtiter plate (Nunc), 3 wells at the concentration resulting from protein resuspension as described above, and 3 wells diluted another 1:3. The diluent in this case was alkaline SDS diluted 1:21 into saline, so the chemical environment of the assay was not changed by dilution. Wells for protein analysis contained 150 µl of protein solution and 150 µl of working BCA reagent; plates were incubated at room temperature for 90 minutes prior to reading absorbance at 562 nm on a Biokinetics EL312e plate reader (Bio-Tek Instruments). Immediately prior to reading the plates, it was necessary to pop any bubbles remaining in the wells with the tip of a hypodermic needle.

Protein standards were dilutions of bovine serum albumin (BSA, Pierce Chemical) in lysis buffer that had been sonicated and diluted 1:10 into TE as for cell samples. Multiple 1.00 ml aliquots of sonicated protein prepared in a single batch were added to 500 ml of 30% TCA in locking microcentrifuge tubes and frozen until use. Two such protein standard tubes were subjected to hot TCA extraction and protein resolubilization in parallel with samples; each tube contributed duplicate 5 point 2-fold dilution series of the protein standard (starting at 1,000 ng/ml or 500 ng/ml depending on the run) for a total of 4 protein standard dilution series per run of protein analysis.

Sonicated reagent blanks and diluent blanks (8 wells each) were included in each run. The sonicated reagent blanks were sonicated lysis buffer diluted into TE and subjected to TCA extraction and resolubilization exactly as for cell material and protein standards. Diluent blanks were a 1:21 dilution of alkaline SDS into saline, chemically identical to the material added to samples, protein standards and sonicated reagent blanks subsequent to TCA extraction. However, color development differed significantly between wells containing sonicated reagent blanks and diluent blanks (probably due to pH or chemical effects of residual TCA, data not shown). Thus, we expect the final absorbance of sample and standard wells to depend not only on protein content, but also on the degree of dilution subsequent to TCA extraction. Therefore, blank values subtracted from the raw absorbance of each well were specific to the dilution of the well, interpolated linearly between the absorbance of sonicated reagent blanks and of diluent blanks. Protein concentrations in sample wells were estimated from blank-corrected A_{562} values

according to a 2nd order standard curve fitted to the blank-corrected A_{562} values of standard wells.

Sample, standard and blank wells were excluded if A_{562} values differed by more than 3 studentized residuals from the mean of replicate wells. If there were significant differences between estimates of the original culture protein concentration derived from the two different assay concentrations, the lower concentration was invariably found to be close to the detection limit of the assay and those wells were excluded. Subsequent measurements from that strain-medium combination omitted the 6 wells at the lower concentration in favor of 2 additional wells at the higher concentration.

Color development in the BCA assay is initially very rapid and then decelerates with time, but does not reach a true endpoint, which had several practical consequences for our measurements. First, differences of about 2 minutes in the time of BCA reagent addition (approximately the time required to add reagent to all wells of the two plates of a single run) causes significant differences in the color intensity of replicate wells assessed simultaneously, until the rate of color development has slowed considerably (data not shown). Practically, this imposes a minimum incubation time on the assay. Second, we found that incubating the plates at 60°C to accelerate color development (as suggested by the manufacturer) led to enhanced color development in the outer wells of the plate (data not shown), presumably because they warmed more quickly than interior wells. We maintained our reagents and incubations at room temperature, and avoided using the outermost wells of our plates as sample or standard wells. Instead, we filled the

outermost wells with water to provide thermal mass against temperature fluctuations. Finally, there were subtle differences in the rate of color development between the BSA protein standard and our samples of total bacterial protein. For all strains, the rate of color development from bacterial protein was slightly less than that from BSA, so that longer incubations resulted in lower estimates of sample protein concentration (data not shown). This result suggested that maintaining consistent incubation times between runs was essential, and that shorter incubations would be more precise. The incubation time of 90 minutes was chosen to minimize the effect of different rates of color development between samples and standards, while permitting adequate color development from samples of low protein content and reducing the effect of slight differences in the time of reagent addition to sample wells.

Identification of outliers in macromolecular data

For both nucleic acid and protein measurements, at least 2 cell pellets (analytical replicates) were analyzed from each of 3-4 separate culture vessels (experimental replicates) representing each strain-medium combination, except that for strain PX3.15, a single cell pellet from 8 separate culture vessels was analyzed for each medium. Measured macromolecular concentrations from different cultures were normalized to $OD_{420} = 0.1$ to correct for the slight differences in culture density between replicate vessels at harvest. If the estimated DNA, RNA or protein content per OD_{420} unit of the first two analytical replicates from a single culture vessel differed by more than 15%, a 3rd and sometimes a 4th analytical replicate was analyzed as well. Values from an analytical replicate were discarded if they differed by more than 3 studentized residuals from the mean of all analytical replicates analyzed for that strain-medium combination;

remaining values were averaged by experimental replicate to obtain the estimated macromolecular content for a particular strain-medium combination. Calculation of the mean and standard error of compound quantities for each strain-medium combination was done by calculating the quantity separately for each experimental replicate.

Culture density

Determination of the number of cells per culture volume generally conformed to standard protocols for counting bacterial cells stained with 4',6-diamidino-2-phenylindole (DAPI) by epifluorescence microscopy [29,30], with some improvements and modifications as follows:

Cells were fixed in formaldehyde (2% final concentration) at least overnight; small, slowly growing cells in particular sometimes appeared dim after only 2-4 hours of fixation. At least 2 filters were prepared per culture vessel within 48 hours of harvest, stained filters stored at -20°C remained countable for at least a year.

The glass filter funnel was treated with Rain-X (Sopus Products), a commercial silanizing reagent to reduce aqueous adhesion. Assembly of the 25 mm diameter glass filter tower (Millipore) included two Teflon gaskets under the stainless steel frit and two GF/F filters (Whatman) beneath a 0.45 mm pore size mixed cellulose ester backing filter (Millipore) to help create a tight seal of the filter funnel against the 0.2 mm pore size Anodisc working filter (Whatman). All filters except the Anodisc (which is bonded to a polyproylene handling ring) were handled from beneath with a spatula to avoid areas of concentrated cell deposition on the working filter from the compression or rupture of

underlying filters. Such areas were frequently observed following even gentle handling of these filters with filter forceps.

20-100 µl of formaldehyde-fixed culture and 5-10 µl of a 0.5 mg/ml working solution of DAPI were added to 2.5 ml particle-free water in the funnel, followed by another 2.5 ml of water to mix the reagents. After 5-7 minutes staining, vacuum was applied continuously at a maximum pressure differential of 25 kPa until 2 minutes after the last solution had been drawn through the filter sandwich. The Anodisc filter was dried for several minutes in a foil-covered petri dish on a 60°C dry block heater, then applied to a ~28 mm diameter circle spread from 2 drops of type A (low viscosity, low fluorescence) immersion oil (Cargille) on a 75 x 38 mm microscope slide. A 50 x 35 mm coverslip with 1 drop of the same oil was applied to the filter, and the edges of the cover slip sealed with cosmetic fingernail polish.

The entire surface of all filters was scanned under low power magnification to look for regions of concentrated cell deposition and areas where liquid apparently passed through the filter beyond the edge of the bore of the glass filter funnel (as indicated by the presence of background DAPI fluorescence). Since the latter event was fairly common (confirmed by the presence of cells in these areas, albeit at lower density) the entire filterable area of each Anodisc filter was counted as follows: The vernier scales of the microscope stage were used to estimate the diameter of the filterable area of the Anodisc filter in both the X and Y direction on >150 filters; the mean value of 18.8 mm was used to calculate the estimated filter area for all filters. Cells were counted within the ocular

grid for 6 randomly chosen, widely dispersed microscope fields between the edge of the filterable area and the edge of the bore of the filter funnel (diameter 16 mm), and for 17 randomly chosen microscope fields on two perpendicular transects within the bore of the filter funnel. Since the ratio of 6:17 is within 2% of the ratio of filter area outside and inside the bore of the funnel, these counts could simply be summed and divided by the total area covered by the 23 ocular grids to provide an estimate of the cell density on the entire filter corrected for the two areas of the filter with different cell densities.

Cell counts were performed on an Axioscop 2 (Carl Zeiss) equipped with a 100 Watt Hg lamp and a 100x Plan NeoFluar objective. Filters were prepared to have an average of between 20 and 100 cells entirely within or touching two sides of the ocular grid when viewing the central area of the filter, so that at least 400 cells were counted per filter. Estimates of cell number per optical density derived from a single filter that differed by more than 3 studentized residuals from the mean of the 6-10 filters prepared for all cultures of a particular strain-medium combination were excluded; remaining estimates were averaged by culture vessel to provide a single estimate of the cell number per optical density for that strain-medium combination.

Cell size

Cell sizes were determined by image analysis of digital photomicrographs of immobilized bacteria under phase-contrast illumination. Prior to biomass harvest, agarose coated slides were prepared in a dust-free environment by pipetting 1 ml of a tempered, molten 1.6% (w/v) solution of quadruple-washed agarose onto cleaned microscope slides on a horizontal surface. After the agarose was solidified, slides were

dried at 60°C and stored without contact to the dried agarose surface. At harvest, 20-26 ml of culture was pipetted onto the dried agarose and covered immediately with a coverslip. The unfixed, unstained cells were gradually immobilized against the coverslip by the swelling agarose, with optimum contrast occurring just as the cells became immobilized. The slides were scanned to find such areas, which were then photographed in 8-bit grayscale with a Spot 2 cooled CCD camera (Diagnostic Instruments) mounted on an Axioscop 2 (Carl Zeiss) using a 100x Plan-Neofluar objective under phase contrast illumination with a green filter. 5-10 such images containing 20-200 mostly non-overlapping cells were obtained from each culture vessel. For strain PX3.15, slides prepared at the time of harvest were flawed; images obtained from earlier cultures in the same media (4 independent culture replicates per medium) were used for biovolume determination instead. The highest quality images from each strain-medium combination were processed in Photoshop 6.0 (Adobe) to produce binary images (pixels either black or white) containing a total of at least 300 nontouching cells; the binary threshold was chosen interactively so that cell sizes corresponded closely to the apparent cell size on the original grayscale images. Nonetheless, some image-to-image variation in measured cell dimensions was observed, reflecting differences in the quality of focus and contrast in the original digital images. Shape-dependent measurement features available in the CMEIAS software package [31] were used with these binary images to produce biovolume estimates. Actual dimensions of the imaged cells were calibrated from a digital photomicrograph of a stage micrometer visualized under identical conditions.

Statistical comparisons of data

Potential relationships between two quantities across all strains were examined using simple linear regression of the mean values of the quantities for each strain. Because data from the same strain growing in multiple media cannot be considered independent, comparisons across strains use only data from cultures grown in R2BV, since all strains were grown in this medium. Differences between the values of a quantity for a single strain in two media were examined with t tests of values measured from all replicate cultures of a strain.

Phylogenetic analysis

The phylogenetic affiliations of bacterial strains in our collection were established using standard protocols for small subunit (16S) ribosomal RNA gene sequence analysis. Genomic DNA was extracted from pure cultures of the 8 soil isolates and *S. alaskensis* using either the UltraClean Soil DNA kit (MoBio Laboratories) or the Bactozol kit (Molecular Research Center) as directed, and used as a template for amplification of a near full length segment of the small subunit (16S) ribosomal RNA gene via the polymerase chain reaction. Standard bacterial primers (8F paired with either 1492R or 1540R, numbers indicate the starting nucleotide of probe according to *E. coli* numbering) were used with the following parameters for thermocycling: 5 minutes at 95°C; 30-35 cycles of 95°C/65°C/72°C held at each temperature for 30 seconds; 5 min at 72°C. (8F: AGAGTTTGATCCTGGCTCAG, 1492R: GGTTACCTTGTTACGACTT, 1540R: AAGGAGGTGATCCARCCGCA) Reaction mixtures were purified with either Wizard PCR Preps (Promega) or Microcon Centrifugal Filters (Millipore) and amplicons were

sequenced directly using Big Dye terminator chemistry (Applied Biosystems) on an ABI 377 or an ABI 3700 Gene Sequencer (Applied Biosystems) via sequencing primers that provided an average of >3x coverage over the length of the amplicon. Assembled sequences have been deposited in GenBank with accession numbers AY337597-AY337605.

16S gene sequences from our own sequencing efforts and 16S sequences from fully sequenced bacterial and archaeal genomes were imported into an Arb database [32] and aligned automatically to full length aligned sequences obtained from release 8.0 of the Ribosomal Database Project [33]. Initial alignments were optimized manually with reference to secondary structure information available at the Comparative RNA Website, www.rna.icmb.utexas.edu [34]. We chose a set of high quality, near full length 16S sequences including our isolates, close relatives of our isolates, and organisms representing all the major bacterial lineages, along with several diverse archaeal lineages to serve as outgroups. A mask was created to exclude positions of ambiguous alignment, leaving 1250 positions for phylogenetic inference using the maximum likelihood algorithm within Arb. The resulting tree containing 166 bacterial and archaeal sequences has been 'pruned' without altering branching order or branch lengths to obtain the phylogenic tree displayed in Figure 2.1. The taxonomic affiliations shown in Table 2.1 are derived only from 16S sequence data, not phenotypic characterization, and includes both the division and family (or genus, where possible) affiliation of each strain.

Figure 2.1: Phylogeny and *rrn* copy number of experimental bacteria

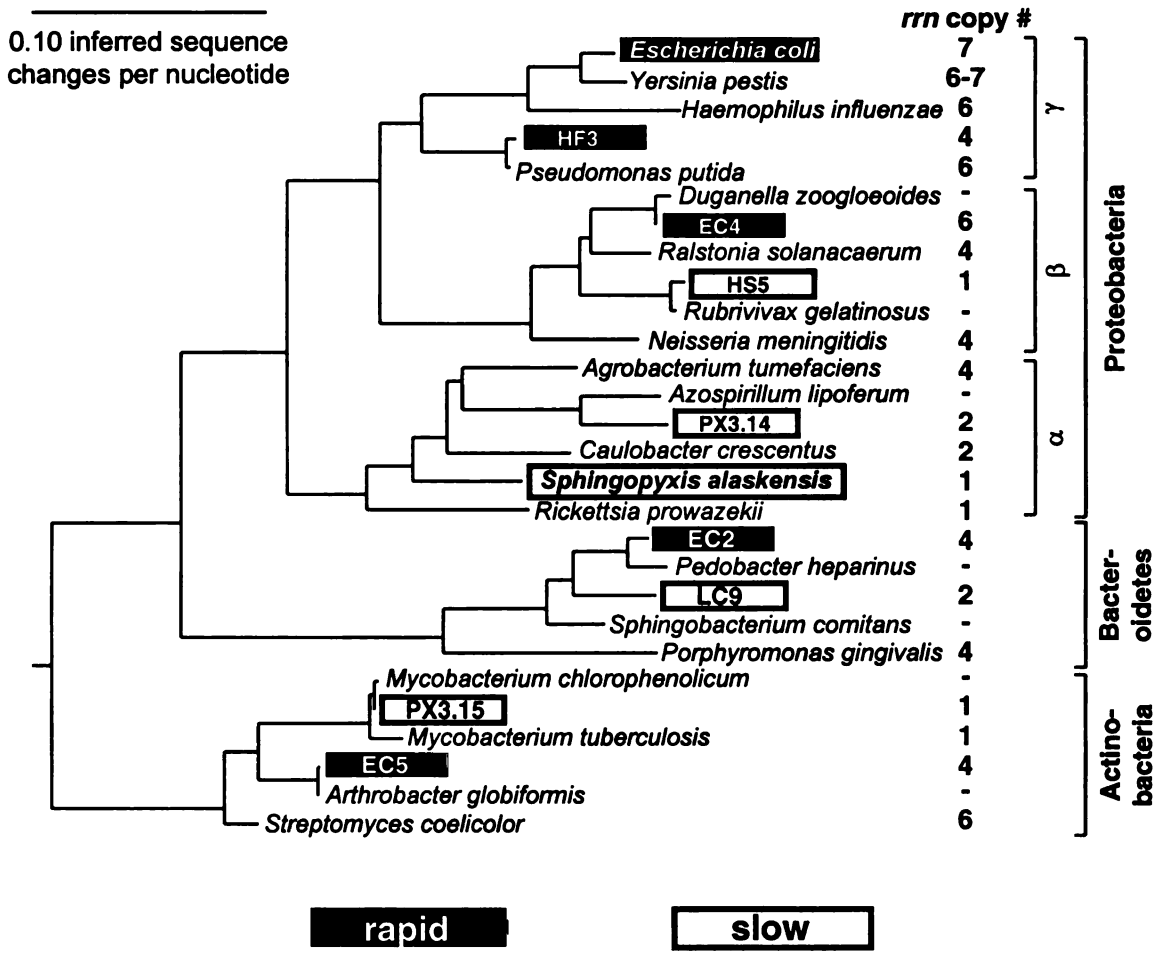


Figure 2.1. Evolutionary relationships among the bacteria examined in this study and selected reference species were examined by a maximum likelihood algorithm using near full-length small subunit ribosomal RNA gene sequences. Rapidly responding bacteria are shown in filled boxes, slowly responding bacteria in open boxes. Sequences from *E. coli* and species shown without boxes were obtained from the Ribosomal Database Project[33]; other sequences were obtained in this study. The number of ribosomal RNA (*rrn*) operons per genome is shown, where known[19,35]; taxonomic affiliations of species are shown at the right.

Table 2.1: Cell and culture traits of experimental bacteria

Strain	Taxonomy[a]	Ref.	*rrn*#[b]	Culture medium	Specific growth rate (hr^{-1})	Cell volume (fl)	Culture density (10^7 cells ml^{-1} at OD_{420}=0.1)
Rapid Responders							
E. coli	γ-Proteobacteria	20	7	R2BV	0.67 ± 0.015	6.4 ± 0.4	2.9 ± 0.1
	Enterobacteriaceae			R2B-GCS	0.65 ± 0.006	4.9 ± 0.2	3.4 ± 0.1
HF3	γ-Proteobacteria	17	4	R2BV	0.710 ± 0.008	5.7 ± 0.5	1.8 ± 0.1
	Pseudomonas			PVY/10	0.645 ± 0.007	3.2 ± 0.2	2.9 ± 0.3
EC4	β-Proteobacteria	19	4	R2BV	0.546 ± 0.005	5.3 ± 0.5	3.6 ± 0.1
	Oxalobacteriaceae			R2B-GCS	0.482 ± 0.003	5.7 ± 0.4	3.4 ± 0.2
EC2	Bacteroidetes	19	6	R2BV	0.431 ± 0.002	2.4 ± 0.1	7.1 ± 0.6
	Sphingobacteriaceae			PVY/10	0.236 ± 0.002	1.5 ± 0.1	12 ± 1.8
EC5	Actinobacteriaceae	19	4	R2BV	0.545 ± 0.004	2.4 ± 0.1	5.6 ± 0.2
	Arthrobacter			PVY/10	0.470 ± 0.008	2.3 ± 0.2	6.2 ± 0.3
Slow Responders							
S. alaskensis	α-Proteobacteria	23,24	1	R2BV	0.217 ± 0.002	1.9 ± 0.2	9.1 ± 0.4
	Sphingomonadaceae			PVY/10	0.130 ± 0.002	0.8 ± 0.1	16.5 ± 0.7
PX3.14	α-Proteobacteria	18	2	R2BV	0.144 ± 0.001	2.8 ± 0.1	3.4 ± 0.1
	Rhodospirillaceae			PVY/10	0.126 ± 0.001	2.9 ± 0.1	2.9 ± 0.1
HS5	β-Proteobacteria	17	1	R2BV	0.081 ± 0.001	3.5 ± 0.2	4.8 ± 0.3
	Comamonadaceae			R2B-GCS	0.067 ± 0.001	2.3 ± 0.1	11.2 ± 0.3
LC9	Bacteroidetes	19	2	R2BV	0.239 ± 0.003	1.3 ± 0.2	12.1 ± 0.4
	Sphingobacteriaceae			R2B-GCS	0.118 ± 0.005	0.9 ± 0.1	9.7 ± 0.6
PX3.15	Actinobacteriaceae	18	2	R2BV	0.106 ± 0.001	1.5 ± 0.1	2.1 ± 0.2
	Mycobacterium			PVY/10	0.040 ± 0.003	1.3 ± 0.1	2.2 ± 0.1

a. Taxonomic placement of soil isolates to family or genus level is tentative, based on small subunit ribosomal RNA gene sequence analysis.

b. Number of copies of the ribosomal RNA operon per genome from references 18 and 19.

Table 2.2: Macromolecular content of experimental bacteria

Strain	Culture medium	DNA		RNA		Protein	
		per cell	per biovolume	per cell	per biovolume	per cell	per biovolume
		(fg cell^{-1})	(fg fl^{-1})	(fg cell^{-1})	(fg fl^{-1})	(fg cell^{-1})	(fg fl^{-1})
Rapid Responders							
E. coli	R2BV	31 ± 0.7	4.9 ± 0.4	162 ± 8	26 ± 2	428 ± 33	69 ± 8
	R2B-GCS	24 ± 0.7	4.9 ± 0.2	123 ± 7	25 ± 1	307 ± 33	63 ± 8
HF3	R2BV	54 ± 2.2	9.7 ± 0.9	249 ± 6	45 ± 3.	582 ± 35	104 ± 8
	PVY/10	16 ± 3.0	4.9 ± 0.9	59 ± 5	19 ± 2	306 ± 9	97 ± 2
EC4	R2BV	29 ± 0.9	5.6 ± 0.5	54 ± 1	10 ± 1	330 ± 26	64 ± 8
	R2B-GCS	29 ± 3.2	5.3 ± 0.9	54 ± 9	10 ± 2	366 ± 42	66 ± 11
EC2	R2BV	13 ± 0.3	5.3 ± 0.1	37 ± 1	15 ± 1	157 ± 8	65 ± 4
	PVY/10	11 ± 0.3	7.3 ± 0.5	18 ± 1	12 ± 1	118 ± 3	81 ± 4
EC5	R2BV	5 ± 0.5	2.0 ± 0.3	23 ± 3	10 ± 1	94 ± 9	39 ± 5
	PVY/10	5 ± 0.6	2.1 ± 0.2	29 ± 2	13 ± 1	130 ± 6	58 ± 4
Slow Responders							
S. alaskensis	R2BV	8 ± 0.5	4.3 ± 0.4	23 ± 1	12 ± 1	93 ± 4	49 ± 4
	PVY/10	5 ± 0.3	6.6 ± 0.7	9 ± 0.7	12 ± 2	51 ± 2	68 ± 8
PX3.14	R2BV	15 ± 1.0	5.3 ± 0.6	38 ± 1	13 ± 1	198 ± 7	71 ± 5
	PVY/10	17 ± 1.0	6.0 ± 0.4	52 ± 1	18 ± 1	254 ± 18	89 ± 7
HS5	R2BV	11 ± 0.8	3.2 ± 0.4	16 ± 1	4.7 ± 0.4	157 ± 5	46 ± 3
	R2B-GCS	12 ± 0.5	5.0 ± 0.3	16 ± 1	6.9 ± 0.2	92 ± 2	39 ± 2
LC9	R2BV	9 ± 0.2	7.4 ± 1.6	22 ± 1	19 ± 4	80 ± 1	66 ± 13
	R2B-GCS	12 ± 0.7	13.2 ± 1.5	20 ± 2	21 ± 1	105 ± 1	111 ± 7
PX3.15	R2BV	17 ± 1.2	11.5 ± 0.8	15 ± 1	10 ± 1	155 ± 7	101 ± 5
	PVY/10	17 ± 0.5	13.1 ± 0.4	16 ± 1	12 ± 1	139 ± 8	108 ± 7

Chapter 2 Results and Discussion

Figure 2.1 shows a phylogenetic tree of the inferred evolutionary relationships among strains in our collection and a selection of familiar bacteria. We were successful in obtaining both rapidly and slowly responding strains from a number of different taxonomic groups, so that any differences between bacteria pursuing different ecological strategies in our analyses will not be explained by evolutionary relatedness among representatives of one strategy to the exclusion of representatives of the alternative strategy. The collection contains representatives of each strategy among the Actinobacteria, the Bacteroidetes and the β-Proteobacteria. However, the soil isolates derived from long-term plating experiments did not include any γ-Proteobacteria that met the criteria for slow responders, nor any α-Proteobacteria that met the criteria for rapid responders. Nonetheless, given the topology of the phylogenetic tree, we can still be certain that the 4 pairs of soil isolates represent 4 evolutionary transitions between the ecological strategies of rapid and slow response.

Taxonomic information and *rrn* copy number are shown for all strains in Table 2.1, as well as the specific growth rate, cell volume, culture density and macromolecular content for each strain in two growth media. We report the macromolecular content in Table 2.2 both per cell and per biovolume. Normalizing to cell or to biovolume, as opposed to OD_{420} unit, facilitates comparison with other published work, although it introduces additional uncertainty by incorporating one or two additional experimental measurements, respectively. (Since the average cell volume is multiplied by the culture density to obtain the denominator for expressing macromolecular measurements per cell

volume, these values are influenced by errors in estimating both cell volume and culture density.) Both in Table 2.2 and in subsequent figures, the uncertainty reported for compound quantities is the standard error of separate estimates of the quantity for each experimental replicate, not an estimate of uncertainty based on propagating the estimated uncertainty of the component measurements.

In general terms, the data conform to expectations, with protein the most abundant and DNA the least abundant of the 3 macromolecules, for 9 of the 10 strains in both growth media. For Strain PX3.15, protein is still the most abundant macromolecule, but the mass of DNA is slightly greater than the mass of RNA in both media tested, although the difference is not statistically significant.

More specifically, our data for *E. coli* are consistent with the data of Bremer and Dennis [36], once a correction is made for differences in growth temperature. It is generally accepted that the composition of the growth medium determines the macromolecular composition of bacterial cells, while changing the incubation temperature influences the growth rate, but not cell size or composition [1]. Calculating from the temperature-dependent growth rate data of Farewell and Neidhardt [37], we estimate the specific growth rates would be accelerated by a factor of 2.33 by shifting from our growth temperature of 25°C to 37°C. Table 2.3 shows that with this correction, our measurements of protein and RNA content per cell are similar to those of Bremer and Dennis [36]; however, our measurements of DNA per cell are 55%-90% higher. We note that despite the conclusion of reference 1 that growth temperature does not influence

Table 2.3: Macromolecular content of *E. coli*

Source of Data	specific growth rate (hr^{-1})	macromolecular content per cell: Protein (fg)	RNA (fg)	DNA (fg)
Bremer & Dennis [36]				
	1.73	448	210	18
	1.39	338	131	14
this study				
	1.57[a]	428	162	31
	1.50 [a]	307	123	24

a. Growth rate adjusted by a factor of 2.33 from actual growth rates at 25°C to obtain estimated growth rate at 37°C for comparison to data from reference 36.

the chemical composition of bacterial biomass, the data of reference 1 show 35%-50% higher DNA content at 25°C than 37°C for 3 of the 5 media tested, and essentially the same DNA content for the 2 remaining media. Given the unknown influence of the different *E. coli* strains examined, as well as the differences in the growth temperatures and the methods used for measuring macromolecular content, we are satisfied with the agreement between our data and the data of reference 36.

The growth rate of a bacterial strain would be expected to influence the time required for it to form a visible colony, and early or late colony appearance was one of the criteria for considering bacteria to be rapid or slow responders. Hence, we are not surprised that growth rates in liquid media differ for bacteria that pursue different ecological strategies. The fastest growth rate observed among the slowly responding strains (0.24 hr^{-1} for strain LC9 in R2BV) is the same as the slowest growth rate observed among the rapidly responding strains (0.24 hr^{-1} for strain EC2 in PVY/10). The richer medium, R2BV, supported faster growth rates than the alternative medium for all strains. However, the difference between the 2 growth rates is greater than 20% for only 1 of the 5 rapidly responding strains, whereas it is greater than 20% for 4 of the 5 slowly responding strains. Because only two growth rates were examined for each strain, and these rates did not differ greatly in half the strains, we are very cautious in interpreting the failure to observe growth rate-related trends within a strain, particularly among the fast-growing strains.

Even for the strains that grew at quite different growth rates in the two media, the ability to distinguish growth rate-related trends is limited by the fact that only two different growth rates are represented in the data. Because the two growth rates for a single strain were obtained by varying the composition of the medium, strictly speaking, we cannot interpret differences in the measurements as a function of growth rate *per se*. Hence, our analysis focuses on trends expected across strains, not within strains, although we do at times use t tests to determine whether the measurements made on a single strain in two media are statistically distinct. Although subsequent figures will show data derived from both media for each strain, the data from a single strain grown in two media cannot be considered independent. Hence, for comparisons across all strains or comparisons between ecological strategies, we will use only the data from each strain grown in R2BV medium. However, it could also be argued that if unknown errors have influenced measurements of a particular strain in R2BV for some quantity, an average of the measurements in both media would be more likely to be representative of the true value of that quantity for the strain. For all the comparisons we report between strains, the use of the average value for the two media instead of the value in R2BV would alter conclusions in only one case, which is noted below.

Microbiologists have long associated increases in growth rate with increases in cell size (reference 1, and references therein), although exceptions have been noted [21,38]. Over all strains, the regression of cell volume on growth rate is significantly positive ($p<0.05$, $R^2=0.54$), determined by the fact that the largest cells occupy a narrow range of rapid growth rates (Figure 2.2). In contrast, the range of growth rates among small cells is

Figure 2.2: Biovolume vs. specific growth rate

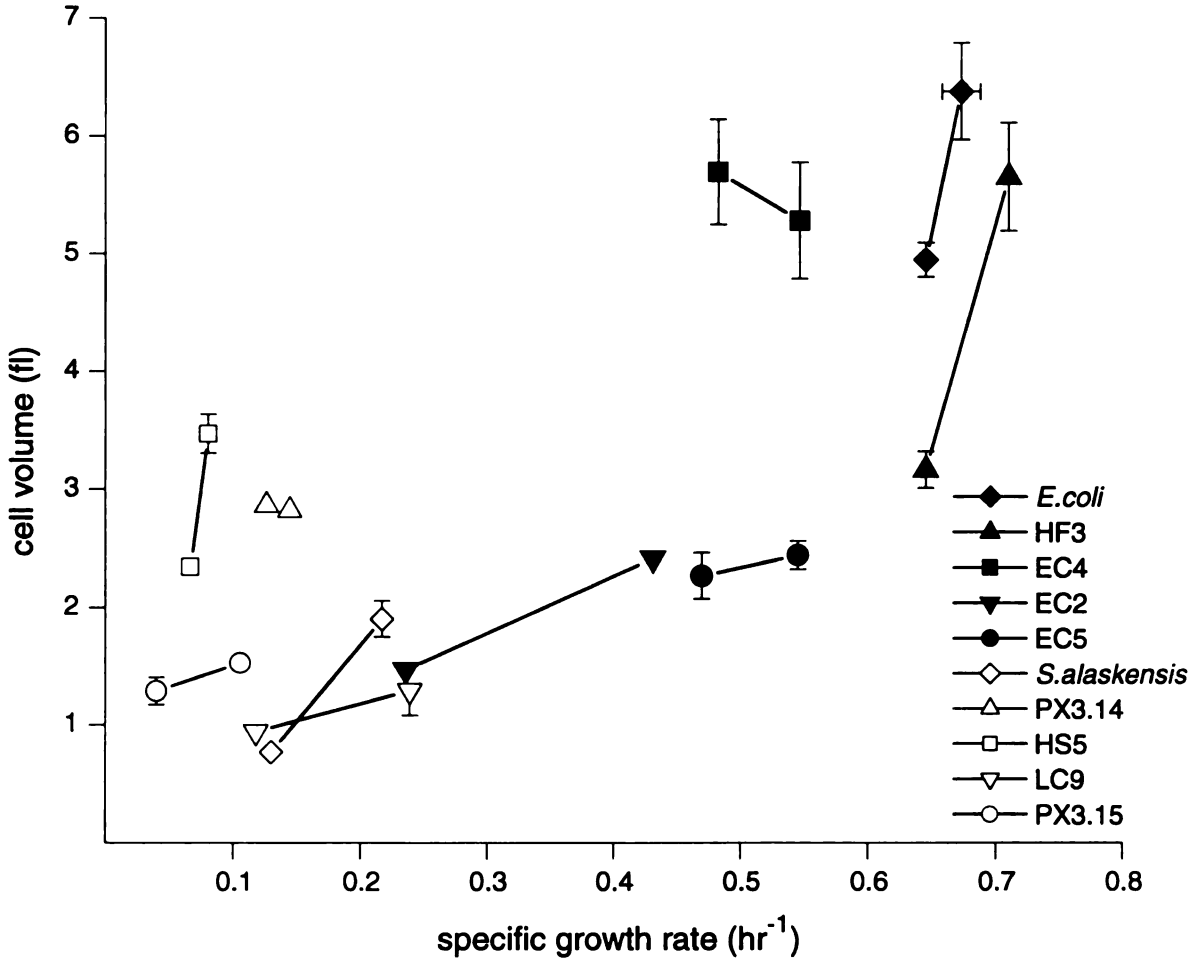


Figure 2.2. Biovolume plotted as a function of specific growth rate for 5 rapidly responding bacteria (solid symbols) and 5 slowly responding bacteria (open symbols). Considering only data from R2BV medium (the medium providing the faster growth rate for each strain), the regression of biovolume on specific growth rate is significant ($p<0.05$, $R^2=0.54$). Data points and error bars represent the means and standard errors of measurements on 3-4 independent replicate cultures per strain medium combination. Error bars smaller than or of comparable size to plot symbols are not shown.

much wider, overlapping at its upper end with the slowest growth rates of the largest cells (compare strains EC4 and EC5). It has been suggested repeatedly that oligotrophic bacteria may benefit from the larger surface:volume ratio conferred by small size [39,40]; our data are consistent with this expectation in that slow responders are smaller than rapid responders (t test, $p < 0.05$, $df = 8$). However, the pattern we observe with respect to cell size may have been influenced by our preference for strains that showed dispersed growth in liquid culture, to permit accurate growth rate measurements via turbidity. Bacteria with filamentous growth forms may have been excluded because of a tendency for flocculent growth, potentially excluding slowly growing strains with large cell volumes.

Considering size variation within each strain, cell volume is significantly larger in the medium supporting faster growth for 5 of the 10 strains, including *E. coli* and HF3, strains that grew in the two media at rates that differed by less than 20%. Strains PX3.15 and LC9 strains grew in the two media at rates that differed by more than a factor of 2, but did *not* differ significantly in cell volume. If we examine the 7 strains with either a significant difference in cell size or at least a 2-fold difference in growth rates, it is clear that the relationship between cell volume and growth rate is dependent on the identity of the strain. Strain PX3.15 shows a (nonsignificant) 20% increase in cell size over a 250% increase in growth rate, whereas HF3 displays an 80% increase in cell size for only a 10% increase in growth rate. Hence, while we do observe a trend across strains of cell volume increasing with growth rate (with the caveat that filamentous bacteria are not represented), our data suggest that inferences of growth rate based on observations of cell

size should be treated cautiously, unless the observations involve a strain for which the relationship between growth rate and cell size has already been established.

The relationship between cell volume and the total macromolecular content per OD_{420} unit (MM/OD, the sum of the DNA, RNA and protein measurements per OD_{420} unit) is shown in Figure 2.3. Although units of optical density are generally assumed to be proportional to dry mass, the proportionality constant may in fact differ between strains and between growth states, particularly if the cells differ in size and shape [41]. The largest components of a bacterial cell that are *not* DNA, RNA and protein, at least in *E. coli,* are components of the cell envelope [42]. Hence, a decline in MM/OD with cell size is expected in cells of approximately uniform shape and similar cell wall structure, since the envelope components would be expected to comprise a larger fraction of cell mass and optical density in smaller cells that have a larger surface:volume ratio. Such a trend is evident in Figure 2.3. However, it is also apparent that while the relationship between MM/OD and biovolume is similar for all cultures containing large cells, it is highly variable among cultures containing small cells. This pattern is not explained by cell size *per se,* but may be explained by cell shape and architecture. All the large cells are regular rods from gram negative lineages, whereas the small cells include a variety of shapes from both gram negative and gram positive lineages (Figure 2.4). The influence of shape on MM/OD may be both direct, in that the shape of a cell may influence its light-scattering properties independently of its composition and total mass, and indirect, since different cell shapes imply different ratios of cell wall material to cytoplasm, and hence different composition.

Figure 2.3: MM/OD_{420} vs. biovolume

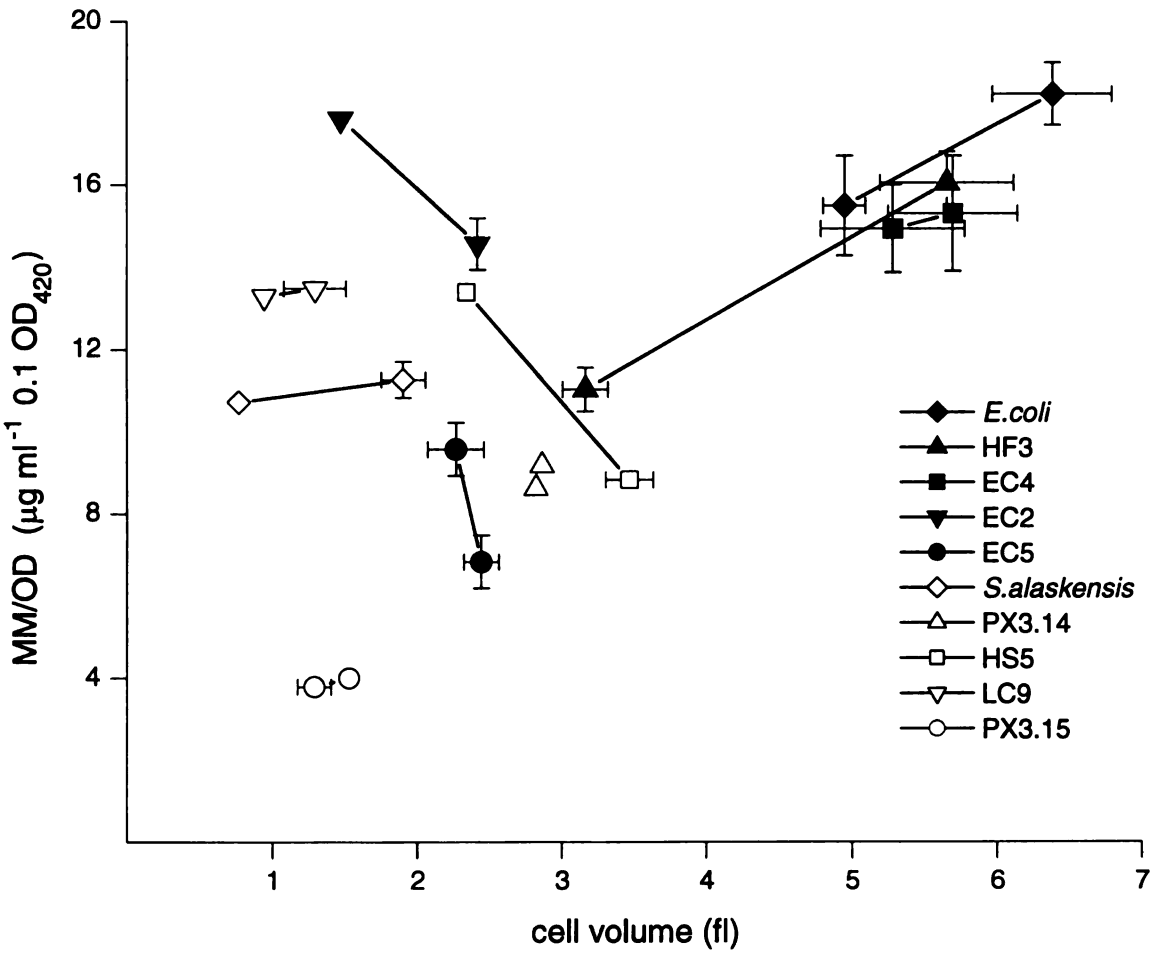


Figure 2.3. MM/OD_{420} (the sum of protein, RNA and DNA per OD_{420} unit) plotted as a function of biovolume for 5 rapidly responding bacteria (solid symbols) and 5 slowly responding bacteria (open symbols). Data points and error bars represent the means and standard errors of measurements on 3-4 replicate cultures per strain-medium combination, except strain PX3.15, represented by 7-8 replicates. Variability in MM/OD_{420} at smaller cell sizes relects greater variation in cell shape and cell wall architecture. Error bars smaller than or of comparable size to plot symbols are not shown.

Figure 2.4: Photomicrographs of experimental strains

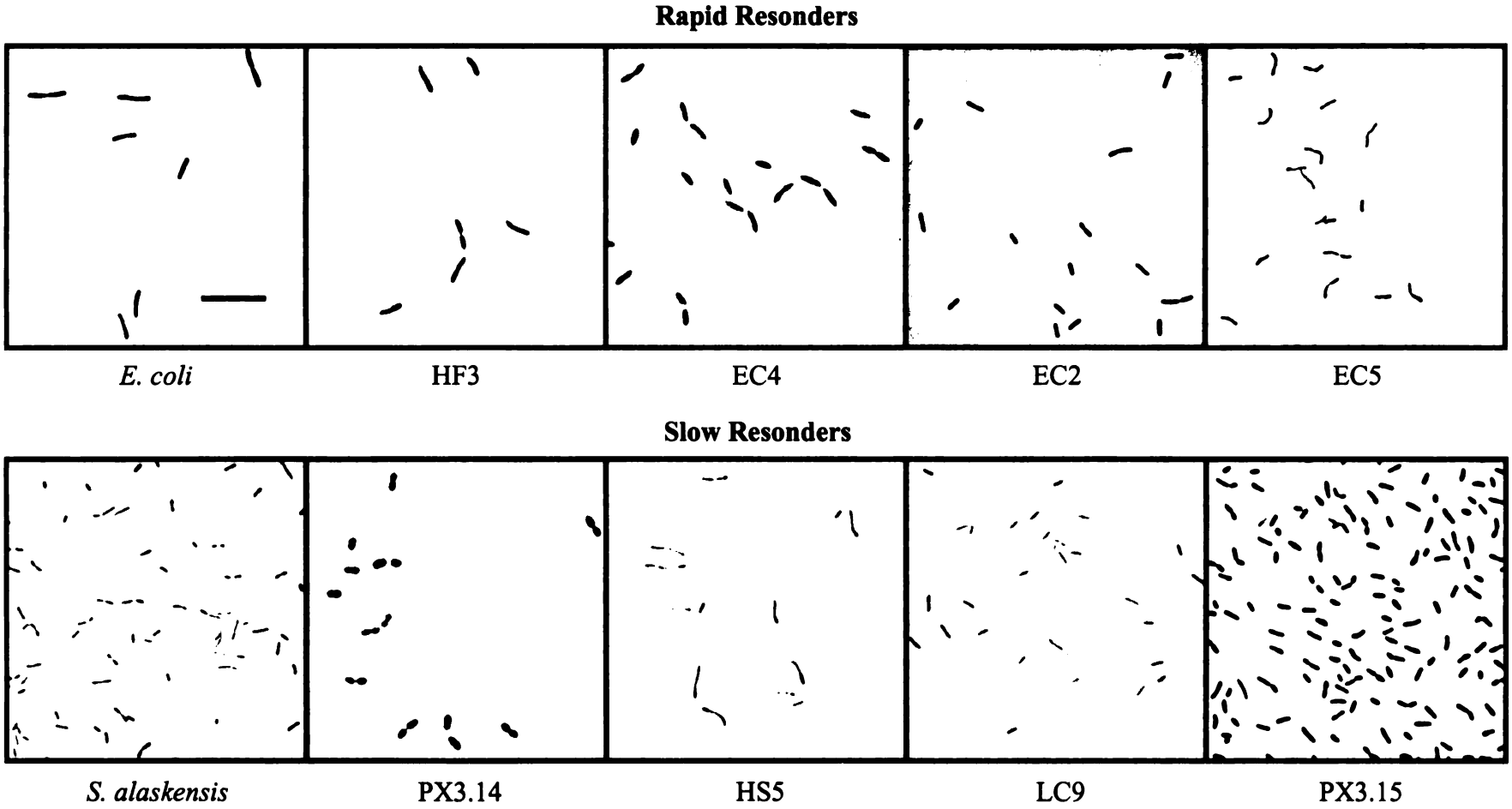


Figure 2.4. Phase contrast micrographs of unfixed, unstained cells immobilized on agarose-coated slides. All photographs taken from cultures in R2BV medium at the time of harvest for macromolecular analysis, except PX3.15 image which was taken from an earlier R2BV culture. The scale bar in the *E. coli* image is 10 micrometers, and applies to all images.

Strain PX3.15 has MM/OD values much lower than those of the other strains. In contrast, the concentration of macromolecules (MM/cell volume) in PX3.15 is in the middle of the range observed in other strains (data not shown), suggesting that the low MM/OD values are a result of high optical density, not low macromolecular content. PX3.15 is a mycobacterial strain, as determined by full-length 16S rRNA sequence analysis; both the unique lipid-rich cell envelope characteristic of *Mycobacterium* and the thicker peptidoglycan layer of gram positive bacteria could contribute to this effect. The one other gram positive organism in our study, strain EC5, has the second lowest values of MM/OD, consistent with there being a larger contribution of the cell wall to optical density in this strain as well. Since the largest cells in our collection are fairly similar in shape and share the gram negative architecture of the Proteobacteria, we do not know whether variability in cell shape and cell wall architecture could influence MM/OD as dramatically among large cells as among small cells.

Strains EC2, EC5 and HS5 display significantly lower MM/OD in the medium which supports larger cells (R2BV in all cases), a change opposite in sign to both the trend observed in most other strains and the trend observed across all strains. The change in biovolume with growth rate is particularly steep in strain HS5 (Figure 2.2), which also shows the greatest decline in MM/OD between the alternative medium and R2BV. These observations could be related if the nature of the carbon substrates in R2BV induces the formation or expansion of an extracellular polysaccharide layer that contributes to cell volume. R2BV contains glucose, casamino acids and soluble starch not found in either alternative medium.

Because of the variability between strains in the extent to which material other than DNA, RNA and protein contribute to optical density, the mass ratios between macromolecules may be the best way to investigate trends in macromolecular content across strains. In *E. coli,* the mass of all 3 macromolecules increases with growth rate (leading to larger cells at faster growth rates), but the increase of RNA mass is the greatest, and the increase of DNA mass is the least for a given increase in growth rate [1,36]. Hence, the RNA:DNA ratio increases with growth rate, and is the macromolecular ratio most sensitive to growth rate. Because of this sensitivity and the conserved roles of these molecules, the RNA:DNA ratio has been suggested as a proxy measurement for the growth rates of bacteria in the environment [43]. A 1995 review of the data available from laboratory-adapted enteric bacteria and several environmental strains found a positive relationship between the RNA:DNA ratio and growth rate across taxa and within taxa, but considerable variability was evident in comparisons of the relationships for individual taxa [44]. In particular, the data showed that low RNA:DNA ratios were obtained over a wide range of slow growth rates, whereas the RNA:DNA ratio was more sensitive to growth rate during faster growth.

The relationship between the RNA:DNA ratio and growth rate derived from our data is shown for all strain-medium combinations in Figure 2.5 ($p<0.01$, $R^2=0.64$, R2BV data). Our data are similar to those in the literature. Specifically, a wide range of growth rates generates RNA:DNA values less than 3, and over this range the identity of the strain is a

better predictor of the RNA:DNA ratio in a culture than the rate at which the culture is growing. We have no satisfactory explanation for the unusual RNA:DNA values of strain EC5. The apparently high value of 6.5 obtained in PVY/10 medium at a growth rate of 0.47 hr^{-1} is near the high end of the range of values reported in the literature at similar growth rates [44], but the steep drop in the ratio with an increase in growth rate is not consistent with other observations.

The protein:DNA ratio in *E. coli* is the macromolecular ratio least sensitive to changes in growth rate, increasing by less than a factor of 2 over a roughly 4-fold increase in growth rate [36]. The plot of the protein:DNA ratio vs. growth rate derived from our data is shown in Figure 2.6; the slight rising trend is not significant. Once again, both the high protein:DNA ratio obtained from strain EC5 growing in PVY/10 and the steep decline in the ratio with an increase in growth rate are unusual. The unusual behavior of EC5 in terms of both the RNA:DNA ratio and the protein:DNA ratio could be explained if the DNA value obtained from biomass grown in PVY/10 medium were low due to some unrecognized experimental artifact.

The protein:RNA ratio in *E. coli* declines nearly 2.5-fold over a 4-fold increase in growth rate [36]. This trend is identified with the need for more ribosomes per protein mass in a fast-growing cell, in order to replicate cell protein more quickly. Figure 2.7 shows the protein:RNA ratio as a function of growth rate for our data. Across all strains, the protein:RNA ratio obtained in R2BV medium shows a significant negative trend with growth rate ($p<0.05$, $R^2=0.48$). However, if the protein:RNA ratio for each strain is

estimated as the average of the values obtained in the two media, the negative trend is not quite significant at the 95% confidence level ($p=0.0575$). The product of growth rate and the protein:RNA ratio is a measure of translational power, as shown in Chapter 1. An exploration of the differences in translational power between the rapidly responding and slowly responding bacteria is the subject of Chapter 3.

Figure 2.5: RNA:DNA ratio vs. specific growth rate

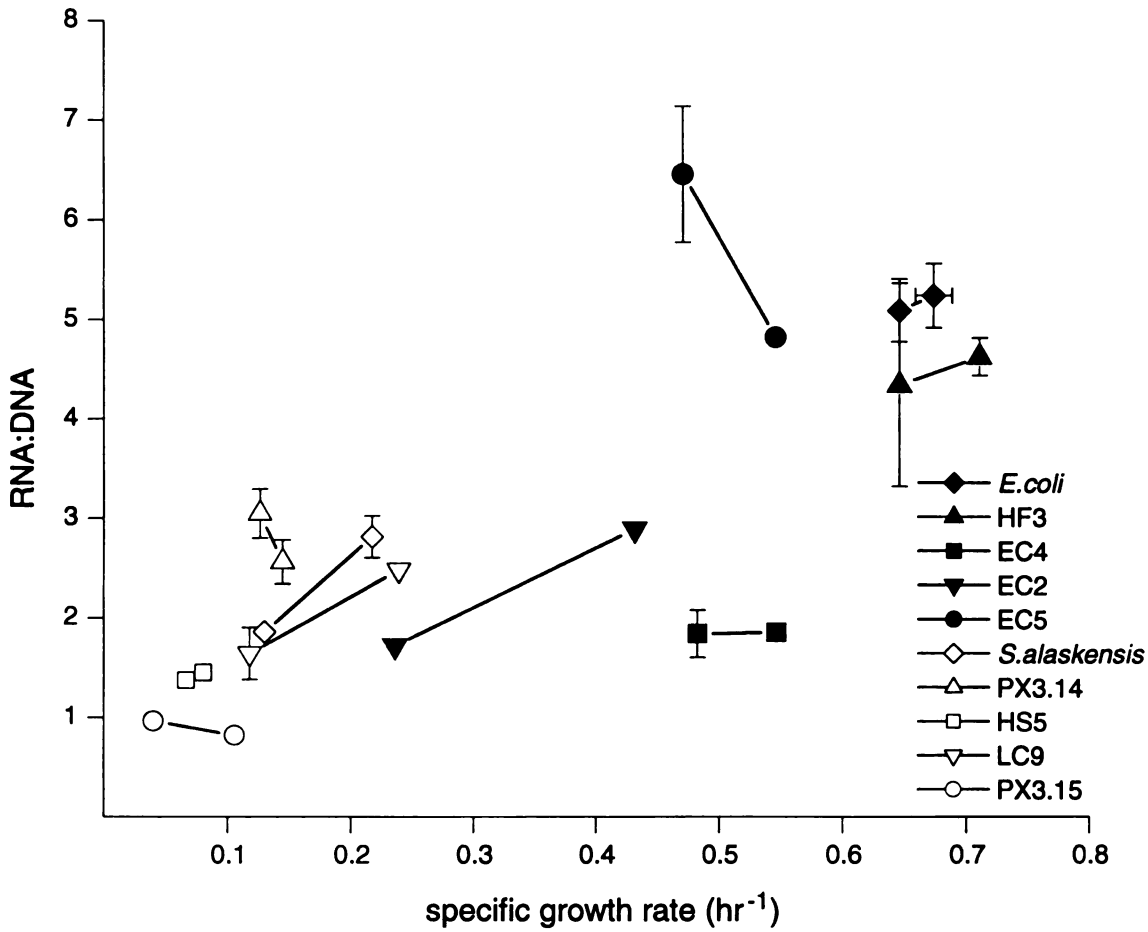


Figure 2.5. RNA:DNA ratio plotted as a function of specific growth rate for 5 rapidly responding bacteria (solid symbols) and 5 slowly responding bacteria (open symbols). Considering only data from R2BV medium (the medium providing the faster growth rate for each strain), the regression of RNA:DNA on specific growth rate is significant ($p<0.01$, $R^2=0.64$). Data points and error bars represent the means and standard errors of measurements on 3-4 replicate cultures per strain-medium combination, except strain PX3.15, represented by 6-7 replicates. Error bars smaller than or of comparable size to plot symbols are not shown.

Figure 2.6: Protein:DNA ratio vs. specific growth rate

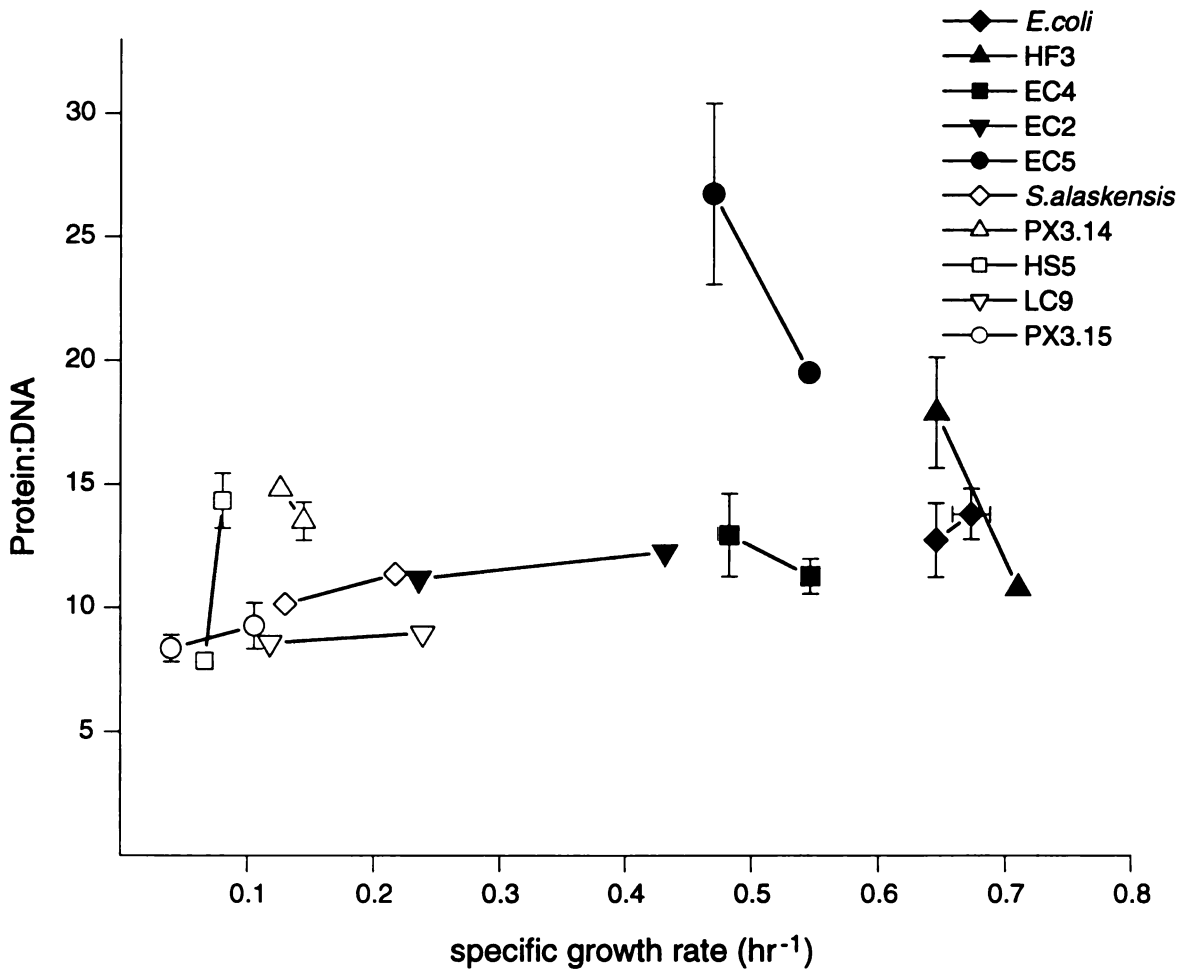


Figure 2.6. Protein:DNA ratio plotted as a function of specific growth rate for 5 rapidly responding bacteria (solid symbols) and 5 slowly responding bacteria (open symbols). Data points and error bars represent the means and standard errors of measurements on 3-4 replicate cultures per strain-medium combination, except strain PX3.15, represented by 7-8 replicates. Error bars smaller than or of comparable size to plot symbols are not shown.

Figure 2.7: Protein:RNA ratio vs. specific growth rate

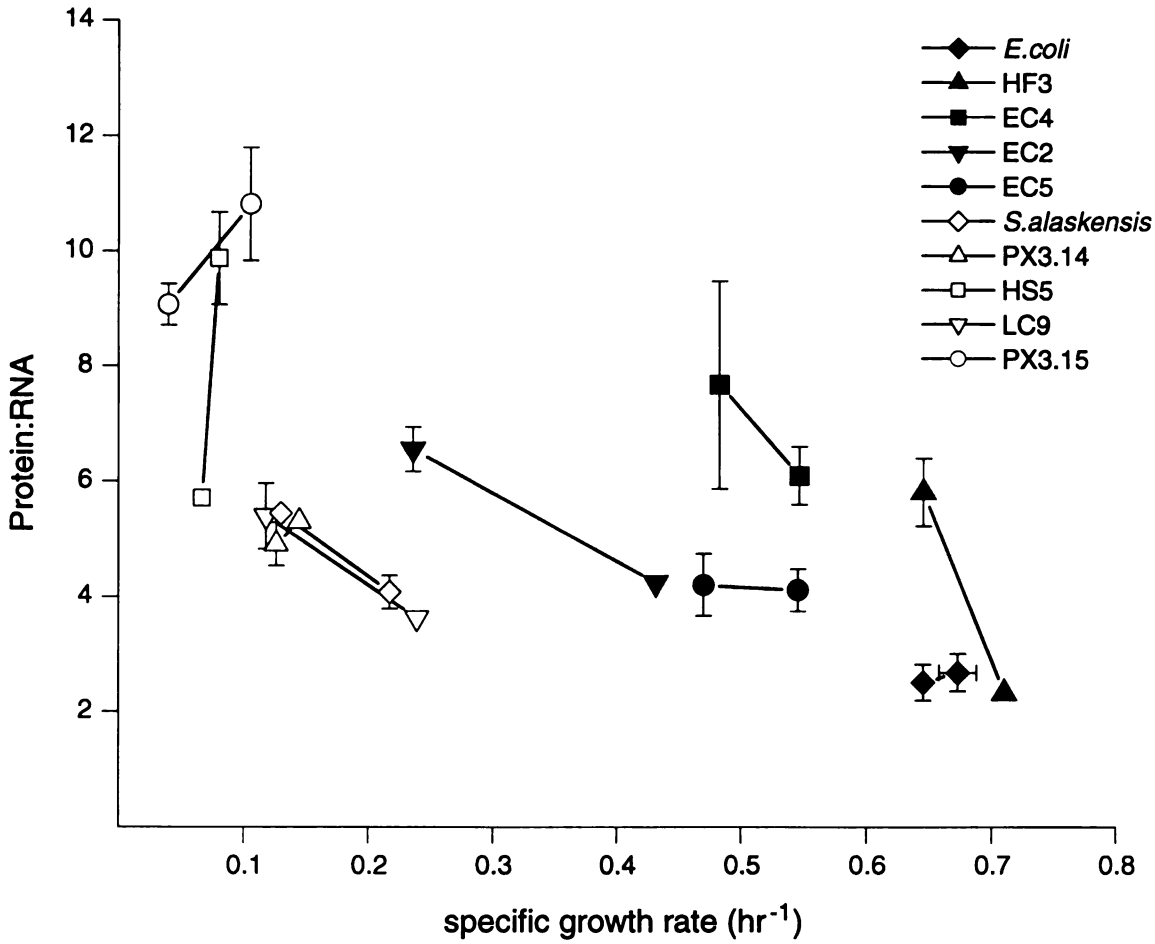


Figure 2.7. Protein:RNA ratio plotted as a function of specific growth rate for 5 rapidly responding bacteria (solid symbols) and 5 slowly responding bacteria (open symbols). Considering only data from R2BV medium (the medium providing the faster growth rate for each strain), the regression of protein:RNA on specific growth rate is significant ($p<0.05$, $R^2=0.48$). Data points and error bars represent the means and standard errors of measurements on 3-4 replicate cultures per strain-medium combination, except strain PX3.15, represented by 7 replicates in each medium. Error bars smaller than or of comparable size to plot symbols are not shown.

Chapter 2 References

1. Schaechter, M., O. Maaloe and N.O. Kjeldgaard, 1958. *Dependency on medium and temperature of cell size and chemical composition during balanced growth of Salmonella typhimurium.* Journal of General Microbiology, **19**(3):592-606.

2. Kjeldgaard, N.O., O. Maaloe and M. Schaechter, 1958. *The transition between different physiological states during balanced growth of Salmonella typhimurium.* Journal of General Microbiology, **19**:607-616.

3. Maaløe, O. and N.O. Kjeldgaard, 1966. *Control of Macromolecular Synthesis.* Microbial and Molecular Biology Series, ed. B.D. Davis. W.A. Benjamin, Inc., New York.

4. Cooper, S., 1993. *The origins and meaning of the Schaechter-Maaloe-Kjeldgaard experiments.* Journal of General Microbiology, **139**:1117-1124.

5. Ingraham, J.L., O. Maaløe and F.C. Neidhardt, 1983. *Growth of the Bacterial Cell.* Sinauer, Sunderland, Massachusetts.

6. Neidhardt, F.C., J.L. Ingraham and M. Schaechter, 1990. *Physiology of the Bacterial Cell.* Sinauer, Sunderland, Massachusetts.

7. Reed, J.L. and B.O. Palsson, 2003. *Thirteen years of building constraint-based in silico models of Escherichia coli.* Journal of Bacteriology, **185**(9):2692-2699.

8. Selinger, D.W., M.A. Wright and G.M. Church, 2003. *On the complete determination of biological systems.* Trends in Biotechnology, **21**(6):251-254.

9. Cho, B.C. and F. Azam, 1990. *Biogeochemical significance of bacterial biomass in the oceans euphotic zone.* Marine Ecology-Progress Series, **63**(2-3):253-259.

10. Ducklow, H.W. and C.A. Carlson, 1992. *Oceanic bacterial production.* Advances in Microbial Ecology, **12**:113-181.

11. Whitman, W.B., D.C. Coleman and W.J. Wiebe, 1998. *Prokaryotes: The unseen majority.* Proceedings of the National Academy of Sciences of the United States of America, **95**(12):6578-6583.

12. Makino, W., J.B. Cotner, R.W. Sterner and J.J. Elser, 2003. *Are bacteria more like plants or animals? Growth rate and resource dependence of bacterial C : N : P stoichiometry.* Functional Ecology, **17**(1):121-130.

13. Chant, J., I. Hui, D. Dejongwong, L. Shimmin and P.P. Dennis, 1986. *The protein synthesizing machinery of the Archaebacterium Halobacterium cutirubrum - molecular characterization.* Systematic and Applied Microbiology, **7**(1):106-114.

14. Pang, H.L. and H.H. Winkler, 1994. *The concentrations of stable RNA and ribosomes in Rickettsia prowazekii.* Molecular Microbiology, **12**(1):115-120.

15. Shahab, N., F. Flett, S.G. Oliver and P.R. Butler, 1996. *Growth rate control of protein and nucleic acid content in Streptomyces coelicolor A3(2) and Escherichia coli B/r.* Microbiology-UK, **142**:1927-1935.

16. Fegatella, F., J. Lim, S. Kjelleberg and R. Cavicchioli, 1998. *Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium Sphingomonas sp. strain RB2256.* Applied and Environmental Microbiology, **64**(11):4433-4438.

17. Gorlach, K., R. Shingaki, H. Morisaki and T. Hattori, 1994. *Construction of eco-collection of paddy field soil bacteria for population analysis.* Journal of General and Applied Microbiology, **40**(6):509-517.

18. Klappenbach, J.A., 2001. *Ecological and evolutionary implications of ribosomal RNA gene copy number in heterotrophic soil bacteria*, Ph.D. Dissertation, Department of Microbiology and Molecular Genetics, Michigan State University. East Lansing.

19. Klappenbach, J.A., J.M. Dunbar and T.M. Schmidt, 2000. *rRNA operon copy number reflects ecological strategies of bacteria.* Applied and Environmental Microbiology, **66**(4):1328-1333.

20. Lenski, R.E., M.R. Rose, S.C. Simpson and S.C. Tadler, 1991. *Long term experimental evolution in Escherichia coli. 1. Adaptation and divergence during 2,000 generations.* American Naturalist, **138**(6):1315-1341.

21. Eguchi, M., T. Nishikawa, K. MacDonald, R. Cavicchioli, J.C. Gottschal and S. Kjelleberg, 1996. *Responses to stress and nutrient availability by the marine ultramicrobacterium Sphingomonas sp. strain RB2256.* Applied and Environmental Microbiology, **62**(4):1287-1294.

22. Ostrowski, M., R. Cavicchioli, M. Blaauw and J.C. Gottschal, 2001. *Specific growth rate plays a critical role in hydrogen peroxide resistance of the marine oligotrophic ultramicrobacterium Sphingomonas alaskensis strain RB2256.* Applied and Environmental Microbiology, **67**(3):1292-1299.

23. Schut, F., E.J. Devries, J.C. Gottschal, B.R. Robertson, W. Harder, R.A. Prins and D.K. Button, 1993. *Isolation of typical marine bacteria by dilution culture - growth, maintenance, and characteristics of isolates under laboratory conditions.* Applied and Environmental Microbiology, **59**(7):2150-2160.

24. Schut, F., J.C. Gottschal and R.A. Prins, 1997. *Isolation and characterisation of the marine ultramicrobacterium Sphingomonas sp. strain RB2256.* FEMS Microbiology Reviews, **20**(3-4):363-369.

25. Singer, V.L., L.J. Jones, S.T. Yue and R.P. Haugland, 1997. *Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation.* Analytical Biochemistry, **249**(2):228-238.

26. Jones, L.J., S.T. Yue, C.Y. Cheung and V.L. Singer, 1998. *RNA quantitation by fluorescence-based solution assay: RiboGreen reagent characterization.* Analytical Biochemistry, **265**(2):368-374.

27. Smith, P.K., R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson and D.C. Klenk, 1985. *Measurement of protein using bicinchoninic acid.* Analytical Biochemistry, **150**(1):76-85.

28. Daniels, L., R.S. Hanson and J.A. Phillips, 1994. *Chemical analysis*; in *Methods for General and Molecular Bacteriology*; P. Gerhardt, R.G.E. Murray, W.A. Wood and N.R. Krieg, eds. ASM Press, Washington, D.C.

29. Kepner, R.L. and J.R. Pratt, 1994. *Use of fluorochromes for direct enumeration of total bacteria in environmental samples - past and present.* Microbiological Reviews, **58**(4):603-615.

30. Kirchman, D., J. Sigda, R. Kapuscinski and R. Mitchell, 1982. *Statistical analysis of the direct count method for enumerating bacteria.* Applied and Environmental Microbiology, **44**(2):376-382.

31. Liu, J., F.B. Dazzo, O. Glagoleva, B. Yu and A.K. Jain, 2001. *CMEIAS: A computer-aided system for the image analysis of bacterial morphotypes in microbial communities.* Microbial Ecology, **41**(3):173-194.

32. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A.W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode and K.-H. Schleifer, 2004. *ARB: a software environment for sequence data.* Nucl. Acids. Res., **32**(4):1363-1371.

33. Cole, J.R., B. Chai, T.L. Marsh, R.J. Farris, Q. Wang, S.A. Kulam, S. Chandra, D.M. McGarrell, T.M. Schmidt, G.M. Garrity and J.M. Tiedje, 2003. *The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy.* Nucleic Acids Research, **31**(1):442-443.

34. Cannone, J., S. Subramanian, M. Schnare, J. Collett, L. D'Souza, Y. Du, B. Feng, N. Lin, L. Madabusi, K. Muller, N. Pande, Z. Shang, N. Yu and R. Gutell, 2002. *The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs.* BMC Bioinformatics, **3**(1):2.

35. Klappenbach, J.A., P.R. Saxman, J.R. Cole and T.M. Schmidt, 2001. *rrndb: the Ribosomal RNA Operon Copy Number Database.* Nucleic Acids Research, **29**(1):181-184.

36. Bremer, H. and P.P. Dennis, 1996. *Modulation of chemical composition and other parameters of the cell by growth rate*; pp. 1553-1569 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

37. Farewell, A. and F.C. Neidhardt, 1998. *Effect of temperature on in vivo protein synthetic capacity in Escherichia coli.* Journal of Bacteriology, **180**(17):4704-4710.

38. Pernthaler, J., T. Posch, K. Simek, J. Vrba, R. Amann and R. Psenner, 1997. *Contrasting bacterial strategies to coexist with a flagellate predator in an experimental microbial assemblage.* Applied and Environmental Microbiology, **63**(2):596-601.

39. Hirsch, P., M. Bernhard, S.S. Cohen, J.C. Ensign, H.W. Jannasch, A.L. Koch, K.C. Marshall, A. Matin, J.S. Poindexter, S.C. Rittenberg, D.C. Smith and H. Veldkamp, 1979. *Life under conditions of low nutrient concentrations group report*; pp. 357-372 in *Strategies of Microbial Life in Extreme Environments: report of the Dahlem Workshop*; M. Shilo, ed. Verlag Chemie, Weinhein.

40. Button, D.K. and B. Robertson, 2000. *Effect of nutrient kinetics and cytoarchitecture on bacterioplankter size.* Limnology and Oceanography, **45**(2):499-505.

41. Koch, A.L., 1994. *Growth measurement*; in *Methods for General and Molecular Bacteriology*; P. Gerhardt, R.G.E. Murray, W.A. Wood and N.R. Krieg, eds. ASM Press, Washington, D.C.

42. Neidhardt, F.C. and H.E. Umbarger, 1996. *Chemical composition of Escherichia coli*; pp. 13-16 in *Escherichia coli and Salmonella Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

43. Dortch, Q., T.L. Roberts, J.R. Clayton, Jr. and S.I. Ahmed, 1983. *RNA/DNA ratios and DNA concentrations as indicators of growth rate and biomass in planktonic marine organisms.* Marine Ecology-Progress Series, **13**:61-71.

44. Kemp, P.F., 1995. *Can we estimate bacterial growth rates from ribosomal RNA content?*; pp. 279-302 in *Molecular Ecology of Aquatic Microbes*; I. Joint, ed. Springer-Verlag, Berlin.

Chapter 3. Translational power and the strength of translational selection vary with ecological strategy in bacteria

Chapter 3 Abstract

All forms of life synthesize protein using a similar process, carried out by similar molecules, but little is known about how translational performance may differ between organisms. Phylogenetically diverse bacteria with contrasting ecological characteristics were used to assess whether variation in translational performance is correlated with differences in the selective regime experienced by the organism. Bacteria capable of rapid growth in response to abundant external resources were assumed to pursue a strategy of exploiting large fluctuations of resource availability. Bacteria incapable of a rapid growth response were assumed to pursue alternative strategies such as exploiting conditions of low resource availability or maximizing starvation survival. Two lines of evidence support a difference in translational performance between rapidly responding bacteria and slowly responding bacteria. First, translational power, the protein output per biomass invested in the translational apparatus, was found to differ between rapidly responding bacteria and slowly responding bacteria. Second, the strength of selection for translational power experienced by a bacterial species was greater for strains with a higher number of ribosomal RNA operons per genome, which measures the extent to which a strain is adapted for a rapid response strategy. These results are consistent with the existence of an evolutionary trade-off between translational power and translational yield, the amount of protein synthesized per energy consumed.

Chapter 3 Introduction

The activity of a growing bacterial cell is dominated by protein synthesis. The polymerization of amino acids accounts for 2/3 of total energy expenditure in *Escherichia coli* growing on glucose [1], and over half the cell's dry mass can be devoted to the protein synthesis system [2]. Hence, even small improvements in the operation of the protein synthesis system – decreasing the energy cost of translation, or increasing the rate of protein production from the biomass invested in the translational apparatus – may have significant effects on bacterial fitness. However, despite recent progress in understanding ribosomal structure [3] and the sequence of events involved in translation [4], interspecies variation in translational performance has received little experimental attention. This lack of attention may result from an assumption that since the translational apparatus is highly conserved, the translational apparatus performs similarly in all organisms.

Since protein synthesis is both essential and ancient, it might be argued that there has been ample opportunity for natural selection to optimize translational performance. However, analysis of macromolecular composition data from the scientific literature indicates that microbes differ in translational power, a measure that reflects both the rate of translating ribosomes and the fraction of ribosomes that are active (Chapter 1). Furthermore, the ribosomes of a number of natural *E. coli* strains show a wide range of translation rates and missense error rates, and these traits respond rapidly to the selection for faster growth experienced by the strains in the laboratory environment [5,6]. In *E. coli,* the evolution of the protein synthesis system is influenced by profound connections between the rate of translation and the frequency of translational errors [7-13]. Hence, if

the relative importance of rapid translation and error-free translation depends on the environment, natural selection may have led to differences in the performance of the translational apparatus between organisms.

Errors of translation include both inaccurate selection of aminoacyl-tRNAs (missense errors) and premature translational termination (processivity errors). Processivity errors are particularly costly, because energy is spent and ribosomes are occupied in the production of truncated, nonfunctional polypeptides that are subsequently degraded [8]. In contrast, most missense errors result in a protein that retains most of the activity of the canonical version [8]. Nonetheless, translational processivity errors are not rare in laboratory-adapted strains of *E. coli,* the only organism for which data are available. An early experiment demonstrated that the products of processivity errors represent about 25% by mass of all protein synthesized from the *lacZ* gene [14]; other experiments with *lacZ* have resulted in similar estimates [10,11]. The majority of processivity errors in *E. coli* can be attributed to dropoff, the loss of peptidyl-tRNA from the ribosome [8,15]. The average likelihood of a processivity error is about 2.5×10^{-4} per codon; at this rate about 10% of the ribosomes that initiate translation of an average-length gene would fail to synthesize a full-length protein. Processivity errors increase the energy expended for protein synthesis. The energy of 4 phosphoanhydride bonds is required to form each peptide bond – 2 for activating an amino acid for attachment to a tRNA molecule, and 2 for adding an amino acid to the growing polypeptide during translation. While the amino acids in a truncated polypeptide are recycled via proteolysis, the energy spent to form the peptide bonds is lost. We define translational yield as the mass of (functional) protein

produced per energy consumed. Processivity errors, by wasting energy in the synthesis of nonfunctional, truncated polypeptides, decrease translational yield.

On the basis of a postulated evolutionary tradeoff between translational power and translational yield, we predicted that translational power would be higher in bacteria adapted to exploit large fluctuations of resource availability through rapid growth, and lower in bacteria adapted to exploit small fluctuations of resource availability, and to maximize starvation survival. The bacterial soil isolates we have identified as rapid responders (Chapters 1 & 2) clearly would be expected to have high translational power. We have fewer reasons to identify the slow responders with high yield strategies, because these strains have been classified by the absence of a rapid response, rather than by positive identification of strategy-revealing traits such as low resource thresholds for growth or viability. Nonetheless, because low *rrn* copy number limits the capacity of these strains to make a rapid growth response, we are confident that these strains are not simply cryptic rapid responders, but instead are pursuing a different ecological strategy. On the basis of the proposed benefit of low *rrn* copy number in perpetually resource-limited habitats and the observed tradeoff between rapid growth and starvation survival in *E. coli*, we predicted that the slow responders would have low translational power.

Codon usage analysis was used to develop a second line of evidence supporting a relationship between translational performance and the ecological strategies of bacteria. Codon bias refers to the use of some codons significantly more frequently than synonymous alternatives; the preferred codons are translated more quickly, and with

fewer errors [16,17]. Selection for translational power has been invoked to explain a common pattern in codon bias among the genes of an organism, that the degree of bias is correlated with the expression level of the gene. Hence, codon usage analysis can be used to infer the strength of selection for translational power experienced by an organism. Because a high number of *rrn* operons per genome is adaptive for a rapid response strategy [18,19], we predicted that the strength of selection for translational power would be correlated with *rrn* copy number.

Chapter 3 Materials and Methods

Translational power

Bacterial strains and laboratory techniques used to obtain growth rate and macromolecular composition data are reported in Chapter 2. These data are the basis of the comparisons of translational power described in this chapter.

Codon usage analysis

The strength of translational selection was measured by comparing codon usage in a set of highly expressed (HE) genes to codon usage in the genome as a whole, using 76 bacteria for which the complete genome sequence is available. A complete list of the genomes analyzed and a phylogenetic tree depicting a hypothesis of their evolutionary history is presented in the Appendix. The 8 genes comprising the HE set were chosen because they are known to be highly expressed and to have high codon bias in *E. coli* [20]. The HE set includes genes for 5 ribosomal proteins *rpsA*/S1, *rpsB*/S2, *rpsI*/S9, *rplA*/L1 and *rplM*/L13, and 3 elongation factors *tuf*/EF-Tu, *tsf*/EF-Ts and *fus*/EF-G. We chose not to include highly expressed *E. coli* genes in other functional categories (e.g., outer membrane protein *ompA* or heat shock protein *groEL*) in the HE set because we were more confident that homologs of translation apparatus genes would be identifiable and would have retained high levels of expression in all organisms. The number of ribosomal protein and elongation factor genes in the HE set was chosen so that a sufficient number of codons were included in each subcategory for reliable independent estimations of codon frequencies. All the results based on the complete HE set are essentially

unchanged if the analysis is done using either the ribosomal protein gene subset or the elongation factor gene subset.

Homologs of the HE set were identified in each bacterial genome via the 'Name Search' function of the Comprehensive Microbial Resource (CMR) [21], a website hosted by The Institute for Genomic Research (www.tigr.org). All homologs of a gene in a single genome were included in our analysis (i.e., multiple copies of elongation factor genes were included), as long as protein parameters (length, pI, etc.) were similar between multiple homologs in a genome. The number of instances of each codon in each gene of the HE set were downloaded from the CMR website and summed to calculate the observed relative codon frequencies in the HE set of each genome. Codon data tabulated for all predicted genes in each genome were downloaded directly from the CMR and used to calculated the expected relative codon frequencies for comparison with the HE set.. Procedures to handle unreliable relative codon frequency estimates when amino acids were observed fewer than 5 times were as suggested by Novembre [22], although the number of codons included in each dataset was sufficient that such corrections were rare.

The comparison between codon use in the HE set and in the genome as a whole was done via Novembre's N'c [22], a generalization of the effective number of codons, Nc, an index of codon bias proposed by Wright [23]. Nc makes an implicit comparison of observed relative codon frequencies to uniform usage of synonymous codons, reported as an effective number of codons between 61, the total number of sense codons, and 20, the minimum number of codons representing all amino acids. A value of 61 indicates no bias

(all synonymous alternative codons used at the same frequency), whereas 20 indicates the maximum possible bias (only a single codon used for each amino acid).

Instead of making a comparison only to the special case of uniform synonymous codon use, Novembre's N'c measures the deviation of the relative codon frequencies observed in a particular set of genes relative to any explicitly specified expected codon frequency distribution, e.g., the frequencies with which codons are used in the genome as a whole. It is based on a χ^2 comparison of observed and expected frequencies over the codons of an amino acid. Low χ^2 values (meaning observed and expected frequencies are similar) are then transformed to a value equal or close to the total number of codons available for that amino acid, whereas high χ^2 values (meaning observed and expected frequencies differ greatly) are transformed to a value approaching 1. The sum of these values over all amino acids is N'c for a sequence, which ranges from 61 to 20, as does Wright's Nc. However, the scale is not interpreted precisely as an effective number of codons. Instead, the value of N'c is maximized if the observed sequence makes complete use all the degrees of coding freedom represented by the expected codon frequencies, even if the expected frequencies themselves reflect preferential use of only a subset of all sense codons. If the expected codon frequency distribution is completely unbiased (i.e., uniform usage of synonymous codons), N'c reduces to Nc. Neither of these measures is biased by differences in amino acid composition or gene length, and the increase in variability of both indices when estimated over short sequences is relatively mild compared to other indices of codon bias [22,24]. Our data are reported using a transformation of Novembre's N'c: $\Delta N'c = 61 - N'c$. We have made this transformation

to obtain an index which equals zero when there is no difference in codon use between the HE set and the genome as a whole, and increases in magnitude as the codon use of the HE set is increasingly biased relative to codon use over the genome as a whole. Hence, $\Delta N'_c$ is positively correlated with the strength of translational selection.

In addition to calculating $\Delta N'_c$, which depends on the disparity of codon bias between the HE set and the genome, we wanted to assess relationship of *rrn* copy number to the codon bias of each component separately. We measured the codon bias of the HE set and the genome separately by calculating Wright's N_c for each.

The total number of identified tRNA genes in each sequenced genome were obtained from the CMR website [21], with pseudogenes and selenocysteine tRNA genes excluded.

Comparative analyses via log likelihood ratios

We make comparisons of *rrn* copy number to 4 different traits across many bacterial genomes. The four traits are $\Delta N'_c$ for a comparison of the HE set to the genome, Wright's N_c for both the HE set and the genome, and tRNA gene copy number. To test the significance of an association between *rrn* copy number and any of these traits, we used a maximum likelihood log ratio method described by Pagel [25] and implemented in the Continuous software program available at www.ams.rdg.ac.uk/zoology/pagel/. The method calculates the model of evolutionary change of measured traits over a specified phylogeny that is most likely to have resulted in the observed distribution of trait values in extant taxa. The likelihood of the best (most likely) evolutionary model representing

an experimental hypothesis (e.g., correlation exists between 2 trait values) is compared to the likelihood of the best evolutionary model representing a null hypothesis (e.g., trait values are uncorrelated) to calculate the probability (p value) that the null hypothesis explains the observed distribution of traits as well as the experimental hypothesis.

The explicit phylogenetic information used in the log ratio tests was based on a phylogeny inferred by a maximum likelihood algorithm [26] from 16S rRNA gene sequences (see Appendix). To ensure that poorly resolved details of the bacterial phylogeny did not influence the results, each comparative analysis using the Continuous software was repeated with 2 different descriptions of bacterial phylogeny: a tree inferred by maximum likelihood analysis of 16S rRNA sequence data with unmodified internal branch lengths, and a tree derived from the previous tree in which short internal branches were collapsed.

Each test for a correlation between two traits using Continuous software followed the same sequence. First, maximum likelihood estimations of the scaling parameters κ, δ, and λ were made with the two traits to be tested included in the model. These scaling parameters describe how the evolutionary model for the traits fits to the tree topology; e.g., the extent to which trait evolution is gradual or punctuated, or the extent to which it has accelerated or decelerated over time (see reference 25 and software documentation for more details). The scaling parameter values specific to each model were used during subsequent hypothesis tests. Second, separate models were estimated assuming either nondirectional (random walk) or directional models of trait evolution, and compared with

a log likelihood ratio test; in no case could a nondirectional model be rejected at the 95% confidence level. Third, the likelihood of the nondirectional model was calculated with the strength of the correlation between the two traits either constrained to be zero (null hypothesis), or estimated by maximum likelihood (experimental hypothesis). The reported significance (p value) for each correlation between two traits represents the log likelihood ratio these two competing evolutionary models. In other words, it represents the likelihood that evolution over the specified phylogeny would have generated the observed trait distribution by chance, if the two traits were unrelated. If the correlation is significant, the reported coefficient of determination (R^2 value) is the square of the maximum likelihood estimate of the correlation coefficient between the two traits.

Chapter 3 Results

Obtaining bacteria with contrasting ecological strategies

Bacteria were tentatively considered to be rapid responders if they were isolated from colonies first visible 1-2 days after plating a dilute soil suspension on solid media; isolates from colonies that were first visible after 7 or more days of incubation were tentatively considered to be slow responders. Several lines of evidence support these designations. First, the trait of early or late colony appearance is displayed consistently during subsequent growth of the strains as purified isolates on various media. If delayed colony appearance during the initial plating experiment were due to variation in the physiological state of the individual cells when first plated, as opposed to heritable variation between strains, such consistency would not be maintained. The consistency of colony appearance time also indicates that late colony formation is not explained by growth inhibition from co-occurring strains or particular components of the medium. Second, the identities of the early and late appearing strains are distinct. Although a single phylogenetic group may contain both rapid and slow responders, the same strain, as determined by small subunit ribosomal RNA gene sequence, was never isolated from both early and late appearing colonies. Finally, the number of ribosomal RNA (*rrn*) operons per genome, which ranges from 1-15 in bacteria [27], is distinct between the isolates from early and late appearing colonies. Early colony forming strains from soil have an average of 5.5 *rrn* operons per genome, significantly more than the average of 1.4 among the late colony forming strains [18]. High *rrn* copy number is thought to be an adaptation for conditions of fluctuating resource availability, by permitting both fast

growth and rapid acceleration from non-growth to maximal growth rates [18,28]. Low *rrn* copy number, on the other hand, may be an adaptation for growth in stable conditions of low nutrient availability [19]. Hence, we considered an *rrn* copy number of 4 or higher as confirmation of a rapid response strategy, whereas a copy number of 2 or lower confirmed that a strain pursued an alternative strategy.

To ensure that the comparison of ecological strategies was not confounded by shared evolutionary history, we sought related pairs of soil isolates, including both a rapid responder and a slow responder, within phylogenetically diverse taxa. Rapid responder-slow responder pairs were found among the Actinobacteria, the Bacteroidetes, and the β-Proteobacteria. A fourth pair included a rapidly responding isolate from the γ-Proteobacteria and a slowly responding isolate from the α-Proteobacteria. In addition to these 8 recent soil isolates, we tested a well-characterized exemplar of each strategy, *E. coli* (γ-Proteobacteria) representing rapid response, and the ultraoligotrophic marine bacterium *Sphingopyxis alaskensis* RB2256 (α-Proteobacteria, originally *Sphingomonas alaskensis*) representing slow response [29,30]. The phylogeny and *rrn* copy number of these strains are depicted in Figure 2.1.

Comparison of translational power

We compared the translational power of the 10 bacterial strains in our collection using the metric introduced by Kjeldgaard and Kurland [31] (originally called 'ribosome efficiency', see Chapter 1 for our rationale for the new terminology):

$$\text{translational power} = \frac{\mu P}{R},$$

where μ is the specific growth rate and P and R are the masses of protein and RNA in a cell or culture (Chapter 1). Using the macromolecular data reported in Chapter 2, we find that the ranges of translational power for organisms with contrasting ecological strategies do not overlap (Figure 3.1). The highest translational power among the slowly responding bacteria is not as high as the lowest translational power among the rapidly responding bacteria. Considering data from only the R2BV medium, the average translational power among the rapidly responding bacteria was 2.2 ± 0.31 hr^{-1}, significantly higher than the average translational power among the slowly responding bacteria of 0.9 ± 0.07 hr^{-1}, as predicted (mean ± se, $p < 0.002$ by one-tailed t-test, df=8). The conclusion is unchanged if the translational power of each strain is estimated as the average value over both media, or if the data from both media are included as separate data points. Our observation that the average translational power of the slowly responding bacteria in R2BV medium is 41% of the average translational power of the rapidly responding bacteria in the same medium falls at one end of the range of comparisons based on data in the literature between various slowly-growing strains and *E. coli,* which ranged from <17% – 42% (Chapter 1). Note, however, that making this comparison between strains in a consistent environment does not control for different levels of investment in the translational apparatus between strains.

A more detailed presentation of the data is made in Figure 3.2, which plots translational output against RNA content for each strain in both media. On this plot, translational

Figure 3.1: Translational power of rapidly responding and slowly responding bacteria

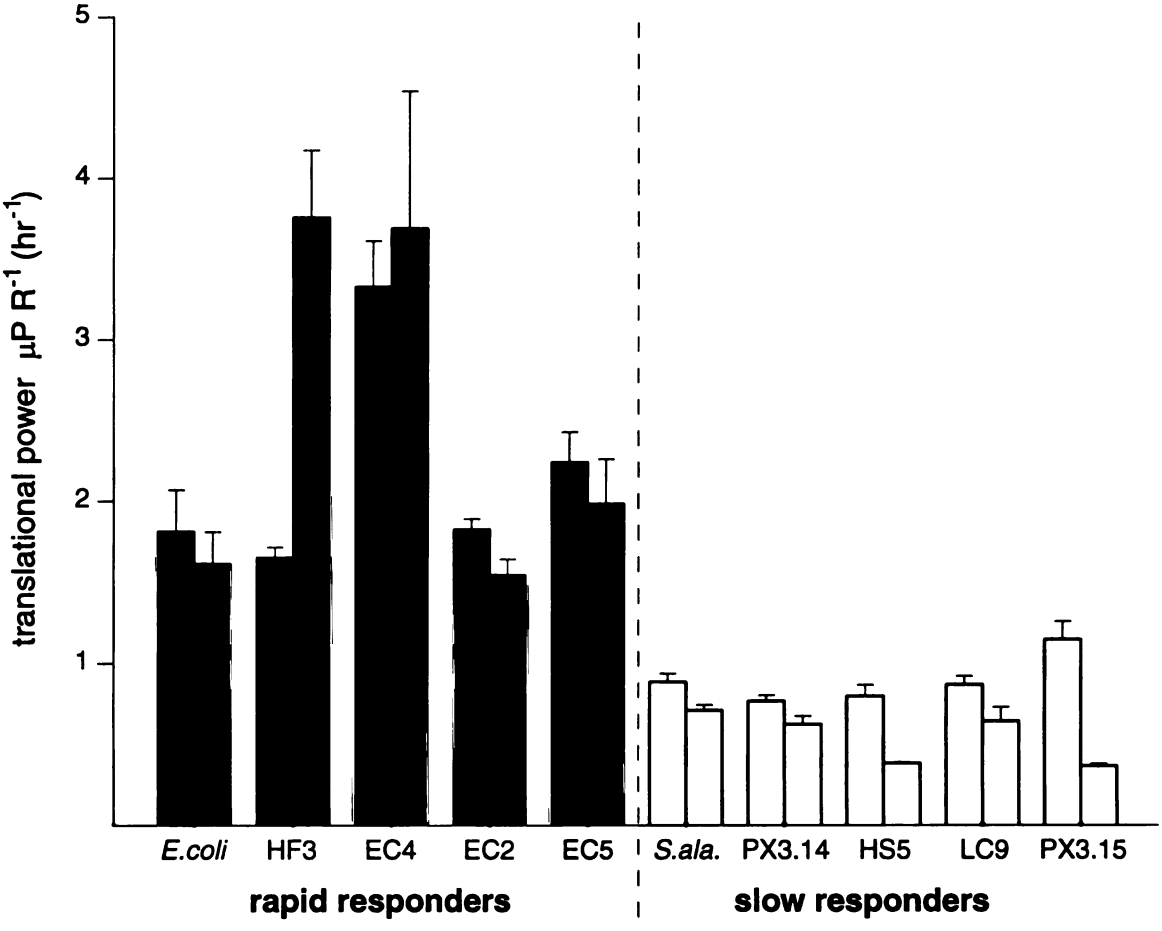


Figure 3.1. Translational power of rapidly responding bacteria (solid bars) and slowly responding bacteria (open bars) grown in two different media, based on data presented in Chapter 2. The bar on the left represents data from growth in R2BV medium for each strain; the bar on the right represents either R2B-GCS (*E. coli*, EC4, HS5, LC9) or PVY/10 (all remaining strains). Media are described in Chapter 2. The height of each bar represents the average of measurements on at least 3 independent cultures of a strain-medium combination; error bars are standard errors of the mean. Based only on data generated in R2BV, translational power is significantly higher among rapid responders than slow responders ($p<0.002$ by one-tailed t test, df=8).

power is the ratio of the ordinate to the abscissa, so any straight line passing through the origin represents a contour of equal translational power. Figure 3.2 shows one such contour that clearly separates all data points derived from rapidly responding bacteria from all data points derived from slowly responding bacteria. Making a comparison of the data in the region where the levels of RNA investment of microbes of both strategies overlap (boxed region of Figure 3.2, between about 1.25 and 2.75 mg ml^{-1} 0.1 OD_{420}^{-1} RNA), the average translational output of rapidly responding bacteria is about 3.6-fold higher than the average translational output of the slowly responding bacteria. This difference is in the middle of the range of comparisons of translational power between *E. coli* and slowly growing microbes presented in Chapter 1.

Correlation of *rrn* copy number with the strength of translational selection

Translational selection refers to the existence of a translation-related benefit, such as faster or less error-prone translation, obtained by using certain codons instead of synonymous alternatives [17]. Because the magnitude of any such benefit varies with expression level, a correlation among the genes in an organism between expression level and the degree of codon bias is evidence of translational selection. In order to compare the strength of translational selection between strains, however, it is necessary to distinguish codon bias that is due to translational selection from codon bias that is due to mutational bias or other factors that are consistent within a strain but vary between strains [17]. We have made this distinction by measuring the codon bias of a set of highly expressed genes relative to the codon bias of the genome from which they were derived; this measure is expressed as the quantity $\Delta N'_c$ (Materials and Methods). Using codon

Figure 3.2: Translational output as a function of RNA content

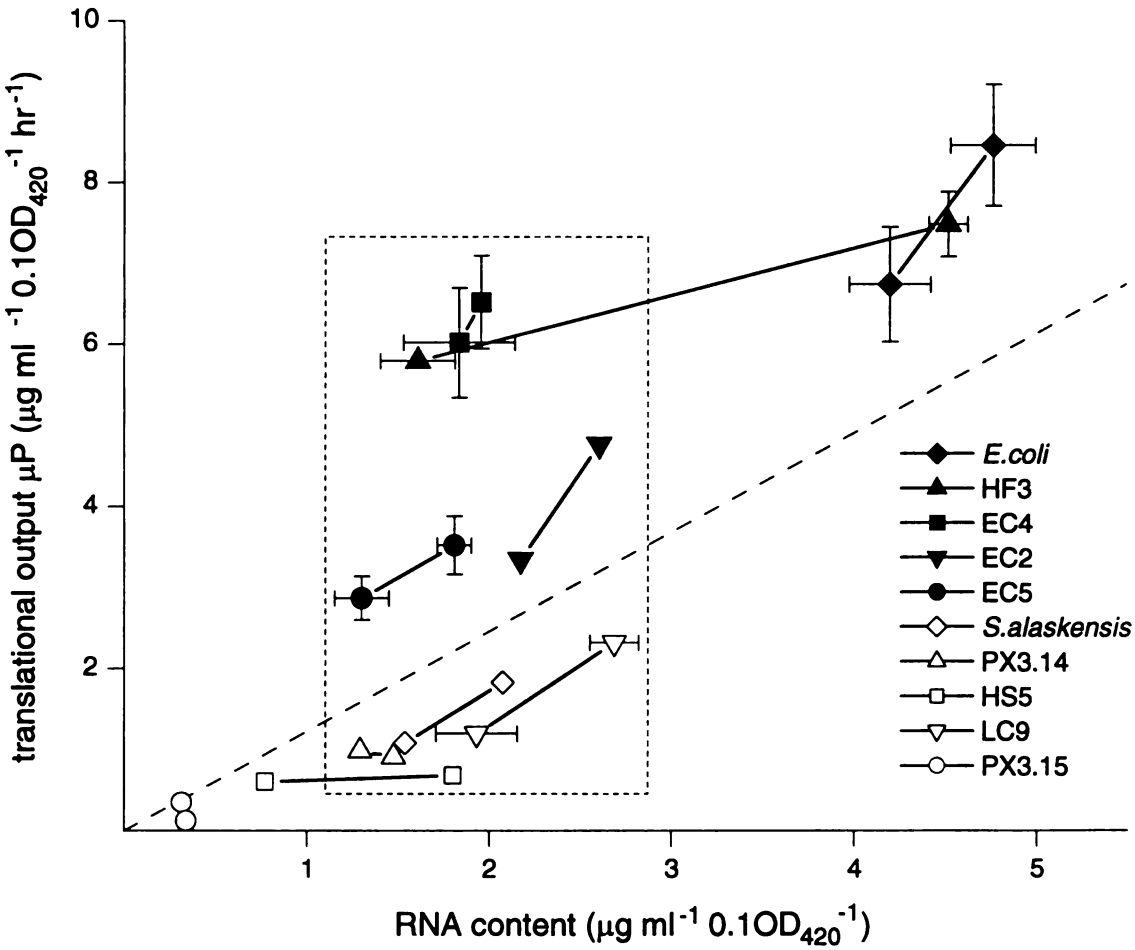


Figure 3.2. Translational output (μP) per optical density (0.1 OD_{420}) as a function of RNA content per optical density is shown for rapidly responding bacteria (solid symbols) and slowly responding bacteria (open symbols). In the range of RNA content represented by both ecological strategies (data points within boxed area), the average translational output of the rapid responders is 3.6-fold higher than that of the slow responders. Because translational power is the ratio of the ordinate to the abscissa on this plot, a straight line through the origin represents a contour connecting points of equal translational power. One such contour has been drawn that separates all data derived from rapid responders (higher translational power, above and left of contour) from all data derived from slow responders (lower translational power, below and right of contour). Data points connected by a line represent the same strain grown in two different media. Data points and error bars are the means and standard errors of measurements on at least 3 independent replicate cultures per strain-medium combination. Error bars smaller than or of comparable size to plot symbols are not shown.

data from 76 completely sequenced bacterial genomes [21], we have found that $\Delta N'_c$ is correlated with *rrn* copy number (Figure 3.3a), implying that stronger translational selection is associated with a rapid growth response strategy, as predicted.

A likelihood ratio method [25] using explicit phylogenetic information (see Appendix) demonstrates that the association of strong translational selection and high *rrn* copy is significant throughout the phylogeny, explaining about a quarter of the evolutionary variation in these traits across bacteria ($R^2=0.29$, $p<10^{-6}$). For this likelihood ratio test as well as the test of tRNA gene number reported below, the conclusions are not altered by the inclusion of poorly resolved details in the bacterial phylogeny (see Appendix). A traditional correlation analysis (which would show a stronger correlation between the traits) is inappropriate for demonstrating a biological relationship between translational selection and *rrn* copy number, since it assumes the statistical independence of data derived from species with various durations of shared evolutionary history [25].

Since our measure of the strength of translational selection depends on both the codon bias of the genome as a whole and the codon bias of the HE set, we investigated whether either of these component traits were correlated with *rrn* copy number, using Wright's N_c as our measure of codon bias. Considered separately, the codon bias of neither the highly expressed genes nor the entire genome is correlated with *rrn* copy number, as shown in Figure 3.4. This result indicates that *rrn* copy number is related specifically to the strength of translational selection, not to trends in codon bias that may be influenced by factors other than translational selection.

Figure 3.3: Correlation between *rrn* copy number and the strength of translational selection

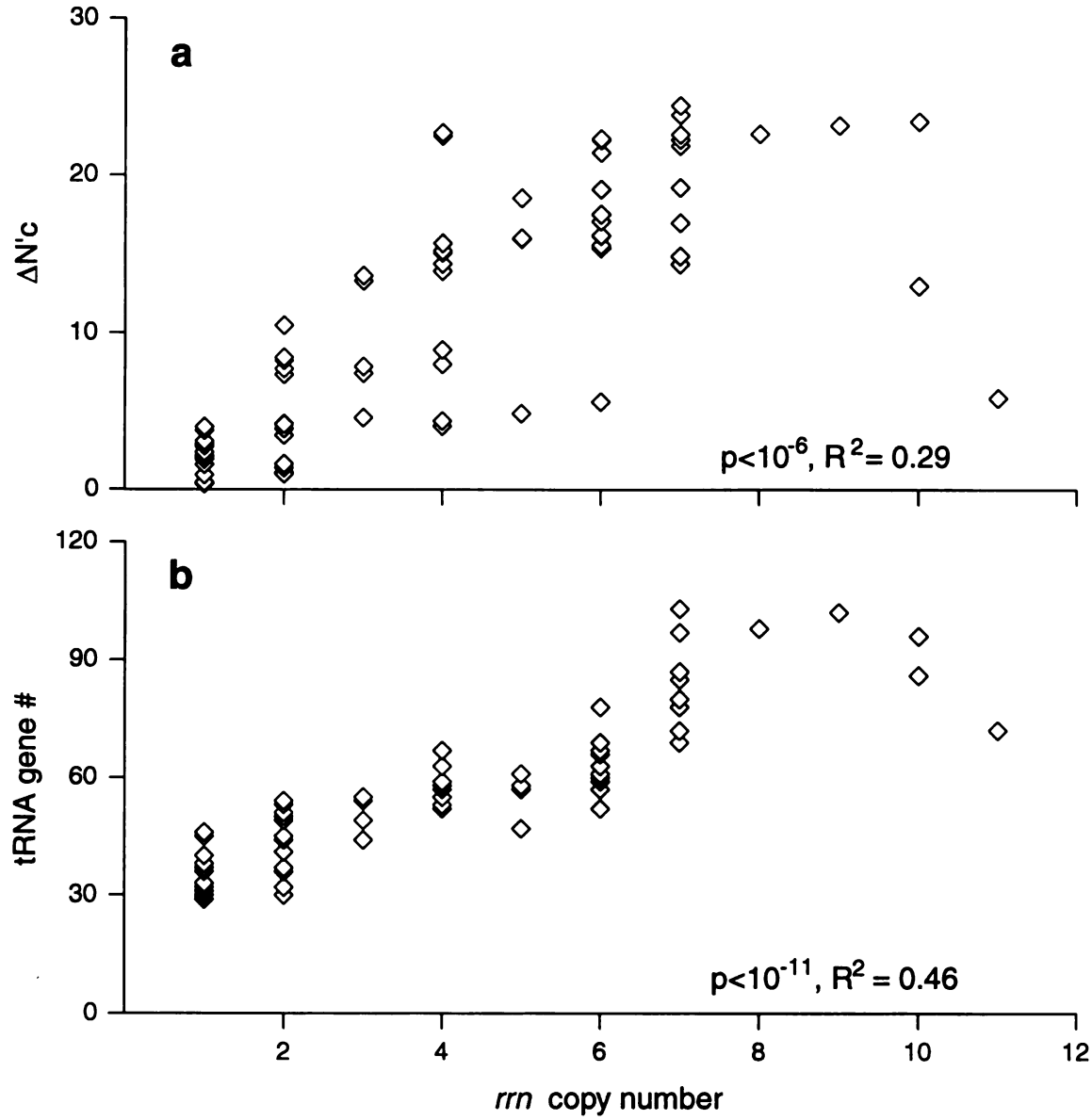


Figure 3.3. Correlation between *rrn* copy number and the strength of translational selection, measured either as ΔN'c (panel a, see Materials and Methods for explanation) or as tRNA gene copy number (panel b). Each data point represents a sequenced genome (n=76); list of taxa and phylogenetic information is provided in the Appendix. Strength and significance of correlations between *rrn* copy number and measures of translational selection across taxa are assessed by log likelihood ratio tests (see Materials and Methods for explanation).

Figure 3.4: Relationship between *rrn* copy number and the codon bias of entire genomes and of highly expressed genes

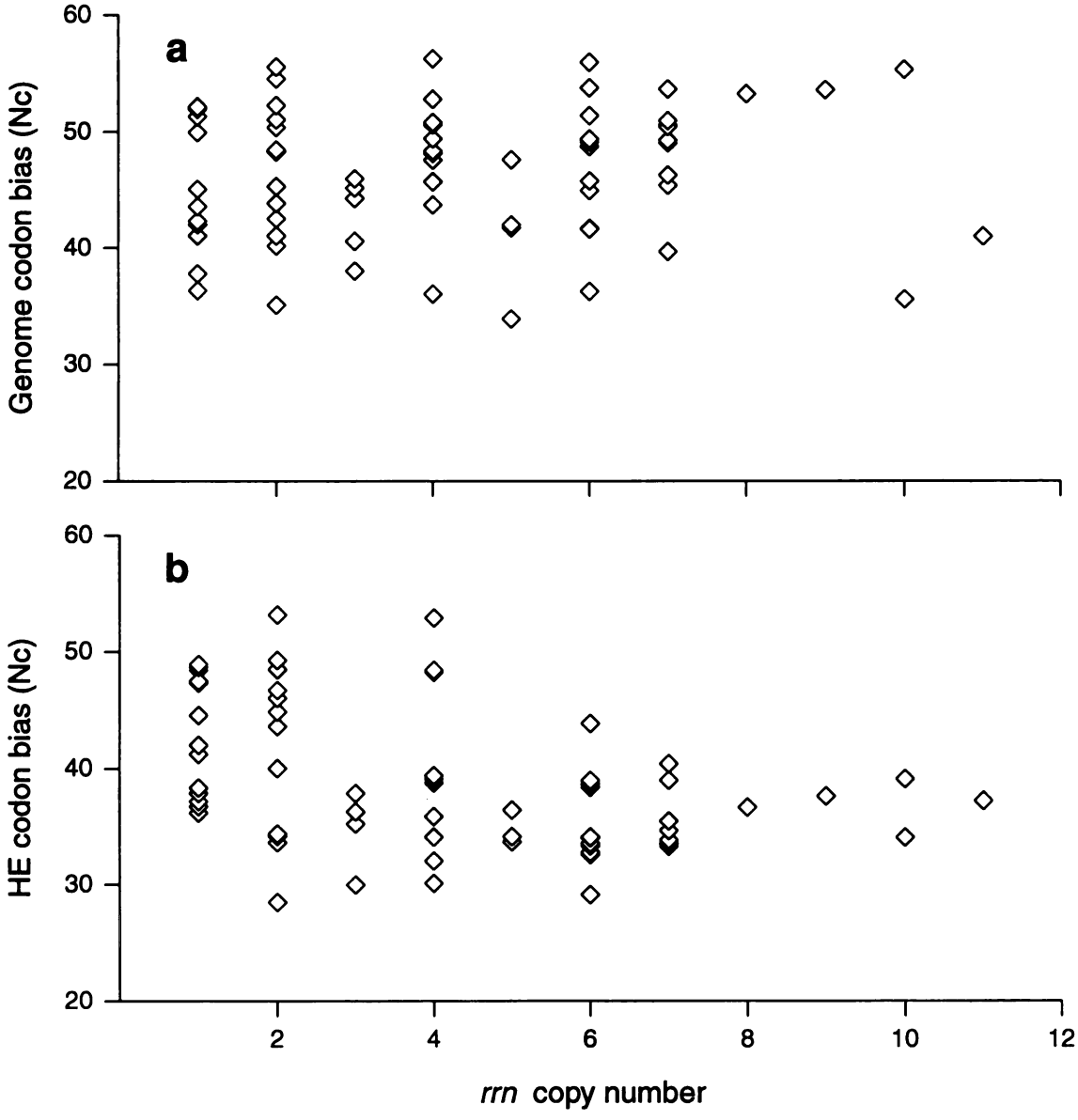


Figure 3.4. Relationships between *rrn* copy number and the degree of codon bias either in all predicted genes of a genome (panel a) or in a set of 8 highly expressed (HE) genes (panel b, see Materials and Methods for a list of genes). Codon bias is measured as Wright's Nc[23]. Each data point represents a sequenced genome (n=76); list of taxa and phylogenetic information is provided in the Appendix. The correlation between *rrn* copy number and codon bias across genomes is significant in neither set of genes, as assessed by log likelihood ratio tests (see Materials and Methods for details).

An independent measure of the strength of translational selection is the number of tRNA genes per genome. All organisms use 'wobble' pairing in the 3rd codon position to translate the 61 sense codons with a reduced number of unique anticodons, typically around 40. However, each anticodon can be represented in the genome by 1 or more tRNA genes. Since higher gene dosage is associated with the abundant tRNAs that translate preferred codons, the total number of tRNA genes reflects the overall strength of translational selection [32,33]. Consistent with a previous study using a smaller set of microorganisms [33], we found that tRNA gene copy number in 76 sequenced bacterial genomes correlates with *rrn* copy number (R^2=0.46, $p<10^{-11}$ by likelihood ratio test), supporting our inference of an association between strong translational selection and the rapid response strategy (Figure 3.3b).

Chapter 3 Discussion

The data of Figures 3.1 and 3.2 demonstrate that variation in translation power among bacteria is consistent with ecological factors, but not phylogeny. Specifically, rapidly responding bacteria have high translational power, whereas slowly responding bacteria have low translational power. Protein synthesis is commonly assumed to be fundamentally equivalent in all bacteria. Indeed, we would agree that all organisms accomplish translation by the same sequential reactions, using homologous components with conserved functions. However, these data do not show equivalent performance, but rather systematic differences between rapidly responding bacteria and slowly responding bacteria either in the translation rate, or in the fraction of active ribosome, or both. Our results are consistent with comparisons made from macromolecular data available in the literature (Table 1.1).

Codon usage analysis provides supporting evidence that translational performance differs between bacteria, and that this variation is influenced by ecological factors. Preferred codons tend to be translated by abundant tRNA species, or to form 3 canonical base pairs with their cognate tRNA, instead of relying on 'wobble' pairing in the third codon position. This allows preferred codons to be translated more quickly and with fewer errors [16,34,35], enhancing translational power. Hence, selection for translational power is one of the two major factors contributing to codon bias. However, to use codon bias as an indicator of the strength of selection for translational power, the effects of translational selection must be distinguished from those of the other major factor affecting codon use, which is mutational bias. Mutational bias establishes a characteristic, organism-specific

nucleotide composition in the chromosome [36]; it is generally supposed to be selectively neutral. It is also fairly uniform throughout a genome; spatial heterogeneity of mutational bias in the genome has been identified as the major source of variability in codon use within an organism only when the mutational bias is not extreme and translational selection is weak or absent [37-39]. This has allowed us to use genome-average codon use to control for differences in mutational bias and other genome-wide processes between organisms; we can measure the strength of translational selection consistently across organisms by the deviation of codon use in the HE set relative to this control.

The complexity of correcting for mutational bias in a codon-based measure of translational selection can be circumvented by using a different measure of the strength of translational selection, the total number of tRNA genes per genome. Because the benefit of codon bias depends in large part on having abundant tRNA molecules cognate to the preferred codons [16], and because the abundance of a particular tRNA species is strongly influenced by gene dosage [40,41], the total number of tRNA genes in an organism is a measure of the benefit it obtains from codon bias. As shown in Figure 3.3, both measures of the strength of translational selection are consistent, indicating that the strength of selection for translational power is strongest in bacteria with high *rrn* copy number. Our results relating both codon bias patterns and tRNA gene copy number to *rrn* copy number are similar to those of an earlier study based on 18 microbial genomes [33]. Both data from the scientific literature (Table 1.1) and our own results (Figure 3.1) indicate that translational power is high in organisms with high *rrn* copy

number, consistent with the predicted ecological benefit of this trait for rapid growth and rapid acceleration of growth [18,19].

Codon bias is not merely a passive indicator of translational selection; codon bias itself directly influences translational power [16]. It is reasonable to ask whether the stronger selective pressure for translational power identified in the rapidly responding bacteria needs any additional mechanism, beyond codon bias, in order to exert its effect. In other words, we are asking if the variation in codon bias between rapid and slow responders is a sufficient explanation for the observed differences in translational power. It is not a trivial task to address this topic. A precise estimate of the extent to which codon bias accelerates translation in an organism would require knowledge of the translation rate for each codon as a function of the concentration of each of its cognate tRNA species, as well as knowledge of both codon expression frequency and individual tRNA concentrations *in vivo* as a function of growth rate. Such data is not yet available for *E. coli,* much less any other organism. However, the available data are sufficient for a rough estimate of the magnitude of the translation rate benefit of codon bias in *E. coli*; this estimate is the subject of Chapter 4. Our conclusion is that translation is at most 60% faster in *E. coli* than it would be in the absence of codon bias. While substantial, this effect is much smaller than the variation in translational power inferred from macromolecular data in the literature (Table 1.1) or from our own measurements (Figures 3.1 and 3.2). Hence, differences in translational performance between microbes adapted for different ecological strategies must reflect differences in the translational apparatus itself.

A tradeoff mediated by processivity errors

One mechanism that could account for varying performance of the translational apparatus between organisms is an evolutionary tradeoff, so that organisms with high translational power would necessarily be worse in some other fitness-related trait. Organisms would be selected to have low translational power relative to others, if they occupied a niche where the alternative trait made a larger relative contribution to their fitness. In Chapter 1, we noted evidence for such a tradeoff involving starvation survival, obtained during the adaptation of natural *E. coli* isolates to laboratory conditions. The introduction of this chapter reviews evidence that processivity error rates, specifically dropoff rates, are surprisingly high in laboratory adapted *E. coli.* Accordingly, we propose the existence of a tradeoff between high translational power and high translational yield, mediated by the frequency of processivity errors.

Although a complex tradeoff in which missense accuracy affects both the translation rate and the dropoff frequency has been proposed on theoretical grounds [9,42], we know of no direct experimental evidence indicating that rapid translation must be associated with frequent dropoff errors. However, evidence for the existence of some sort of mechanistic connection between the missense error rate, the processivity error rate, and the translation rate is strong. Ribosome mutations derived from laboratory-adapted strains of *E. coli* that alter the frequency of missense errors in either direction consistently decrease the translation rate and increase the frequency of processivity errors, most of which are dropoff events [8,10]. *In vitro* measurements of the rates of peptidyl-tRNA dissociation from two different sites on the ribosome suggest the involvement of two classes of

dropoff events: missense-restrictive mutations may facilitate dropoff from the A-site, whereas missense-enhancing mutations may facilitate dropoff from the P-site [12]. Hence, the laboratory 'wild-type' ribosome of *E. coli* appears to combine optimality according to two criteria simultaneously: rapid translation, and a low frequency of processivity errors, obtained by balancing two different classes of dropoff events. However, the frequency of processivity errors for these rapidly translating ribosome remains high enough that only about 90% of ribosomes successfully translate the length of an average gene. If the selective pressure for rapid translation were relaxed, further reductions in the rate of processivity errors may be possible. The discovery that selection for rapid translation is linked to a decline in the ability to survive starvation among natural *E. coli* strains [5,6] is consistent with such a tradeoff between translation rate and the frequency of dropoff errors. If starvation survival requires energy derived from endogenous resources, some of which is used for protein synthesis, enhancing translational yield by reducing the frequency of processivity errors could promote survival by conserving endogenous resources.

There may be a second, independent link between processivity errors and translational power, mediated by protein factors involved in normal translational termination. The tradeoff would involve not translation rate, but the active fraction of ribosomes, the second component of translational power. When a stop codon occupies the ribosomal A-site, either release factor 1 (RF1) or release factor 2 (RF2) enters the A-site to facilitate cleavage of the nascent protein from the P-site tRNA; release factor 3 (RF3) then displaces RF1 or RF2 from the ribosome [43,44]. Ribosome release factor (RRF) replaces

RF3 in the A-site, and with the help of EF-G, it releases the deacylated tRNA from the P-site and allows the ribosome either to dissociate from or slide along the mRNA [45,46]. However, RF3, RRF and EF-G can also act with a sense codon in the A-site and peptidyl-tRNA in the P-site, generating a dropoff error [47]. Nonetheless, at an individual sense codon, elongation is far more likely than dropoff, which may be explained, at least in part, by the fact that amino acyl-tRNA selection is much faster than release factor activity. The ribosome typically requires ~20-50 ms to find a cognate ternary complex during elongation [48,49], whereas protein release by RF1/2 during termination occupies ~1-2 s [50] and the release of the ribosome appears to be even slower [51,52]. However, the inherent competition between elongation and factor-catalyzed dropoff at sense codons may explain why *E. coli* has not evolved to reduce the amount of time ribosomes spend sequestered at the end of messenger RNA. Decreasing the fraction of inactive ribosomes in the cell by accelerating post-termination ribosomal release (whether by regulatory changes to increase the concentration of RF3 and/or RRF, or by structural changes that accelerate the activity of these factors) may also increase the frequency of dropoff errors during elongation. A similar argument can be made that the slow action of RF1 and RF2 in releasing the nascent polypeptide is related to avoidance of false stops, when these factors act inappropriately during translational elongation. However, misrecognition of sense codons by RF1 and RF2 occurs less frequently than dropoff errors [50].

An alternative explanation

We are not the first to note a link between patterns of codon bias and ecological factors. Sharp and his colleagues have suggested that differences between strains in the strength

of translational selection (recognized as variation in patterns of codon bias) reflect differences in bacterial life history [17], suggesting that the importance of translational power varies with ecological strategy. More recently, this group has specifically identified a gradient across species in the strength of translational selection on codon use with a gradient in the rate of exponential growth or importance of exponential growth for the fitness of the organism [17,53]. *Helicobacter pylori* provides the most extreme contrast to the paradigm of translational selection represented by *E. coli*; reportedly it shows no evidence of translational selection in its codon use, despite the fact that its mutational bias is not strong enough to obscure translational selection if it exists [54]. Lafay *et al.* argue that translational power offers no benefit in the niche inhabited by *H. pylori*, suggesting that it does not experience competitive exponential growth [54]. This appears to be consistent with the harsh environment of the gastric mucosa, where no other microbial species are known to persist [54]. In effect, these authors have conceived of a spectrum of ecological strategies very similar to the spectrum we describe between rapidly responding and slowly responding bacteria, and suggested that the relative positions of strains on this spectrum could be inferred by assessing the strength of translational selection. In this light, the correlations shown in Figures 3.3 can be seen as quantitative support for their expectation, confirming the utility of *rrn* copy number as an index of ecological strategy along this axis.

The explanation offered by Sharp and colleagues appears to be simpler than an explanation involving tradeoffs. It depends solely on variation in the strength of positive selection for translational power, rather than invoking as-yet untested mechanisms that

link translational power to other traits. We would certainly concur that the benefit of translational power varies between strains according to their ecological strategy, and that such variation contributes to patterns in codon use and translational power. However, we consider it unlikely that a benefit of translational power is completely absent in any organism. In *H. pylori,* for example, Lafay *et al.* acknowledge that the population structure of this organism reflects high rates of interstrain recombination, with well documented coinfection by multiple strains in a single individual [54]. These results make it more difficult to accept the claim [54] that episodes of intraspecific competition are completely absent.

As explained in Chapter 1, even conditions where growth is severely constrained by resource availability do not necessarily eliminate selection for higher translational power. Furthermore, the Ehrenberg-Kurland model of the cell [7] (see also Chapter 1) suggests that the contribution of a cell component to the fitness of an organism can be measured by its mass. If protein synthesis has little influence on the fitness of an organism, why would it devote a large amount of biomass to the translational apparatus? A large biomass investment in the protein synthesis system suggests that the rate of protein synthesis has limited the growth rate of the strain, at least in some of the conditions it has experienced over its evolutionary history. We have no information on the protein or RNA content of *H. pylori;* however, we know that some microbes with low translational power are making substantial biomass investments in ribosomes or in RNA (Table 1.1, Figure 2.7). The phenomenon of a large investment in the protein synthesis machinery along with low translational power becomes easier to understand if high translational

power has not only beneficial effects (which may or may not be realized depending on the ecological context), but deleterious effects as well.

One small, puzzling detail remains. If, as we suggest, reduced but positive selection for high translational power is opposed in some conditions by stronger selective pressure to avoid processivity errors, strains adapted to these conditions might be expected to have a translational apparatus with a low inherent frequency of processivity errors, but also to have biased codon use to increase translational power. If the translational apparatus itself is constrained to adopt a higher yield, lower power phenotype, selection for higher translational power will be focused on codon use and tRNA abundance. Why do we not find stronger evidence of translational selection among the slowly responding, low *rrn* copy number strains? It may be that selection to avoid processivity errors directly influences the degree of codon bias as well the phenotype of the translational apparatus. In order to supply a larger concentration of tRNAs cognate to preferred codons, either the total tRNA concentration must increase (hence diminishing the translation rate benefit of codon bias by imposing an increased biomass cost), or the concentration of tRNAs cognate to nonpreferred codons must decrease. However, the expression of codons translated by rare tRNAs increases the frequency of processivity errors (at least in *E. coli*), an effect that can be reversed by increasing the supply of the rare tRNAs [34]. The selective pressure to restrict processivity errors in slowly responding bacteria may constrain the disparity in abundance of different tRNA species, which in turn reduces the potential for a co-evolutionary response in codon frequency and tRNA abundance distributions that is the hallmark of translational selection.

Conclusion

Translation is an ancient process, predating the last common ancestor of all known life. It is both essential and expensive, representing a significant fraction of all biological activity, particularly for unicellular organisms. Most improvements to the translational apparatus that do not involve compromising one beneficial feature for another may well have become universal through more than 3.5 billion years of evolution by natural selection. Thus, engineers, ecologists, and evolutionary biologists might not be surprised at the suggestion that optimizing this complex, well-integrated biological apparatus for one aspect of performance involves sacrifices according to another criterion. We believe that evidence for such a tradeoff is found in correlations between the ecological strategy of a bacterial species and both the power of its translational machinery and the strength of selection influencing its codon use. Certainly, the specific mechanism we have proposed may be confirmed or refuted as our knowledge of translation becomes more complete. Nonetheless, these results demonstrate that the performance characteristics of the translational apparatus are not invariant, but adapt in response to selection. Comparative analysis of the translational apparatus among bacteria with different ecological strategies may improve our understanding of translation, and recognizing that translational performance is adaptive will contribute to our understanding of the diversity and distribution of microbial life on Earth.

Chapter 3 References

1. Russell, J.B. and G.M. Cook, 1995. *Energetics of bacterial growth - balance of anabolic and catabolic reactions.* Microbiological Reviews, **59**(1):48-62.

2. Ingraham, J.L., O. Maaløe and F.C. Neidhardt, 1983. *Growth of the Bacterial Cell.* Sinauer, Sunderland, Massachusetts.

3. Yusupov, M.M., G.Z. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H.D. Cate and H.F. Noller, 2001. *Crystal structure of the ribosome at 5.5 angstrom resolution.* Science, **292**(5518):883-896.

4. Ramakrishnan, V., 2002. *Ribosome structure and the mechanism of translation.* Cell, **108**(4):557-572.

5. Kurland, C.G. and R. Mikkola, 1993. *The impact of nutritional state on the microevolution of ribosomes*; pp. 225-237 in *Starvation in Bacteria*; S. Kjelleberg, ed. Plenum Press, New York.

6. Mikkola, R. and C.G. Kurland, 1992. *Selection of laboratory wild-type phenotype from natural isolates of Escherichia coli in chemostats.* Molecular Biology and Evolution, **9**(3):394-402.

7. Ehrenberg, M. and C.G. Kurland, 1984. *Costs of accuracy determined by a maximal growth rate constraint.* Quarterly Reviews of Biophysics, **17**(1):45-82.

8. Kurland, C.G., D. Hughes and M. Ehrenberg, 1996. *Limitations of translational accuracy*; pp. 979-1004 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

9. Kurland, C.G. and M. Ehrenberg, 1985. *Constraints on the accuracy of messenger RNA movement.* Quarterly Reviews of Biophysics, **18**(4):423-450.

10. Dong, H.J. and C.G. Kurland, 1995. *Ribosome mutants with altered accuracy translate with reduced processivity.* Journal of Molecular Biology, **248**(3):551-561.

11. Jorgensen, F. and C.G. Kurland, 1990. *Processivity errors of gene expression in Escherichia coli.* Journal of Molecular Biology, **215**(4):511-521.

12. Karimi, R. and M. Ehrenberg, 1996. *Dissociation rates of peptidyl-tRNA from the P-site of E. coli ribosomes.* EMBO Journal, **15**(5):1149-1154.

13. Karimi, R. and M. Ehrenberg, 1994. *Dissociation rate of cognate peptidyl-transfer RNA from the A-site of hyper-accurate and error-prone ribosomes.* European Journal of Biochemistry, **226**(2):355-360.

14. Manley, J.L., 1978. *Synthesis and degradation of termination and premature termination fragments of β-Galactosidase in vitro and in vivo.* Journal of Molecular Biology, **125**:407-432.

15. Menninger, J.R., 1976. *Peptidyl transfer RNA dissociates during protein synthesis from ribosomes of Escherichia coli.* Journal of Biological Chemistry, **251**:3392-3398.

16. Andersson, S.G.E. and C.G. Kurland, 1990. *Codon preferences in free-living microorganisms.* Microbiological Reviews, **54**(2):198-210.

17. Sharp, P.M., M. Stenico, J.F. Peden and A.T. Lloyd, 1993. *Codon usage - mutational bias, translational selection, or both?* Biochemical Society Transactions, **21**(4):835-841.

18. Klappenbach, J.A., J.M. Dunbar and T.M. Schmidt, 2000. *rRNA operon copy number reflects ecological strategies of bacteria.* Applied and Environmental Microbiology, **66**(4):1328-1333.

19. Stevenson, B.S. and T.M. Schmidt, 1998. *Growth rate-dependent accumulation of RNA from plasmid-borne rRNA operons in Escherichia coli.* Journal of Bacteriology, **180**(7):1970-1972.

20. Eyre-Walker, A., 1996. *Synonymous codon bias is related to gene length in Escherichia coli: Selection for translational accuracy?* Molecular Biology and Evolution, **13**(6):864-872.

21. Peterson, J.D., L.A. Umayam, T. Dickinson, E.K. Hickey and O. White, 2001. *The Comprehensive Microbial Resource.* Nucleic Acids Research, **29**(1):123-125.

22. Novembre, J.A., 2002. *Accounting for background nucleotide composition when measuring codon usage bias.* Molecular Biology and Evolution, **19**(8):1390-1394.

23. Wright, F., 1990. *The effective number of codons used in a gene.* Gene, **87**(1):23-29.

24. Comeron, J.M. and M. Aguade, 1998. *An evaluation of measures of synonymous codon usage bias.* Journal of Molecular Evolution, **47**(3):268-274.

25. Pagel, M., 1999. *Inferring the historical patterns of biological evolution.* Nature, **401**(6756):877-884.

26. Felsenstein, J., 1981. *Evolutionary trees from DNA sequences - a maximum likelihood approach.* Journal of Molecular Evolution, **17**(6):368-376.

27. Klappenbach, J.A., P.R. Saxman, J.R. Cole and T.M. Schmidt, 2001. *rrndb: the Ribosomal RNA Operon Copy Number Database.* Nucleic Acids Research, **29**(1):181-184.

28. Condon, C., D. Liveris, C. Squires, I. Schwartz and C.L. Squires, 1995. *Ribosomal RNA operon multiplicity in Escherichia coli and the physiological implications of rrn inactivation.* Journal of Bacteriology, **177**(14):4152-4156.

29. Cavicchioli, R., M. Ostrowski, F. Fegatella, A. Goodchild and N. Guixa-Boixereu, 2003. *Life under nutrient limitation in oligotrophic marine environments: An eco/physiological perspective of Sphingopyxis alaskensis (formerly Sphingomonas alaskensis).* Microbial Ecology, **45**(3):203-217.

30. Schut, F., J.C. Gottschal and R.A. Prins, 1997. *Isolation and characterisation of the marine ultramicrobacterium Sphingomonas sp. strain RB2256.* FEMS Microbiology Reviews, **20**(3-4):363-369.

31. Kjeldgaard, N.O. and C.G. Kurland, 1963. *The distribution of soluble and ribosomal RNA as a function of growth rate.* Journal of Molecular Biology, **6**:341-348.

32. Ikemura, T., 1985. *Codon usage and transfer RNA content in unicellular and multicellular organisms.* Molecular Biology and Evolution, **2**(1):13-34.

33. Kanaya, S., Y. Yamada, Y. Kudo and T. Ikemura, 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**(1):143-155.

34. Sorensen, H.P., H.U. Sperling-Petersen and K.K. Mortensen, 2003. *Production of recombinant thermostable proteins expressed in Escherichia coli: Completion of protein synthesis is the bottleneck.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, **786**(1-2):207-214.

35. Akashi, H. and A. Eyre-Walker, 1998. *Translational selection and molecular evolution.* Current Opinion in Genetics & Development, **8**(6):688-693.

36. Sueoka, N., 1988. *Directional mutation pressure and neutral molecular evolution.* Proceedings of the National Academy of Sciences of the United States of America, **85**(8):2653-2657.

37. McInerney, J.O., 1998. *Replicational and transcriptional selection on codon usage in Borrelia burgdorferi.* Proceedings of the National Academy of Sciences of the United States of America, **95**(18):10698-10703.

38. Lafay, B., A.T. Lloyd, M.J. McLean, K.M. Devine, P.M. Sharp and K.H. Wolfe, 1999. *Proteome composition and codon usage in spirochaetes: Species-specific and DNA strand-specific mutational biases.* Nucleic Acids Research, **27**(7):1642-1649.

39. Romero, H., A. Zavala and H. Musto, 2000. *Codon usage in Chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces.* Nucleic Acids Research, **28**(10):2084-2090.

40. 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 Escherichia coli translational system.* Journal of Molecular Biology, **151**(3):389-409.

41. Dong, H.J., L. Nilsson and C.G. Kurland, 1996. *Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates.* Journal of Molecular Biology, **260**(5):649-663.

42. Kurland, C.G. and M. Ehrenberg, 1987. *Growth-optimizing accuracy of gene expression.* Annual Review of Biophysics and Biophysical Chemistry, **16**:291-317.

43. Freistroffer, D.V., M.Y. Pavlov, J. MacDougall, R.H. Buckingham and M. Ehrenberg, 1997. *Release factor RF3 in E. coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner.* EMBO Journal, **16**(13):4126-4133.

44. Kisselev, L., M. Ehrenberg and L. Frolova, 2003. *Termination of translation: interplay of mRNA, rRNAs and release factors?* EMBO Journal, **22**(2):175-182.

45. Janosi, L., S. Mottagui-Tabar, L.A. Isaksson, Y. Sekine, E. Ohtsubo, S. Zhang, S. Goon, S. Nelken, M. Shuda and A. Kaji, 1998. *Evidence for in vivo ribosome recycling, the fourth step in protein biosynthesis.* EMBO Journal, **17**(4):1141-1151.

46. Kaji, A., M.C. Kiel, G. Hirokawa, A. Muto, Y. Inokkuchi and H. Kaji, 2001. *The fourth step of protein synthesis: Disassembly of the posttermination complex is catalyzed by elongation factor G and ribosome recycling factor, a near-perfect mimic of tRNA*; pp. 515-529 in *The Ribosome*; B. Stillman, ed. Cold Spring Harbor Laboratory Press.

47. Heurgue-Hamard, V., R. Karimi, L. Mora, J. MacDougall, C. Leboeuf, G. Grentzmann, M. Ehrenberg and R.H. Buckingham, 1998. *Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome.* EMBO Journal, **17**(3):808-816.

48. Bilgin, N., F. Claesens, H. Pahverk and M. Ehrenberg, 1992. *Kinetic properties of Escherichia coli ribosomes with altered forms of S12.* Journal of Molecular Biology, **224**(4):1011-1027.

49. Bilgin, N., L.A. Kirsebom, M. Ehrenberg and C.G. Kurland, 1988. *Mutations in ribosomal proteins L7/L12 perturb EF-G and EF-Tu functions.* Biochimie, **70**(5):611-618.

50. Freistroffer, D.V., M. Kwiatkowski, R.H. Buckingham and M. Ehrenberg, 2000. *The accuracy of codon recognition by polypeptide release factors.* Proceedings of the National Academy of Sciences of the United States of America, **97**(5):2046-2051.

51. Pavlov, M.Y., D.V. Freistroffer, V. HeurgueHamard, R.H. Buckingham and M. Ehrenberg, 1997. *Release factor RF3 abolishes competition between release factor RF1 and ribosome recycling factor (RRF) for a ribosome binding site.* Journal of Molecular Biology, **273**(2):389-401.

52. Pavlov, M.Y., D.V. Freistroffer, J. MacDougall, R.H. Buckingham and M. Ehrenberg, 1997. *Fast recycling of Escherichia coli ribosomes requires both ribosome recycling factor (RRF) and release factor RF3.* EMBO Journal, **16**(13):4134-4141.

53. Grocock, R.J. and P.M. Sharp, 2002. *Synonymous codon usage in Pseudomonas aeruginosa PA01.* Gene, **289**(1-2):131-139.

54. Lafay, B., J.C. Atherton and P.M. Sharp, 2000. *Absence of translationally selected synonymous codon usage bias in Helicobacter pylori.* Microbiology-UK, **146**:851-860.

Chapter 4. Differences in codon bias cannot explain differences in translational power between bacteria pursuing different ecological strategies

Chapter 4 Abstract

Translational power is a comparison of translational output to the biomass invested in translational machinery. Translational power is higher among bacteria adapted to exploit large fluctuations in resource availability via rapid growth, and lower among bacteria adapted to exploit small fluctuations in resource availability (Chapters 1 and 3). The selective benefit of high translational power among *Escherichia coli* and other microbes capable of rapid growth is believed to explain a correlation between the degree of codon usage bias and gene expression level in such organisms. Conversely, the absence of such a benefit has been inferred from the absence of a correlation between codon bias and gene expression level among at least some slowly growing microbes. To investigate whether differences in the degree of codon usage bias could explain differences in translational power between bacteria pursuing these contrasting strategies, we estimated the extent to which codon bias could accelerate translation in *E. coli*. We conclude that while codon bias offers a substantial benefit in terms of faster translation, it is insufficient to explain the observed differences in translational power.

Chapter 4 Introduction

Translational power is the translational output of a cell or culture normalized to the amount of biomass invested in the translational apparatus. The concept was first introduced to facilitate quantitative comparisons of translational performance between different growth rates within a single bacterial strain [1]. (Translational power is a new term we are introducing to replace the terms 'ribosome efficiency' and 'translational efficiency' that have previously been used for this concept; the rationale for this semantic distinction is provided in Chapter 1.) The initial belief that translational power is nearly constant across a wide range of growth rates, based both on empirical data and ecological arguments [2,3], has gradually given way to the current understanding that translational power increases with growth rate, at least in *E. coli* [4-7]. The question of whether translational power varies between microbial species has been addressed only rarely, in four studies that have made comparisons between *E. coli* and one other strain [8-11]. Each of these studies found that translational power was higher in *E. coli* than in a slowly growing strain; reanalysis of these studies as well as other data available in the literature suggests that the translational power of bacteria incapable of rapid growth ranges from less than 17% to 42% of the translational power of *E. coli* (Chapter 1). Furthermore, an experiment comparing bacteria of contrasting ecological strategies showed higher translational power among strains adapted for rapid growth in response to resource abundance, and lower translational power among strains incapable of such a response (Chapter 3). The quantitative differences were similar in magnitude to those found in data from the literature.

One factor capable of affecting translational power is biased use of synonymous alternative codons. In the standard translational code, 18 of the 20 amino acids are encoded by more than a single codon, but in many microorganisms, synonymous codons are not equally abundant [12]. The pattern first found in *E. coli* and *Bacillus subtilis* turns out to be common: the majority of genes within an organism show a preference for the same subset of codons, but the degree of bias towards the preferred subset is correlated with the expression level of the gene [13,14]. For some time, the consensus has been that such a pattern reflects selection for translational power [15,16]. Codon bias enhances translational power because preferred codons are translated more rapidly than synonymous alternatives, either because the cognate tRNA species are relatively abundant, or because the codon-tRNA interaction is relatively efficient, or both [15,16]. Additional translation-related selective pressures, such as reducing the frequency and severity of translational errors, may have influenced the choice of preferred codons in some instances [17], but have not been demonstrated generally.

Since the benefit of using a preferred codon in a particular gene depends on how often the gene is translated, selection for preferred codons increases with expression level, accounting for the correlation noted above between the degree of codon bias and gene expression level. Although a number of examples have been documented where particular rare codons play a regulatory role [18-20], most non-preferred codons are not under strong selection [21,22], and a general role for non-preferred codons as down-regulators of gene expression has been rejected [16,23]. Rather, non-preferred codons become more abundant in genes expressed at lower levels because the benefit obtained

from preferred codons is reduced, allowing a higher proportion of random synonymous mutations to become fixed by genetic drift [24].

In contrast to codon bias due to translational selection, codon bias that is consistent in both magnitude and direction in genes that vary widely in expression level is explained most easily by mutational bias acting on DNA [15]. Generally, mutational bias has been associated with deviations from 50% G+C content that are uniform throughout the genome, but a regional variation in G+C content has been identified in *Mycoplasma genitalium* [25] and strand-specific mutational bias has been identified in a number of organisms [26]. However, spatial heterogeneity of mutational bias in the genome has been identified as the major source of variability in codon use within an organism only when translational selection is weak or absent [27-29]. While the effects of both translational selection and mutation bias are evident in some microbial genomes with moderately biased G+C content [30,31], several organisms with extremely high or low G+C content have been reported to show very little [32] or no [33-35] evidence of translational selection. Theoretical calculations indicate that if the strength of mutational bias exceeds a certain critical threshold, any pre-existing codon preferences that conflict with the mutational bias will be reversed [36]. In this case, codon use is almost entirely determined by the mutational bias, which influences genes equally regardless of expression level. Note that while a comparison between codon use and expression level within such a genome would not provide evidence of translational selection, this does not imply that deviations from the average codon usage would be selectively neutral in such an organism, or that the fitness penalty of such deviations would be independent of gene expression level. Some

organisms with moderate G+C content have also been reported to lack correlations between codon usage and gene expression level, particularly the spirochete *Treponema pallidum* [28] and the proteobacteria *Helicobacter pylori* [37]. The absence of evidence for translational selection is much more striking in these organisms, since they lack a strong mutational bias that could obscure such evidence. It has been suggested that for such organisms, competitiveness during exponential growth confers little or no fitness benefit [15,37], consistent with their slow growth rate and other characteristics of their ecological niche.

If variation in the strength of selection for translational power leads to differences in the degree of codon bias between microbes (superimposed on any differences in codon bias that can be attributed to variation in mutational bias), we wondered whether differences in codon bias could in turn explain the observed differences in translational power between microbes. To address this issue, we frame the following question: To what extent does codon bias accelerate the translation rate of *E. coli*, in comparison to a hypothetical organism with the same proteome composition and the same investment in the translational apparatus, but which lacks bias in codon use? For convenience, we will refer to this hypothetical *E. coli*-like organism with uniform use of synonymous alternative codons as 'Uni'. By 'same proteome composition', we mean that over a cell generation, the number of times each amino acid is used in translation is the same in Uni and in *E. coli*, although the frequencies of the codons directing that translation will differ for the 18 amino acids specified by multiple codons. By 'same investment in the translational apparatus', we mean that the total biomass of the translational apparatus is

the same in Uni as in *E. coli*, although ideally the allocation of that biomass among the translational components would be optimized for the unbiased codon use postulated for Uni. However, in order to apply empirical codon-specific translation rate data, we will impose a more stringent requirement on Uni, that the abundance of each individual component of the translational apparatus will be unchanged in comparison to *E. coli*. Due to this restriction and for other reasons, we make no claim to be able to answer our question precisely. However, our approximations are sufficient to conclude that differences in codon bias alone are unlikely to account for differences in translational power of the magnitude inferred to exist from macromolecular analysis of slowly growing and rapidly growing microbes.

Chapter 4 Materials and Methods

Calculation of the translation rate benefit of codon bias

Consider a cell in which a total of C_i codons of type i are translated during a single cell generation, so that the sum over all sense codons $C = \Sigma C_i$ is the total number of codons translated during a cell generation. (Hereafter we refer to the translational output over a cell generation as the proteome.) If we define $c_i = C_i/C$ as the proportion of all codons of type i in the proteome and r_i as the average translation rate of codons of type i, the total time required for replication of the proteome (i.e., the proteome generation time) will be

$$g_p = \frac{\sum C_i / r_i}{R} = C / R \sum c_i / r_i \qquad \text{Equation 1}$$

where R is the average number of ribosomes active in translation over the cell cycle. Codon bias in favor of rapidly translated codons will reduce g_p in comparison to uniform codon use. If a mutation changes the fitness of an organism from w to w', the benefit of the mutation is typically described as s, where $w'/w = 1 + s$. By analogy, and considering g_p to be inversely related to fitness, we can express the translation rate benefit of codon bias as

$$s_{bias} = \frac{g_p \text{ (uniform codon use)}}{g_p \text{ (biased codon use)}} - 1$$

The protein content is the same in Uni as in *E. coli* by hypothesis, and with the restrictive condition that the abundance of each individual component of the translational apparatus is unchanged in Uni, ribosome content will be the same also. Hence, the C/R term cancels from g_p in both the numerator and denominator of s_{bias}, leading to

$$s_{bias}(E.\ coli) = \frac{\sum c_i / r_i\ (\text{Uni})}{\sum c_i / r_i\ (E.\ coli)} - 1 \qquad \text{Equation 2}$$

Since amino acid frequencies are identical in *E. coli* and Uni, the disparity in translation rates between synonymous codons largely determines the magnitude of the translation rate benefit of codon bias.

We will use the same codon-specific translation rates (the r_i's) for both Uni and *E. coli,* again invoking the restrictive stipulation that the abundance of each individual tRNA species is unchanged. If rate constants for the interaction of each codon with each of its cognate tRNA species were known, we could calculate the optimal tRNA abundance distribution for the codon frequencies of Uni, and infer the resulting codon-specific translation rates [38,39]. However, *in vivo* codon-specific translation rate data are available only as codon averages, including translation from all tRNA species cognate to each codon. Hence, rate constants specific to each codon-cognate tRNA pair cannot be calculated from the available data for the codons translated by multiple tRNA species, and we cannot calculate an optimal tRNA abundance distribution for Uni. Insofar as the codon-specific translation rates measured in *E. coli* reflect an allocation of tRNA

abundance that would be sub-optimal for Uni (as we argue below), using the same codon-specific translation rates in both cases leads to an overestimate of the benefit of codon bias, a conservative error for our purposes.

To investigate the influence of comparing *E. coli* to a biologically realistic standard, instead of imposing strictly uniform use of synonymous codons on Uni, we also apply Equation 2 while allowing Uni to display as much codon bias as might be found in an actual organism with limited codon bias. We take *T. pallidum* as our example of a low-bias microbe, since it is a slowly-growing bacterium with little mutational bias (52.7% G+C), which reportedly lacks evidence of translational selection as well [28]. *T. pallidum* has the second-most uniform codon use over all predicted genes in the genome (assessed as Wright's effective number of codons [40]) for 108 bacterial and archaeal species for which complete genome sequences were available in June, 2003 (data not shown).

Data Sources

For the codon frequencies used in synthesizing the proteome of *E. coli*, we rely on the data of Dong *et al.* at 2.5 dbl hr^{-1} [41], compiled from public gene sequence databases and protein abundance data derived from 2D gel electrophoresis studies [42,43]. The absolute frequencies shown in Table 4.1 have been recalculated from [41] with initiation and stop (including selenocysteine) codons removed. To compare the effect of codon bias in *E. coli* to the effect of a small, but biologically plausible, amount of codon bias, we generated a low bias set of codon frequencies (Table 4.1) that retain the same amino acid frequencies as *E. coli,* as well as the same rank order of codon frequency within

Table 4.1 footnotes

a. Proteome codon frequencies from reference 41 for *E. coli* growing at 2.5 dbl hr^{-1}, modified slightly as described in Materials and Methods.

b. Low bias codon frequencies representing the degree of codon bias present in the genome of *T. pallidum,* generated as described in Materials and Methods.

c. Summed abundance of all tRNA species cognate to the listed codon, expressed as a percentage of total tRNA, based on tRNA abundance data of references 41 and 44 and cognate specificity of reference 45, modified slightly as described in Materials and Methods. Values for all codons sum to >100%, a result of the partially overlapping codon specificity of many tRNA species.

d. Empirically determined relative rates of ternary complex selection at the listed codon from reference 46, expressed relative to the rate of a uniform competing frameshift event. Rate for codons CGC and CGA modified as described in Materials and Methods. Estimates of s_{bias} were made using the rates as listed or with a correction for the duration of translocation, as described in Materials and Methods.

e. Predicted relative translation rates based on the empirical rates of column 6 and scenarios as described in Materials and Methods. Estimates of s_{bias} were made using empirical rates of column 6 for YNN codons in preference to the predicted rates shown in parentheses; predicted rates are shown for comparison only.

f. Predicted relative translation rates from theory of reference 39, modified slightly as described in Materials and Methods, using the codon frequency data of column 3 and the cognate tRNA abundance data of column 5.

Table 4.1: Codon data

Codon	AA	Codon Frequency		tRNA	Empirical	Predicted Rel. Translation Rates			
		E. coli[a] (x 10^{-3})	low bias[b] (x 10^{-3})	Abund.[c] (%)	Rel. Trans. Rates[d]	Sc. 2[e]	Sc. 3[e]	Sc. 4[e]	Sc. 5[f]
UUU	Phe	8.0	9.6	1.5	8.5	(7.6)	(6.3)	(3.2)	3.7
UUC	Phe	23.4	21.8	1.5	12.0	(12.0)	(10.8)	(10.8)	6.4
UUA	Leu	2.8	6.6	2.7	4.3	(5.1)	(5.2)	(1.9)	7.1
UUG	Leu	4.3	17.1	3.8	8.7	(6.0)	(6.5)	(6.5)	8.9
UCU	Ser	16.5	10.7	3.4	11.6	(10.3)	(9.5)	(9.5)	7.1
UCC	Ser	11.8	7.2	1.2	14.7	(9.0)	(8.0)	(7.3)	6.0
UCA	Ser	2.0	5.4	2.1	7.0	(4.6)	(3.3)	(4.6)	2.5
UCG	Ser	2.5	7.1	2.6	9.0	(5.0)	(3.7)	(0.4)	2.8
UAU	Tyr	6.8	11.3	2.7	4.3	(7.2)	(6.1)	(2.8)	7.3
UAC	Tyr	16.6	12.1	2.7	8.4	(10.4)	(9.5)	(9.5)	11.5
UGU	Cys	2.8	3.1	2.1	4.0	(5.2)	(5.4)	(1.9)	7.6
UGC	Cys	3.8	3.6	2.1	7.0	(5.8)	(6.3)	(6.3)	8.9
UGG	Trp	7.1	7.1	1.5	5.0	(7.3)	(7.3)	(7.3)	6.4
CUU	Leu	3.9	15.2	2.6	8.4	(5.8)	(3.9)	(9.8)	6.3
CUC	Leu	4.1	16.0	1.7	11.0	(5.9)	(4.1)	(12.0)	6.5
CUA	Leu	0.8	4.4	0.9	0.6	(3.6)	(0.6)	(0.6)	2.9
CUG	Leu	61.2	17.9	7.3	14.4	(18.6)	(15.7)	(15.7)	24.9
CCU	Pro	4.4	9.0	1.8	8.4	(6.1)	(4.5)	(7.3)	2.7
CCC	Pro	1.1	6.2	1.1	9.6	(3.9)	(2.3)	(0.5)	1.3
CCA	Pro	5.2	12.1	0.8	1.6	(6.5)	(4.9)	(9.0)	2.9
CCG	Pro	29.0	12.4	1.5	2.5	(13.2)	(11.6)	(11.6)	6.8
CAU	His	6.8	9.0	1.2	4.0	(7.2)	(6.3)	(2.7)	3.5
CAC	His	14.3	12.1	1.2	8.0	(9.7)	(9.1)	(9.1)	5.1
CAA	Gln	7.1	10.7	1.2	5.6	(7.3)	(5.8)	(3.4)	3.3
CAG	Gln	27.5	23.9	2.3	10.0	(12.9)	(11.4)	(11.4)	6.4
CGU	Arg	44.2	21.4	7.5	14.0	(16.0)	(13.7)	(13.7)	31.3
CGC	Arg	20.8	19.1	7.5	11.5	(11.4)	(9.4)	(10.5)	21.4
CGA	Arg	0.7	11.3	7.5	3.0	(3.5)	(1.7)	(0.6)	3.9
CGG	Arg	0.6	5.1	0.6	0.8	(3.4)	(1.6)	(0.6)	2.6

a-f. See facing page for footnotes.

Table 4.1: continued

Codon	AA	Codon Frequency		tRNA	Empirical	Predicted Rel. Translation Rates			
		E. coli[a]	low bias[b]	Abund.[c]	Rel. Trans.				
		(x 10^{-3})	(x 10^{-3})	(%)	Rates[d]	Sc. 2[e]	Sc. 3[e]	Sc. 4[e]	Sc. 5[f]
AUU	Ile	15.9	21.5	6.8		10.2	8.2	4.1	17.1
AUC	Ile	44.2	28.0	6.8		16.0	13.7	13.7	28.6
AUA	Ile	0.5	11.2	0.3		3.3	1.5	0.6	1.4
AUG	Met	21.8	21.8	1.4		11.7	10.5	10.5	5.7
ACU	Thr	20.8	15.0	3.8		11.4	9.9	8.7	7.6
ACC	Thr	26.9	18.7	1.8		12.8	11.3	11.3	8.7
ACA	Thr	2.6	10.0	2.0		5.1	3.5	0.5	2.7
ACG	Thr	4.2	10.8	2.9		6.0	4.5	7.1	3.4
AAU	Asn	5.7	16.7	2.1		6.7	5.1	3.5	3.9
AAC	Asn	29.4	18.4	2.1		13.3	11.7	11.7	8.9
AAA	Lys	55.4	39.6	3.1		17.8	15.0	15.0	12.8
AAG	Lys	17.4	33.2	1.2		10.6	8.4	4.5	2.9
AGU	Ser	2.2	6.8	1.7		4.8	3.8	2.3	3.4
AGC	Ser	9.4	7.2	1.7		8.2	7.9	7.9	7.0
AGA	Arg	0.6	5.4	1.1		3.4	4.9	0.6	3.7
AGG	Arg	0.0	4.5	1.7		2.6	1.1	0.6	0.8
GUU	Val	43.5	38.8	7.9		15.9	13.6	13.6	23.3
GUC	Val	7.7	13.2	2.0		7.5	5.7	0.6	9.8
GUA	Val	22.5	20.1	6.0		11.8	9.8	10.5	16.8
GUG	Val	15.1	16.8	6.0		9.9	8.0	8.5	13.7
GCU	Ala	39.8	38.6	7.2		15.3	13.2	13.2	19.4
GCC	Ala	11.9	15.7	1.1		9.0	7.2	0.5	10.6
GCA	Ala	25.1	30.6	6.1		12.4	10.4	10.1	15.4
GCG	Ala	24.3	16.2	6.1		12.2	10.3	8.2	15.1
GAU	Asp	19.4	22.3	4.4		11.1	9.3	3.7	14.0
GAC	Asp	34.0	31.1	4.4		14.2	12.4	12.4	18.5
GAA	Glu	58.3	39.8	8.5		18.2	15.3	15.3	35.7
GAG	Glu	17.1	35.6	3.4		10.5	8.3	4.5	7.7
GGU	Gly	45.9	23.7	9.3		16.3	13.9	13.9	21.0
GGC	Gly	34.4	23.0	7.3		14.3	12.1	10.7	18.2
GGA	Gly	1.3	17.8	1.9		4.1	2.3	0.6	3.5
GGG	Gly	2.4	19.5	3.2		4.9	3.2	8.7	4.8

a-f. See page 126 for footnotes.

synonymous groups of codons. The low bias frequencies were generated from relative codon frequencies over all predicted genes in the complete genome sequence of *T. pallidum* [47] (obtained from the website of the National Center for Biological Information, www.ncbi.nlm.nih.gov). By relative codon frequencies, we mean the absolute frequency of a codon divided by absolute frequency of the amino acid it encodes. The set of *T. pallidum* relative codon frequencies for a particular amino acid were multiplied by the absolute frequency of that amino acid in the *E. coli* proteome; the resulting set of absolute codon frequency values were assigned to the codons of that amino acid in the low-bias set so as to retain the same rank order of codon frequency among synonyms as exists in the *E. coli* proteome. For example, the total number of isoleucine codons and the identity of the 1st, 2nd and 3rd most common isoleucine codons is the same in the low bias set as in the *E. coli* proteome. However, the relative frequencies of the 1st, 2nd and 3rd most common isoleucine codons in the low bias set are the same as the relative frequencies of the 1st, 2nd and 3rd most common isoleucine codons in the *T. pallidum* genome.

To represent codon-specific translation rates, we use the relative rate data (the quantity R_{tRNA}/R_{shift}) of Curran and Yarus [46] for the 29 sense codons beginning with U or C (YNN codons, Y = pyrimidine). Although incomplete, this is by far the largest data set available for *in vivo* translational kinetics. The original publication transposed values reported for two arginine codons, CGC and CGA [48]; we have corrected this error. We also revised the rate measured for CGA downward, to account for interference from the bulky wobble position inosine-adenine base pair in the P site that results from translation of a CGA codon. Such interference is strongly suggested to slow selection of a ternary

complex at the codon *subsequent* to CGA [48]; such an effect would not have been measured with the experimental system of reference 46, but is appropriate to include as a codon-specific effect of CGA on translation rate. In the absence of more precise data, we reduced the rate measured for CGA by a factor of 3, the factor by which CGA reduces read-through of a following stop codon by a suppressor tRNA in comparison to CGC [48]. This adjustment to the CGA rate brings these results into rough agreement with those of Sorensen and Pedersen [49], who used an experimental approach that would have detected a consistent effect of CGA on the translation rate of the subsequent codon, attributing it to slow translation of CGA itself. The relative rates of reference 46, modified as described above, are listed in Table 4.1.

The relative rates reported by Curran and Yarus [46] do not reflect the entire translational cycle, but rather the time required for selecting a cognate ternary complex at an empty, codon-programmed ribosomal A site, which is believed to occupy the majority of the elongational cycle [50,51]. Although peptide bond formation may be very rapid, the time required for the EF-G-catalyzed translocation of the ribosome to the subsequent codon (and the associated movement of P- and A-site tRNAs) may not be much shorter than the time needed for EF-Tu-catalyzed ternary complex selection [50]. Hence, in addition to calculations made using ternary complex selection to represent an entire cycle of translational elongation (assuming, in effect, that the duration of translocation is negligible), we also made calculations after modifying the reported rates by adding an invariant 'translocation time' to the variable 'ternary complex selection time' for all codons. The duration of translocation per codon was set at 40% of the average time

required to select a ternary complex containing $tRNA^{phe}$ at a UUU codon, consistent with the only quantitative measure of translocation rate that has been made in conditions approximating those *in vivo* [50]. Results from both sets of calculations are presented for each scenario (described below) that is based on these ternary complex selection rates. For convenience, hereafter we will refer to the relative rates of reference 46 as translation rates, rather than using the more accurate but cumbersome expression 'ternary complex selection rates'.

To calculate the total abundance of cognate tRNA for each codon, we assign cognate specificity largely according to Björk [45], and use the tRNA abundance data from Dong *et al.* (at 2.5 dbl hr^{-1}) [41]. We differ from Björk only in assuming that the leucine and glycine tRNAs with uridine in the anticodon wobble position (for which nucleotide modifications have not been characterized) will read codons ending in U, A and G, instead of A and G only. This would be the case if the wobble position U is modified to cmO^5U, as is done for each of the other 6 amino acids encoded by a full box of the translational code (i.e., amino acids for which the four XXN codons are synonyms). Following Björk, we assume that 40% of the tRNAs for glutamate, glutamine and lysine with uridine in the anticodon wobble position are modified to mnm^5Se^2U and thus read codons ending in A or G; the balance of these tRNA species are assumed to have mnm^5S^2U in the wobble position and read A-ending codons only [45]. The abundance of two pairs of isoaccepting tRNA species (Gln1 + Gln2 and Ile1 + Ile2) were reported as summed values by Dong *et al.* [41], since these individual species were not separated under the experimental conditions applied. We have resolved the summed values to the

abundance of individual species using the ratios of the individual abundance values as reported by Ikemura [44]. We show cognate tRNA abundance data in Table 4.1 as a percentage of total tRNA, omitting initiator and selenocysteine tRNAs; the sum of all values is greater than 100%, reflecting the partially overlapping specificity of many tRNA species.

Scenarios for applying incomplete empirical data

We address the incompleteness of codon-specific translation rate data in several ways. In Scenario 1, we assume that the effects of biased use of YNN codons on translation rate can be used to represent the effects of bias over all codons, without assigning particular translation rates to the unmeasured codons. However, since the YNN codons are almost half of all sense codons but only account for about a third of all expression (Table 4.1), they must be less highly expressed, on average, than the RNN codons (R = purine). Consequently, selection for translational power may have been weaker among YNN codons than RNN codons. Scenarios 2-4 address this potential deficiency by applying various strategies of assigning translation rates to the unmeasured codons that are consistent with observed patterns, but that could allow the effect of codon bias on translation rate to be greater among RNN codon than YNN codons. Scenario 5 abandons empirical codon-specific translation rate measurements completely, assigning translation rates to all codons on the basis of the proteome codon frequency and cognate tRNA abundance of *E. coli*, assuming optimality (i.e., maximal translation rate) according to theory developed by Solomovici *et al.* [39]. For the scenarios based on the empirical rates of reference 46 (i.e., Scenarios 1-4), we calculate the translation rate benefit of codon bias in two ways, making different assumptions about the duration of translocation. The two

assumptions are either that translocation is essentially instantaneous in comparison to the duration of ternary complex selection, or that translocation requires a short amount of time that is the same for all codon-cognate tRNA pairs. For all scenarios, we calculate the translation rate benefit of codon bias in *E. coli* in comparison to both uniform synonymous codon usage and a biologically plausible set of low-bias codon frequencies, derived as described above. The comparisons using the low bias codon frequencies assume a short, codon-independent duration of translocation.

Scenario 1: The 29 YNN codons encode 10 amino acids, 9 of which have multiple codons. For 7 of these 9 amino acids, the most common synonym is the codon with the fastest translation rate. One of the remaining amino acids is serine, for which the two fastest-translated codons are the two most abundant, although in reverse order, with relatively small differences between the two in both rate and abundance. Only proline appears to be anomalous; the 2 most abundant codons encode over 90% of all proline residues in the proteome [41], but support ternary complex selection about 3.5-fold more slowly than the 2 least abundant codons [46]. It has been suggested [46] that this anomaly could be adaptive; if proline, because of its unique structure, is found preferentially between protein domains [52] where slow translation may be important to permit cotranslational folding [53,54]. If proline is the only amino acid for which such contrarian selection pressure is more important than selection for translational power, including proline codons in a sample intended to represent all codons will lead to an underestimate of s_{bias}. Hence, in Scenario 1 we apply Equation 2 over YNN codons, with the calculated translation time for non-proline YNN codons weighted by a factor of 3.2,

which scales the expression level of these codons to the expression level of all non-proline codons. In other words, we assume the effects of codon bias on translation rate among the 25 non-proline YNN sense codons are representative of the effects of codon bias among all 57 non-proline sense codons, whereas the translation rates measured for proline codons are applied only to themselves.

Scenario 2: Curran and Yarus noted that among highly expressed genes, there is a significant tendency for rapidly-translated codons to be used frequently, although the relationship appears to be nonlinear [46]. We observe the same pattern comparing their relative rate data to the proteome codon frequency data of Dong *et al.* [41] at the highest growth rate. For non-proline YNN codons, the best fit ($R^2 = 0.56$) of a quadratic relationship passing through the origin between the codon frequency and translation rate data of Table 4.1 is $c_i = 0.205\ r_i - 0.522\ r_i^2$. We use this equation to predict translation rates from codon frequency for all RNN codons, as shown in Table 4.1. Since our objective is to obtain a reasonable estimate the codon-specific translation rate for codons which have not been measured, not to defend a particular model of the relationship between codon frequency and translation rate, we make no attempt to justify a quadratic fit in comparison to other possible functional relationships. The predicted rates for RNN codons and the measured rates for YNN codons (Table 4.1) are used with Equation 2 to estimate the translation rate benefit of codon bias under Scenario 2.

Scenario 3: The preceding scenario applied to the YNN codons tends to predict translation rates among synonymous alternatives that are not as disparate as those

actually observed. Furthermore, the fit of a functional relationship between codon frequency and translation rate among YNN codons is better when only preferred codons are considered, instead of all codons. Hence, we fit a quadratic relationship passing through the origin to data from 10 preferred non-proline YNN codons, obtaining $c_i = 0.352\ r_i - 1.611\ r_i^2$ ($R^2 = 0.81$). Among the 10 preferred codons, we include UGG, the sole tryptophan codon, and UUG, the preferred leucine codon within the UUR split box although not the preferred leucine codon overall. We then applied this equation to predict translation rates from codon frequencies for 12 preferred RNN codons, including AUG, the sole methionine codon, and AGG and AGC, the preferred arginine and serine codons within their respective split boxes, although not the preferred codons overall. For non-preferred RNN codons, translation rate is predicted by multiplying the predicted rate for the preferred synonym (within the full or split box) by the ratio of the square roots of the codon frequencies for the non-preferred and preferred codons:

$$\text{predicted } r_{nonpref} = \text{predicted } r_{pref} \frac{\sqrt{c_{nonpref}}}{\sqrt{c_{pref}}} \qquad \text{Equation 3}$$

This relationship was chosen both because a dependence on the square root of codon frequency has been suggested repeatedly in theoretical investigations of optimal translation rates [38,39,55-57], and because for all non-preferred RNN codons, this relationship leads to a greater disparity of predicted translation rates compared to the preferred synonym than the regression of Scenario 2. (It also predicts a greater translation rate disparity than is observed for the majority of non-preferred YNN codons.)

When both the quadratic regression for preferred codons and Equation 3 for non-preferred codons are applied to predict the translation rate of non-proline YNN codons, the correlation of predicted with measured translation rates is comparable to that attained with Scenario 2 ($R^2 = 0.57$). The predicted rates for RNN codons and the measured rates for YNN codons (Table 4.1) are used with Equation 2 to estimate the translation rate benefit of codon bias under Scenario 3.

Scenario 4: This scenario is generated in three steps, with the goal of generating an estimate of the translation rate benefit of codon bias that is consistent with the most extreme empirical observations. First, three rare RNN codons (AGG and AGA for arginine and AUA for isoleucine, all with $c_i < 0.1\%$) are assigned the slowest relative translation rate observed among YNN codons ($r_i = 0.6$ for the rare leucine codon CUA). Second, the translation rates for preferred RNN codons within full or split boxes (except AGG) are estimated according to the regression equation described for Scenario 3. Finally, the translation rates for non-preferred codons (except AGA and AUA) are predicted from the preferred synonym using the ratios of the most disparate translation rates observed empirically among synonymous alternatives, treating split boxes and full boxes of the translational code separately. The most extreme ratio observed among translation rates in a split box is 3.375, for glutamate codons in the study of Sorensen and Pedersen [49]. The most extreme ratios observed for translation rates of codons in a full box is 1:1.3:1.6:24 for the CUN leucine codons in the study of Curran and Yarus [46]. (Exploring other rate values $1 \leq x \leq y \leq 24$ in ratios of the form $1{:}x{:}y{:}24$ failed to find any that greatly increased the estimated benefit beyond that using the leucine ratios, data not

shown.) Although this scenario is based on extreme observations, applying these 3 rules to the non-proline YNN codons leads to a correlation of predicted and measured translation rates ($R^2 = 0.67$) somewhat better than that obtained under Scenarios 2 and 3. The predicted rates for RNN codons and the measured rates for YNN codons (Table 4.1) are used with Equation 2 to estimate the translation rate benefit of codon bias under Scenario 4.

Scenario 5: In contrast to the preceding scenarios that extend codon-specific translation rate measurements of a subset of codons to make an estimate of the translation rate benefit of codon bias over all codons, Scenario 5 incorporates a theoretical prediction of optimal translation rates based only on empirical codon frequency and cognate tRNA abundance data. Solomovici *et al.* [39] assume that selection on synonymous codon frequencies reflects intrinsic differences in rate constants for a cognate tRNA interacting with preferred and non-preferred codons, while the total tRNA abundance and amino acid composition are fixed. They demonstrate that the fastest overall translation rate is obtained when the square roots of synonymous codon frequencies are proportional to the rate constants for cognate tRNA interacting with the codons. They assume further that the rate constants for the interaction of all non-degenerate or preferred codons with their preferred cognate tRNA are identical, so the translation rate for these codons is proportional to cognate tRNA abundance. We modified the approach of reference 39 to allow for greater degeneracy in translation [45] (see also the earlier comment regarding the codon specificity of leucine and glycine tRNAs), and applied it using the codon frequency and tRNA abundance data of Dong *et al.* [41], modified as shown in Table 4.1.

The predicted relative translation rates for YNN codons (i.e., the recalculated quantities d_{ij} and d_{im_j} of reference 39 for codons with single or multiple cognate tRNAs, respectively) are not in good agreement with observed relative rates of Curran and Yarus [46] ($R^2 = 0.30$). However, the empirical codon frequencies of Dong *et al.* [41] are correlated more closely with predicted relative rates of Scenario 5 ($R^2 = 0.70$) than with the empirical relative rates of Curran and Yarus [46] ($R^2 = 0.31$). A good correlation between the predicted translation rates and the empirical codon frequencies is expected, since the latter were used in generating the former. However, the poor correlation between predicted and empirical translation rates could reflect the inadequacies in 1) the assumptions of Solomovici *et al.* [39], 2) the rate measurements of Curran and Yarus [46], and/or 3) the codon and tRNA data of Dong *et al.* [41]. Alternatively, the phenotype of *E. coli* may not be perfectly optimized for maximal translation rates [38], either because of genetic drift or because of conflicting selection pressures. Nonetheless, the disparity between the relative rates of synonymous preferred and non-preferred codons for most amino acids are greater with the predicted rates of Scenario 5 than with the observed rates. Hence, Scenario 5 will overestimate the translation rate benefit of codon bias compared to a strict application of the empirical codon-specific translation rates. (None of our scenarios are, in fact, strict applications of the empirical rates. Scenarios 1-4 extend the empirical rates in ways that will increase the estimated benefit of codon bias.) The predicted translation rates for all codons (Table 4.1) are used with Equation 2 to estimate the translation rate benefit of codon bias under Scenario 5.

Chapter 4 Results

As expected, the translation rate benefit of codon bias increases monotonically with growth rate, when calculated by any of the scenarios described in Materials and Methods, using proteome codon frequencies and tRNA abundance data from the range of growth rates reported in reference 41 (data not shown). This increase reflects simply the increasing bias in both proteome codon usage and relative tRNA abundance with increasing growth rate. Since we are interested in the maximum effect of codon bias, we report results from only the highest growth rate for which data are available, 2.5 doublings per hour.

We define s_{bias}, the translation rate benefit of codon bias in *E. coli,* as the fractional increase in the time required to replicate the *E. coli* proteome if the actual codon bias in *E. coli* were replaced with uniform use of synonymous codons (Equation 2). Estimates of s_{bias} in *E. coli* according to several scenarios (described in Materials and Methods) are presented in Figure 4.1. All scenarios are based on the data shown in Table 4.1, but use different assumptions for extending incomplete codon-specific relative translation rate data to estimate the benefit of bias over all codons (Scenarios 1-4), or use a theoretical prediction of optimal codon-specific translation rates based on reasonably complete codon frequency and cognate tRNA abundance data (Scenario 5). Since the empirical rate data used in Scenarios 1-4 do not account for translocation, for these scenarios two estimates are shown that make different assumptions regarding the relative duration of translocation and ternary complex selection. The estimated values of s_{bias} range from 0.6 – 1.4 if translocation time is neglected, or from 0.4 – 1.1 with the more realistic

Figure 4.1: Translation rate benefit of codon bias in *E. coli*

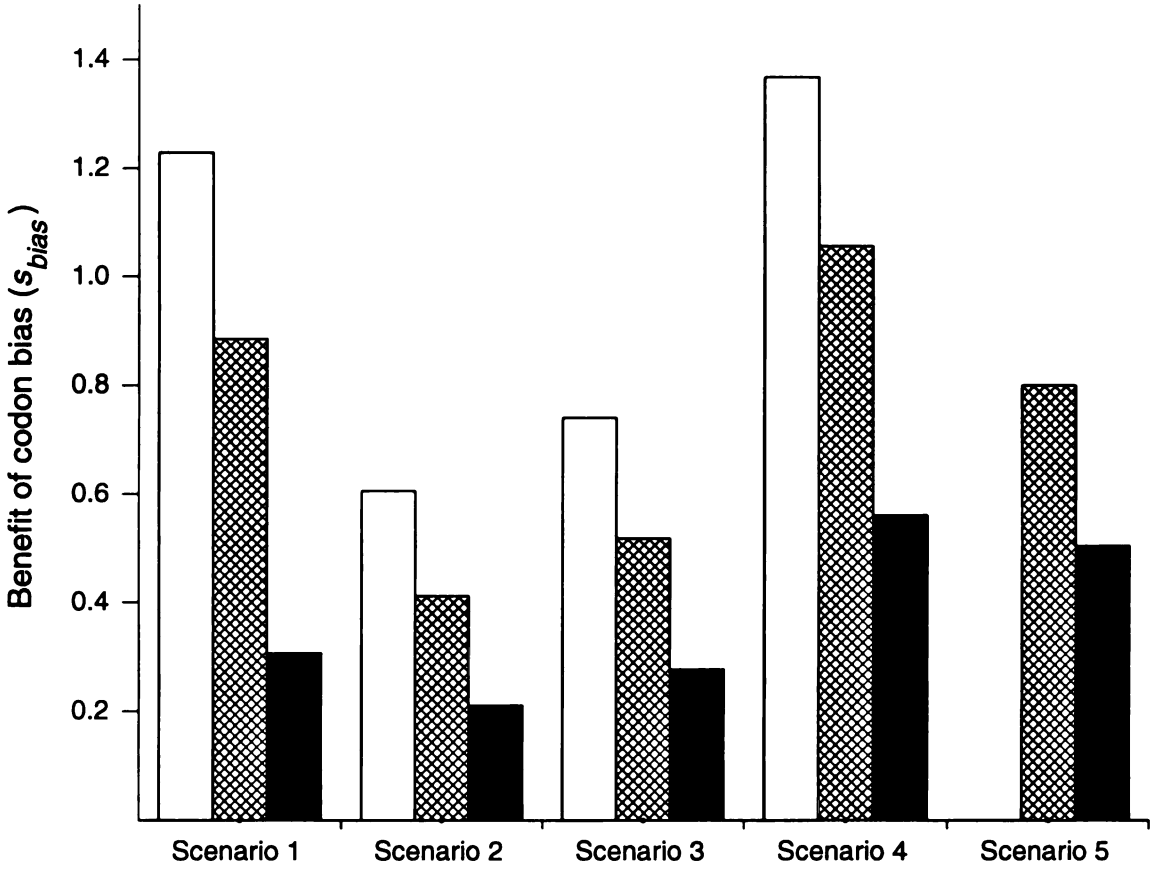


Figure 4.1. The estimated translation rate benefit of codon bias in *E. coli*, using 5 different scenarios to obtain an complete set of relative codon-specific translation rates from incomplete empirical data (see Materials and Methods for explanation). White bars: duration of translocation assumed to be negligible in comparison to the duration of ternary complex selection. Cross-hatched bars: duration of translocation assumed to be constant for all codons and short in comparison to the duration of ternary complex selection. Both white and cross-hatched bars: benefit of codon bias in *E. coli* estimated in comparison to strictly uniform codon use. Black bars: duration of translocation assumed to be constant for all codons and short in comparison to the duration of ternary complex selection; benefit of codon bias in *E. coli* estimated in comparison to degree of codon bias found in an actual low-bias organism. The white, cross-hatched and black bars represent a series of increasingly realistic estimates.

assumption that translocation requires a short amount of time. In a further attempt to introduce biological reality to our calculations, we also estimated s_{bias} for *E. coli,* not in comparison to strictly uniform synonymous codon use, but in comparison to the limited degree of codon bias that might be found in an actual low-bias organism. In this case, the estimated benefits are lower still, ranging from 0.2 – 0.6, again assuming a short duration of translocation. The estimates of s_{bias} from the more theoretically-based Scenario 5 (using codon frequency and tRNA abundance data) fall into the middle of the range of s_{bias} estimates from the more empirically-based Scenarios 1-4 (using codon frequency and codon-specific translation rate data). Hence, we are confident that our conclusions are not overly sensitive to specific details of the empirical translation rate measurements or of the methods used to extend the empirical measurements to unmeasured codons.

Our definition of s_{bias} can be applied over any subset of codons, in particular, it can be applied to the codons of each amino acid separately. While all amino acids with multiple codons except proline contribute positively to s_{bias} in all scenarios, the magnitude of that contribution is highly variable between amino acids (Figure 4.2). Codon bias accelerates the translation of most amino acids only slightly in *E. coli*, because most non-preferred codons are not particularly rare in the *E. coli* proteome, compared to the preferred synonym. For example, among the 9 amino acids encoded by 2 codons, there is an average 2.9-fold difference in the frequency of preferred and non-preferred codons; asparagine shows the greatest difference with GAC being 5.2-fold more abundant than GAU. Even if the disparity in translation rates is unrealistically large, the ratio of preferred to non-preferred codons in *E. coli* constrains the maximum possible value of

s_{bias}. For asparagine, even if the preferred codon were translated instantaneously (i.e., infinitely faster than the non-preferred codon), the difference between 50% usage of the slowly translated codon in Uni compared to 16% usage in *E. coli* corresponds to only about a 3-fold acceleration of translation ($s_{bias} \approx 2$) for this amino acid. With more realistic disparities in the translation rates of the preferred and non-preferred codons, the largest estimate of s_{bias} for asparagine in any of our scenarios is less than 0.2, i.e., codon bias leads to a 20% increase in translation rate over all asparagine codons (Figure 4.2).

The amino acids making the largest contribution to s_{bias} are leucine, isoleucine, and arginine (Figure 4.2). Although these amino acids are not rare, they possess between them the six rarest codons in *E. coli*, each encoding less than 0.1% of the proteome. The disparities in frequency between the most and least abundant synonym for leucine, isoleucine and arginine are 74-fold, 83-fold, and 1460-fold, respectively. (The ratio for arginine reflects the extreme rarity of AGG, which is 17-fold less abundant than the second rarest *E. coli* codon, AUA encoding isoleucine.) Since the translation rates measured or assumed for these rare codons are quite slow, their increased abundance in Uni accounts for the much of the additional time required for replicating the Uni proteome. If these six codons were as rare in Uni as they are in *E. coli,* while all other synonymous codons were used without bias, the translation rate benefit estimated under Scenario 4 (the scenario producing the largest benefit estimates) would be reduced by almost half (data not shown). The influence of these 6 codons is such that the estimate of s_{bias} is quite sensitive to the translation rates assigned to them, in contrast to the relative insensitivity of s_{bias} to the exact translation rates assigned to most codons.

Figure 4.2: Translation rate benefit of codon bias by amino acid

Figure 4.2. The translation rate benefit of codon bias in *E. coli,* estimated separately for each amino acid, is shown by the ordinate; the frequency of the amino acid in the *E. coli* proteome is shown by the abscissa. Each amino acid is represented by its one-letter abbreviation. Panels a-e represent Scenarios 1-5 respectively, described in Materials and Methods. The estimate of s_{bias} is made in comparison to uniform use of synonymous codons, and assumes that the duration of translocation is negligible (corresponding to the first set of bars in Figure 4.1). Only a few amino acids, those encoded by one or more rare codons, contribute disproportionately to the total translation rate benefit of codon bias.

Figure 4.2: Translation rate benefit of codon bias by amino acid

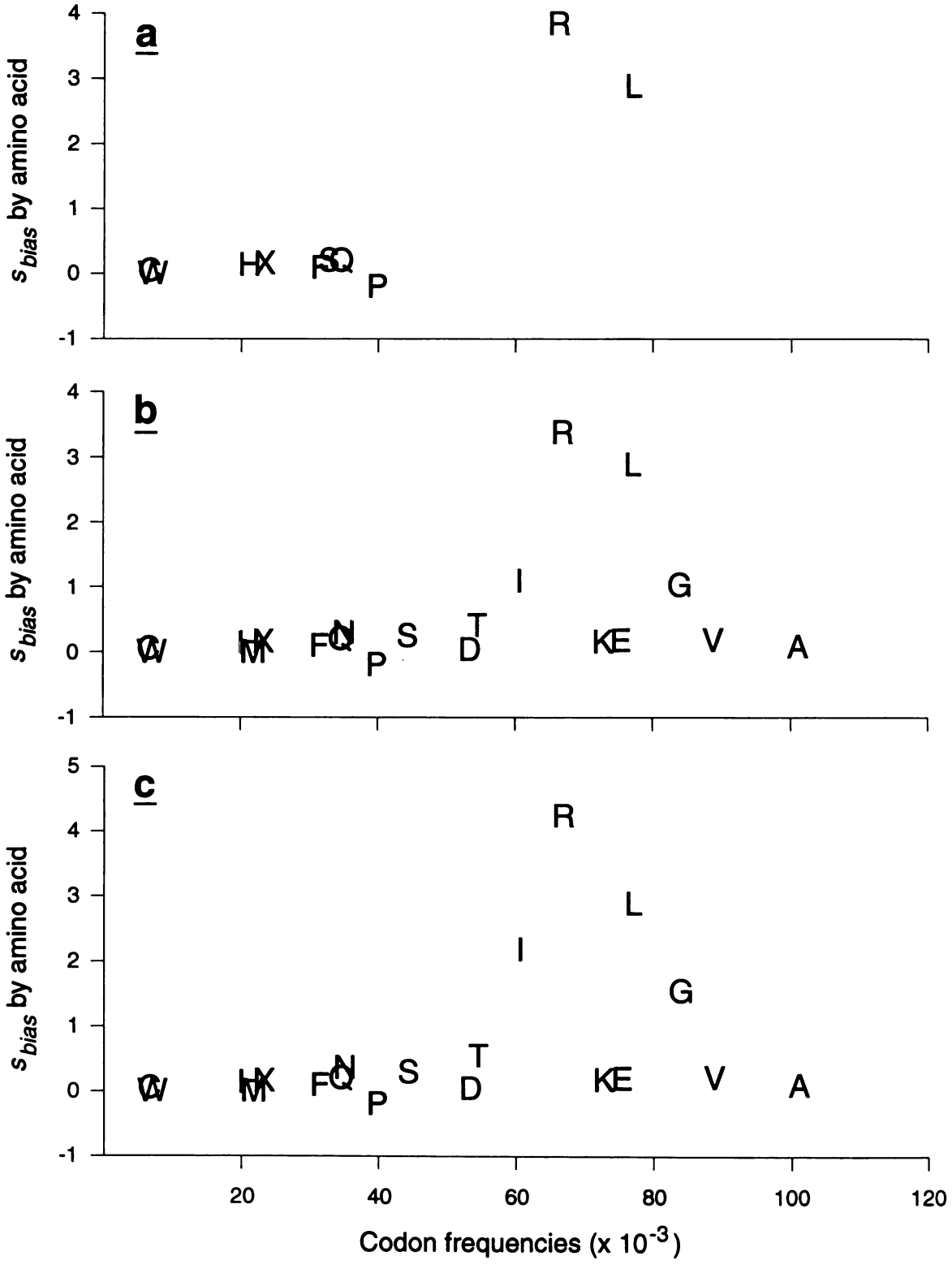
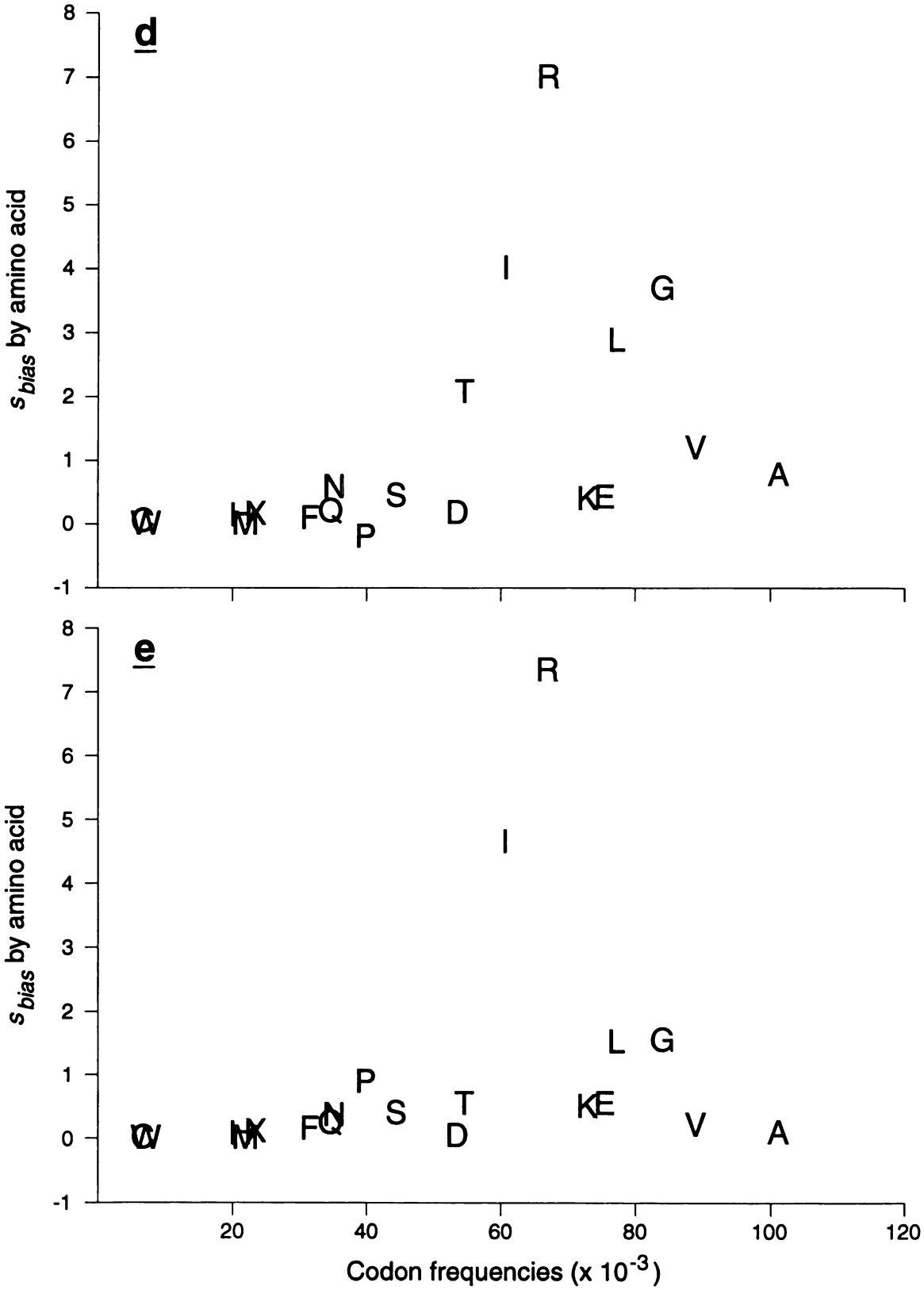


Figure 4.2: continued



Chapter 4 Discussion

We want to know whether reduced codon bias could account for the lower translational power measured in at least some slowly growing bacteria, compared to *E. coli.* We approach this issue by its converse, calculating the acceleration of the translation rate due to codon bias in *E. coli.* If we take our estimates at face value, we would conclude that even during rapid growth when the proteome is most biased, s_{bias} is unlikely to be as large as 1, corresponding to a doubling of the average translation rate. However, there are two reasons to think that the benefit of codon bias for *E. coli,* in comparison to most actual slow-growing organisms, is even less than this.

The first reason is that we have prevented our hypothetical Uni from adapting to the codon frequencies we have assigned to it, by keeping the abundance of each component of the translational apparatus fixed. The data do not suggest that maximizing translational power has been the only selective pressure influencing codon use in *E. coli* [38,46]. If it had been, the codon with the highest rate constant for ternary complex selection among synonymous alternatives would always be the preferred codon, since it would permit faster translation with a lower investment in cognate tRNA. Of 10 amino acids with multiple codons for which codon-specific translation rate measurements exist [46,58], leucine, serine and proline are not consistent with this prediction. On the other hand, it seems clear that selection for rapid translation has exerted some, and perhaps the major influence on the evolution of codon frequencies and tRNA abundance in *E. coli.* The codon with the highest rate constant *is* the preferred codon for 7 of the 10 amino acids for which data are available. Other considerations (possibly including error

avoidance [17], interactions between adjacent tRNA anticodons [59], or factors unrelated to translation [60]) may have been more influential than the inherent characteristics of the codon-anticodon interactions for determining the preferred codons encoding leucine, serine and proline. However, the importance of rapid translation remains evident in that *E. coli* still translates the preferred codons quickly for 2 of these 3 amino acids, albeit with a larger investment in tRNA than would be necessary if the interaction between the preferred codon and its cognate tRNA occurred more readily.

At a larger scale, the correlation across all codons between frequency and cognate tRNA abundance [41,44] is best explained as a response to selection for rapid translation, as is the pattern of increased bias towards rapidly translated codons with increased gene expression [46]. Without asserting that the distribution of tRNA abundance in *E. coli* necessarily produces the fastest possible translation rate for the *E. coli* codon frequency distribution, it is clear that selection for translational power has been a significant factor in the *co*evolution of codon frequencies and cognate tRNA abundances in *E. coli*. Thus, it is very unlikely that the distribution of tRNA abundance values in *E. coli*, and the resulting codon-specific translation rates, produces the fastest possible translation rate when matched with different codon frequencies in Uni. For this reason, our estimates confound the translation rate benefit of codon bias in *E. coli* with the penalty of a suboptimal allocation of translational resources in Uni.

The second reason that the benefit of codon bias in *E. coli* is unlikely to be as large as our estimates, when the comparison is made to actual slow-growing organisms, is that actual

microbes are not completely devoid of codon bias. Assessing s_{bias} in *E. coli* in comparison to a biologically plausible standard for low codon bias, instead of the implausible standard of no codon bias whatsoever, reduces the estimated benefit in *E. coli* by about half (Figure 4.1). The low bias codon frequencies derived from *T. pallidum* can accelerate translation about half as much as the more highly biased *E. coli* codon frequencies because only a few codons in *E. coli* are translated much more slowly than the median rate. Moderate avoidance of these few codons can provide a considerable acceleration of the average translation rate without generating a dramatic bias in overall codon use.

Our estimate of biologically plausible low-bias codon frequencies is deliberately conservative, underestimating the degree of bias expected in most slowly growing microbes, for two reasons. First, we used the genome codon frequencies of *T. pallidum*, as if all predicted genes in the genome were expressed equally. Although analysis of codon use at the level of individual genes failed to uncover evidence of translational selection in *T. pallidum* [28], calculating codon frequencies over a set of putative high expression genes shows that codon use in such genes is, indeed, more biased than codon use in the genome as a whole. This conclusion is based on a comparison of Wright's effective number of codons (Nc) [40] for codon frequencies summed over ribosomal proteins and translation elongation factors (Nc = 52.7) or for codon frequencies summed over all predicted genes in the genome (Nc = 55.2). The failure to observe this low level of codon bias in the earlier analysis [28] can probably be attributed to high variability in estimates of codon frequencies derived from small samples of codons in each gene.

Thus, we believe that even for *T. pallidum,* proteome codon frequencies will be more biased than the low-bias codon frequencies derived from the *T. pallidum* genome that were used to generate the estimates shown in Figure 4.1. Second, the choice of *T. pallidum* to represent slow-growing strains is conservative, because its codon bias is essentially free of the influence of mutational bias, with a genome G+C content of 52.7%. In contrast, many slow-growing microbes have more extensive codon bias that can be attributed mostly or entirely to biased nucleotide composition (e.g., *R. prowazekii* [35], *Helicobacter pylori* [37], *Borrelia burgdorferi* [27,28], *Buchnera aphidicola* [61], *Mycoplasma genitalium* [25], and *Chlamydia* species [29,62]). If codon bias derived from biased nucleotide composition, like codon bias derived from translational selection, permits more rapid translation, the use of low bias codon frequencies derived from *T. pallidum* will underestimate the translation rate of many slow growing strains.

The advantage of codon bias depends to a large extent on matching the preferred codons with abundant cognate tRNAs. Even if codon use is determined by mutational bias in the DNA replication and repair systems [63], not by selection acting simultaneously on individual codons and their cognate tRNAs via translation-associated effects, selection for translational power can influence the relative abundance of tRNA species. Relatively few mutations may suffice to influence the relative tRNA abundance in an organism, in comparison to the number of mutations required to influence proteome codon frequencies. (Consider that ~40 mutations could allow a single mutation in the regulatory region of many or even all tRNA genes, depending on the organism, but could alter the identity of less than 0.5% of the >9,000 codons in genes encoding ribosomal proteins and

translational elongation factors.) Hence, the mutation-selection balance argument invoked to explain diminished codon bias in genes expressed at low levels in many strains [15,24] also suggests that the tRNA abundance distribution can be influenced by translational selection that may be too weak to create a dramatic effect on codon usage. In fact, if codon use is highly biased in the same direction in all genes (as expected if the source of codon bias is mutational bias), instead of being biased only in highly expressed genes, it increases the selective pressure for adaptation of the tRNA pool. Hence, it would be very surprising if the anticodon sequences and the tRNA abundance distribution in organisms with high or low G+C content did *not* reflect their biased use of codons. This prediction is confirmed by the only two studies of tRNA abundance in microbes with extreme G+C content, involving *Mycoplasma capricolum* (25% G+C) [64] and *Micrococcus luteus* (74% G+C) [65]. *M. capricolum,* but not *M. luteus,* can be considered a constitutively slow-growing strain. As expected, cognate tRNA abundance in both organisms is correlated with codon frequency, both across all codons and within synonymous codon families. For *M. capricolum,* this is accomplished largely without the tRNA gene dosage effects that are important for *E. coli* [41] and *B. subtilis* [66], since 28 of the 29 *M. capricolum* tRNA genes are present in only a single copy [64]. These examples indicate that selection for translational power is operative even for organisms in which the codon bias is determined by mutational bias instead of translational selection, and even if the organisms are slow growers. Because codon bias from any source can be exploited to obtain higher translational power, the estimates of s_{bias} for *E. coli* compared to codon frequencies derived from *T. pallidum* will overstate the benefit expected in comparison to most other slowly growing microbes.

In summary, we believe the translation rate benefit of codon bias in *E. coli* is likely to be less than 0.6 when the comparison is made to an actual slow-growing organism that shows limited codon bias, such as *T. pallidum,* and substantially less than 0.6 in comparison to a slow-growing organism with more extensive codon bias, regardless of its source. We do not mean to suggest that an increase of up to 60% in the overall translation rate is unimportant. Clearly, the aggregate benefit of codon bias must be substantial, considering that many thousands of preferred codons are stably maintained in the *E. coli* genome, despite the randomizing influence of mutation acting at each individual codon.

On the other hand, the influence of codon bias on the average translation rate is far smaller than the differences in translational power observed between microbes pursuing different ecological strategies. In order to explain the difference in translational power between rapidly and slowly responding bacteria growing in R2BV medium (Chapter 3), s_{bias} would have to be ~1.4, and in order to explain the comparison of translational power made at similar levels of RNA investment, s_{bias} would have to be ~2.6. The latter value also corresponds to about the middle of the range of comparisons of translational power between *E. coli* and slowly growing microbes presented in Chapter 1. Our estimate of the maximum value of s_{bias} in *E. coli* is less than half of the lowest of these estimates of translational power differences. Hence, the parsimonious explanation that known differences in the degree of codon bias in the proteome account for observed differences in translational power between microbes of contrasting ecological strategies is not

plausible. Instead, the macromolecular data that demonstrate differences in translational power must be interpreted as evidence for some more fundamental difference in translational performance. The hypothesis proposed in Chapter 3, that there is an evolutionary tradeoff between translational power and translational yield mediated by processivity errors, is one example of such a fundamental difference.

Chapter 4 References

1. Kjeldgaard, N.O. and C.G. Kurland, 1963. *The distribution of soluble and ribosomal RNA as a function of growth rate.* Journal of Molecular Biology, **6**:341-348.

2. Maaløe, O., 1979. *Regulation of the protein-synthesizing machinery - ribosomes, tRNA, factors, and so on*; pp. 487-542 in *Gene Expression*; R.F. Goldberger, ed. Plenum Press, New York.

3. Ingraham, J.L., O. Maaløe and F.C. Neidhardt, 1983. *Growth of the Bacterial Cell.* Sinauer, Sunderland, Massachusetts.

4. Ehrenberg, M. and C.G. Kurland, 1984. *Costs of accuracy determined by a maximal growth rate constraint.* Quarterly Reviews of Biophysics, **17**(1):45-82.

5. Koch, A.L. and C.S. Deppe, 1971. *In vivo assay of protein synthesizing capacity of Escherichia coli from slowly growing chemostat cultures.* Journal of Molecular Biology, **55**:549-562.

6. Bremer, H. and P.P. Dennis, 1996. *Modulation of chemical composition and other parameters of the cell by growth rate*; pp. 1553-1569 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

7. Wada, A., K. Igarashi, S. Yoshimura, S. Aimoto and A. Ishihama, 1995. *Ribosome modulation factor - stationary growth phase-specific inhibitor of ribosome functions from Escherichia coli.* Biochemical and Biophysical Research Communications, **214**(2):410-417.

8. Chant, J., I. Hui, D. Dejongwong, L. Shimmin and P.P. Dennis, 1986. *The protein synthesizing machinery of the Archaebacterium Halobacterium cutirubrum - molecular characterization.* Systematic and Applied Microbiology, **7**(1):106-114.

9. Pang, H.L. and H.H. Winkler, 1994. *The concentrations of stable RNA and ribosomes in Rickettsia prowazekii.* Molecular Microbiology, **12**(1):115-120.

10. Fegatella, F., J. Lim, S. Kjelleberg and R. Cavicchioli, 1998. *Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium Sphingomonas sp. strain RB2256.* Applied and Environmental Microbiology, **64**(11):4433-4438.

11. Shahab, N., F. Flett, S.G. Oliver and P.R. Butler, 1996. *Growth rate control of protein and nucleic acid content in Streptomyces coelicolor A3(2) and Escherichia coli B/r.* Microbiology-UK, **142**:1927-1935.

12. Sharp, P.M. and W.H. Li, 1986. *An evolutionary perspective on synonymous codon usage in unicellular organisms.* Journal of Molecular Evolution, **24**(1-2):28-38.

13. Sharp, P.M., E. Cowe, D.G. Higgins, D.C. Shields, K.H. Wolfe and F. Wright, 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 Research, **16**(17):8207-8211.

14. Akashi, H. and A. Eyre-Walker, 1998. *Translational selection and molecular evolution.* Current Opinion in Genetics & Development, **8**(6):688-693.

15. Sharp, P.M., M. Stenico, J.F. Peden and A.T. Lloyd, 1993. *Codon usage - mutational bias, translational selection, or both?* Biochemical Society Transactions, **21**(4):835-841.

16. Andersson, S.G.E. and C.G. Kurland, 1990. *Codon preferences in free-living microorganisms.* Microbiological Reviews, **54**(2):198-210.

17. Parker, J., T. Johnston, P. Borgia, G. Holtz, E. Remaut and W. Fiers, 1983. *Codon usage and mistranslation - in vivo basal level misreading of the MS2 coat protein message.* Journal of Biological Chemistry, **258**(16):10007-10012.

18. Yanofsky, C., 1987. *Operon-specific control by transcription attenuation.* Trends in Genetics, **3**(12):356-360.

19. Takano, E., M. Tao, F. Long, M.J. Bibb, L. Wang, W. Li, M.J. Buttner, Z.X. Deng and K.F. Chater, 2003. *A rare leucine codon in adpA is implicated in the morphological defect of bldA mutants of Streptomyces coelicolor.* Molecular Microbiology, **50**(2):475-486.

20. Dobrindt, U. and J. Hacker, 2001. *Regulation of the tRNA(5)(Leu)-encoding gene leuX that is associated with a pathogenicity island in the uropathogenic Escherichia coli strain 536.* Molecular Genetics and Genomics, **265**(5):895-904.

21. Sharp, P.M. and W.H. Li, 1987. *The rate of synonymous substitution in Enterobacterial genes is inversely related to codon usage bias.* Molecular Biology and Evolution, **4**(3):222-230.

22. Sharp, P.M. and W.H. Li, 1986. *Codon usage in regulatory genes in Escherichia coli does not reflect selection for rare codons.* Nucleic Acids Research, **14**(19):7737-7749.

23. Kurland, C.G., 1993. *Major codon preference - theme and variations.* Biochemical Society Transactions, **21**(4):841-846.

24. Bulmer, M., 1991. *The Selection-Mutation-Drift theory of synonymous codon usage.* Genetics, **129**:897-907.

25. Kerr, A.R.W., J.F. Peden and P.M. Sharp, 1997. *Systematic base composition variation around the genome of Mycoplasma genitalium, but not Mycoplasma pneumoniae.* Molecular Microbiology, **25**(6):1177-1179.

26. Frank, A.C. and J.R. Lobry, 1999. *Asymmetric substitution patterns: A review of possible underlying mutational or selective mechanisms.* Gene, **238**(1):65-77.

27. McInerney, J.O., 1998. *Replicational and transcriptional selection on codon usage in Borrelia burgdorferi.* Proceedings of the National Academy of Sciences of the United States of America, **95**(18):10698-10703.

28. Lafay, B., A.T. Lloyd, M.J. McLean, K.M. Devine, P.M. Sharp and K.H. Wolfe, 1999. *Proteome composition and codon usage in spirochaetes: Species-specific and DNA strand-specific mutational biases.* Nucleic Acids Research, **27**(7):1642-1649.

29. Romero, H., A. Zavala and H. Musto, 2000. *Codon usage in Chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces.* Nucleic Acids Research, **28**(10):2084-2090.

30. Andersson, S.G.E. and P.M. Sharp, 1996. *Codon usage in the Mycobacterium tuberculosis complex.* Microbiology-UK, **142**:915-925.

31. Shields, D.C. and P.M. Sharp, 1987. *Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases.* Nucleic Acids Research, **15**(19):8023-8040.

32. Wright, F. and M.J. Bibb, 1992. *Codon usage in the G+C-rich Streptomyces genome.* Gene, **113**(1):55-65.

33. Ohama, T., A. Muto and S. Osawa, 1990. *Role of GC-biased mutation pressure on synonymous codon choice in Micrococcus luteus, a bacterium with a high genomic GC content.* Nucleic Acids Research, **18**(6):1565-1569.

34. Ohkubo, S., A. Muto, Y. Kawauchi, F. Yamao and S. Osawa, 1987. *The ribosomal protein gene cluster of Mycoplasma capricolum.* Molecular & General Genetics, **210**(2):314-322.

35. Andersson, S.G.E. and P.M. Sharp, 1996. *Codon usage and base composition in Rickettsia prowazekii.* Journal of Molecular Evolution, **42**(5):525-536.

36. Shields, D.C., 1990. *Switches in species-specific codon preferences - the influence of mutation biases.* Journal of Molecular Evolution, **31**(2):71-80.

37. Lafay, B., J.C. Atherton and P.M. Sharp, 2000. *Absence of translationally selected synonymous codon usage bias in Helicobacter pylori.* Microbiology-UK, **146**:851-860.

38. Berg, O.G. and C.G. Kurland, 1997. *Growth rate-optimised tRNA abundance and codon usage.* Journal of Molecular Biology, **270**(4):544-550.

39. Solomovici, J., T. Lesnik and C. Reiss, 1997. *Does Escherichia coli optimize the economics of the translation process?* Journal of Theoretical Biology, **185**(4):511-521.

40. Wright, F., 1990. *The effective number of codons used in a gene.* Gene, **87**(1):23-29.

41. Dong, H.J., L. Nilsson and C.G. Kurland, 1996. *Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates.* Journal of Molecular Biology, **260**(5):649-663.

42. Vanbogelen, R.A., P. Sankar, R.L. Clark, J.A. Bogan and F.C. Neidhardt, 1992. *The gene-protein database of Escherichia coli - edition 5.* Electrophoresis, **13**(12):1014-1054.

43. Pedersen, S., P.L. Bloch, S. Reeh and F.C. Neidhardt, 1978. *Patterns of protein synthesis in Escherichia coli - catalog of amount of 140 individual proteins at different growth rates.* Cell, **14**(1):179-190.

44. 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 Escherichia coli translational system.* Journal of Molecular Biology, **151**(3):389-409.

45. Björk, G.R., 1996. *Stable RNA modification*; pp. 861-886 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*; F.C. Neidhardt, *et al.*, eds. ASM Press, Washington, D. C.

46. Curran, J.F. and M. Yarus, 1989. *Rates of aminoacyl-tRNA selection at 29 sense codons in vivo.* Journal of Molecular Biology, **209**(1):65-77.

47. Fraser, C.M., S.J. Norris, G.M. Weinstock, O. White, G.G. Sutton, R. Dodson, M. Gwinn, E.K. Hickey, R. Clayton, K.A. Ketchum, E. Sodergren, J.M. Hardham, M.P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J.K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M.D. Cotton, C. Fujii, S. Garland, B. Hatch, K. Horst, K. Roberts, M. Sandusky, J. Weidman, H.O. Smith and J.C. Venter, 1998. *Complete genome sequence of Treponema pallidum, the syphilis spirochete.* Science, **281**(5375):375-388.

48. Curran, J.F., 1995. *Decoding with the A-I wobble pair is inefficient.* Nucleic Acids Research, **23**(4):683-688.

49. Sorensen, M.A. and S. Pedersen, 1991. *Absolute in vivo translation rates of individual codons in Escherichia coli - the 2 glutamic acid codons GAA and GAG are translated with a threefold difference in rate.* Journal of Molecular Biology, **222**(2):265-280.

50. Bilgin, N., L.A. Kirsebom, M. Ehrenberg and C.G. Kurland, 1988. *Mutations in ribosomal proteins L7/L12 perturb EF-G and EF-Tu functions.* Biochimie, **70**(5):611-618.

51. Schillingbartetzko, S., A. Bartetzko and K.H. Nierhaus, 1992. *Kinetic and thermodynamic parameters for transfer RNA binding to the ribosome and for the translocation reaction.* Journal of Biological Chemistry, **267**(7):4703-4712.

52. Thanaraj, T.A. and P. Argos, 1996. *Ribosome-mediated translational pause and protein domain organization.* Protein Science, **5**(8):1594-1612.

53. Cortazzo, P., C. Cervenansky, M. Marin, C. Reiss, R. Ehrlich and A. Deana, 2002. *Silent mutations affect in vivo protein folding in Escherichia coli.* Biochemical and Biophysical Research Communications, **293**(1):537-541.

54. Komar, A.A., T. Lesnik and C. Reiss, 1999. *Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation.* FEBS Letters, **462**(3):387-391.

55. von Heinje, G., 1979. *The concentration dependence of the error frequencies and some related quantities in protein synthesis.* Journal of Theoretical Biology, **78**:113-120.

56. Garel, J.P., 1974. *Functional adaptation of tRNA population.* Journal of Theoretical Biology, **43**:211-225.

57. Xia, X., 1998. *How Optimized Is the Translational Machinery in Escherichia coli, Salmonella typhimurium and Saccharomyces cerevisiae?* Genetics, **149**:37-44.

58. Sorensen, M.A., C.G. Kurland and S. Pedersen, 1989. *Codon usage determines translation rate in Escherichia coli.* Journal of Molecular Biology, **207**(2):365-377.

59. Smith, D. and M. Yarus, 1989. *Transfer RNA-transfer RNA interactions within cellular ribosomes.* Proceedings of the National Academy of Sciences of the United States of America, **86**(12):4397-4401.

60. Antezana, M.A. and M. Kreitman, 1999. *The nonrandom location of synonymous codons suggests that reading frame-independent forces have patterned codon preferences.* Journal of Molecular Evolution, **49**(1):36-43.

61. Wernegreen, J.J. and N.A. Moran, 1999. *Evidence for genetic drift in endosymbionts (Buchnera): Analyses of protein-coding genes.* Molecular Biology and Evolution, **16**(1):83-97.

62. Dalevi, D.A., N. Eriksen, K. Eriksson and S.G.E. Andersson, 2002. *Measuring genome divergence in bacteria: A case study using Chlamydian data.* Journal of Molecular Evolution, **55**(1):24-36.

63. Sueoka, N., 1988. *Directional mutation pressure and neutral molecular evolution.* Proceedings of the National Academy of Sciences of the United States of America, **85**(8):2653-2657.

64. Yamao, F., Y. Andachi, A. Muto, T. Ikemura and S. Osawa, 1991. *Levels of transfer RNAs in bacterial cells as affected by amino acid usage in proteins.* Nucleic Acids Research, **19**(22):6119-6122.

65. Kano, A., Y. Andachi, T. Ohama and S. Osawa, 1991. *Novel anticodon composition of transfer RNAs in Micrococcus luteus, a bacterium with a high genomic G+C content - correlation with codon usage.* Journal of Molecular Biology, **221**(2):387-401.

66. Kanaya, S., Y. Yamada, Y. Kudo and T. Ikemura, 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**(1):143-155.

Appendix. Phylogeny of bacteria used with log likelihood ratio tests of correlated trait evolution

Full length small subunit ribosomal RNA (16S) gene sequences were obtained either from the Comprehensive Microbial Resource [1] of The Institute for Genomic Research (www.tigr.org) or from GenBank (www.ncbi.nlm.nih.gov) for 78 completely sequenced bacterial genomes (Table A.1); the operon labeled 'A' or '1' was chosen from genomes containing multiple operons. Sequences were imported into Arb software [2] and aligned automatically to related full length sequences obtained from release 8.0 of the Ribosomal Database Project [3]; initial alignments were optimized manually with reference to secondary structure information available at the Comparative RNA Website, www.rna.icmb.utexas.edu [4]. Additional near full length sequences from RDP release 8.0 were included in the phylogenetic analysis to ensure that all deeply branching bacterial lineages were represented by at least 3 sequences [5]. A mask was created to exclude positions of ambiguous alignment, leaving 1250 positions for phylogenetic inference using the maximum likelihood algorithm within Arb. The resulting tree containing 166 bacterial and archaeal taxa has been 'pruned' without altering branching order or branch lengths to obtain the phylogenic hypothesis displayed in Figure A.1a, which is rooted with archaeal sequences that are not shown on the tree.

To ensure that poorly resolved details of the bacterial phylogeny did not influence results of the comparative analyses, a second tree was derived in which short internal branches corresponding to 8 or fewer inferred nucleotide changes (in 1250 positions) were

collapsed, depicted in Figure A.1b. The choice of 8 changes as the threshold for collapse was based on a comparison with the assessment in reference 5 of the reliability of maximum likelihood rRNA-based phylogenetic inference for reconstructing ancient events in bacterial evolution. Collapsed branches were assigned the minimal nonzero length since the Continuous software program used for our analysis requires a bifurcating tree; these collapsed lengths represented less than 0.01% of a single nucleotide change between sequences. For both descriptions of bacterial phylogeny, branches of zero length at terminal nodes (i.e., separating strains with identical 16S gene sequences at the 1250 positions analyzed) were changed to a length representing half a nucleotide change between sequences.

Of the 78 bacterial genome sequences depicted on the phylogenetic trees of Figure A.1, codon data was not available for *Mycoplasma pulmonis* and *Ureaplasma urealyticum*, and tRNA gene annotation was not available for *Streptococcus pneumoniae* R6 and *Pseudomonas putida*. Hence, 76 taxa were analyzed for the strength of translational selection either using $\Delta N'_c$, based on codon use (Figure 3.3a), or using tRNA gene copy number (Figure 3.3b).

Table A.1: Bacterial taxa analyzed for translational selection

Agrobacterium tumefaciens C58 Cereon
Agrobacterium tumefaciens C58 Uwash
Aquifex aeolicus VF5
Bacillus halodurans C-125
Bacillus subtilus 168
Borrelia burgdorferi B31
Brucella melitensis 16M
Brucella suis 1330
Buchnera aphidicola Sg
Buchnera APS
Campylobacter jejuni NCTC11168
Caulobacter crescentus CB15
Chlamydia muridarum strain Nigg
Chlamydia pneumoniae AR39
Chlamydia pneumoniae CWL029
Chlamydia pneumoniae J138
Chlamydia trachomatis serovar D
Chlorobium tepidum TLS
Clostridium acetobutylicum ATCC 824
Clostridium perfringens 13
Corynebacterium glutamicum ATCC13032
Deinococcus radiodurans R1
Enterococcus faecalis V583
Escherichia coli K12-MG1655
Escherichia coli O157:H7 EDL933
Escherichia coli O157:H7 VT2-Sakai
Fusobacterium nucleatum ATCC 25586
Haemophilus influenzae KW20
Helicobacter pylori 26695
Helicobacter pylori J99
Lactococcus lactis subsp. *Lactis* IL1403
Listeria innocua CLIP 11262
Listeria monocytogenes EGD-e
Mesorhizobium loti MAF303099
Mycobacterium leprae TN
Mycobacterium tuberculosis CDC1551
Mycobacterium tuberculosis H37Rv
Mycoplasma genitalium G37
Mycoplasma pneumoniae M129
Mycoplasma pulmonis UAB CTIP
Neisseria meningitidis MC58
Neisseria meningitidis serogroup A Z2491
Nostoc sp. PCC 7120
Oceanobacillus iheyensis HTE831
Pasturella multocida PM70
Porphyromonoas gingivalis W83
Pseudomonas aeruginosa PAO1
Pseudomonas putida KT2440
Ralstonia solanacearum GMI1000
Rickettsia conorii Malish 7
Rickettsia prowazekii Madrid E
Salmonella enterica serovar Typhi CT18
Salmonella typhimurium LT2 SGSC1412
Shewanella oneidensis MR-1
Sinorhizobium meliloti 1021
Staphylococcus aureus COL
Staphylococcus aureus Mu50
Staphylococcus aureus MW2
Staphylococcus aureus N315
Streptococcus agalactiae 2603V/R
Streptococcus pneumoniae R6
Streptococcus pneumoniae TIGR4
Streptococcus pyogenes MGAS315
Streptococcus pyogenes MGAS8232
Streptococcus pyogenes SF350 serotype M1
Streptomyces coelicolor A3(2)
Synechocystis sp. PCC6803
Thermoanaerobacter tengcongensis MB4(T)
Thermosynechococcus elongatus BP1
Thermotoga maritima MSB8
Treponema pallidum Nicholas
Ureaplasma urealyticum serovar 3
Vibrio cholerae El Tor N16961
Xanthomonas axonopodis pv citri 306
Xanthomonas campestris pv.campestris
Xylella fastidiosa 9a5c
Yersinia pestis CO92
Yersinia pestis KIM

Table A.1. 78 bacterial genome sequences analyzed for the strength of translational selection, according to two measures. $\Delta N'_c$ could not be assessed for *Mycoplasma pulmonis* and *Ureaplasma urealyticum;* tRNA copy number could not be assessed for *Streptococcus pneumoniae* R6 and *Pseudomonas putida*. (Chapter 3).

Figure A.1: Phylogenetic trees of bacteria

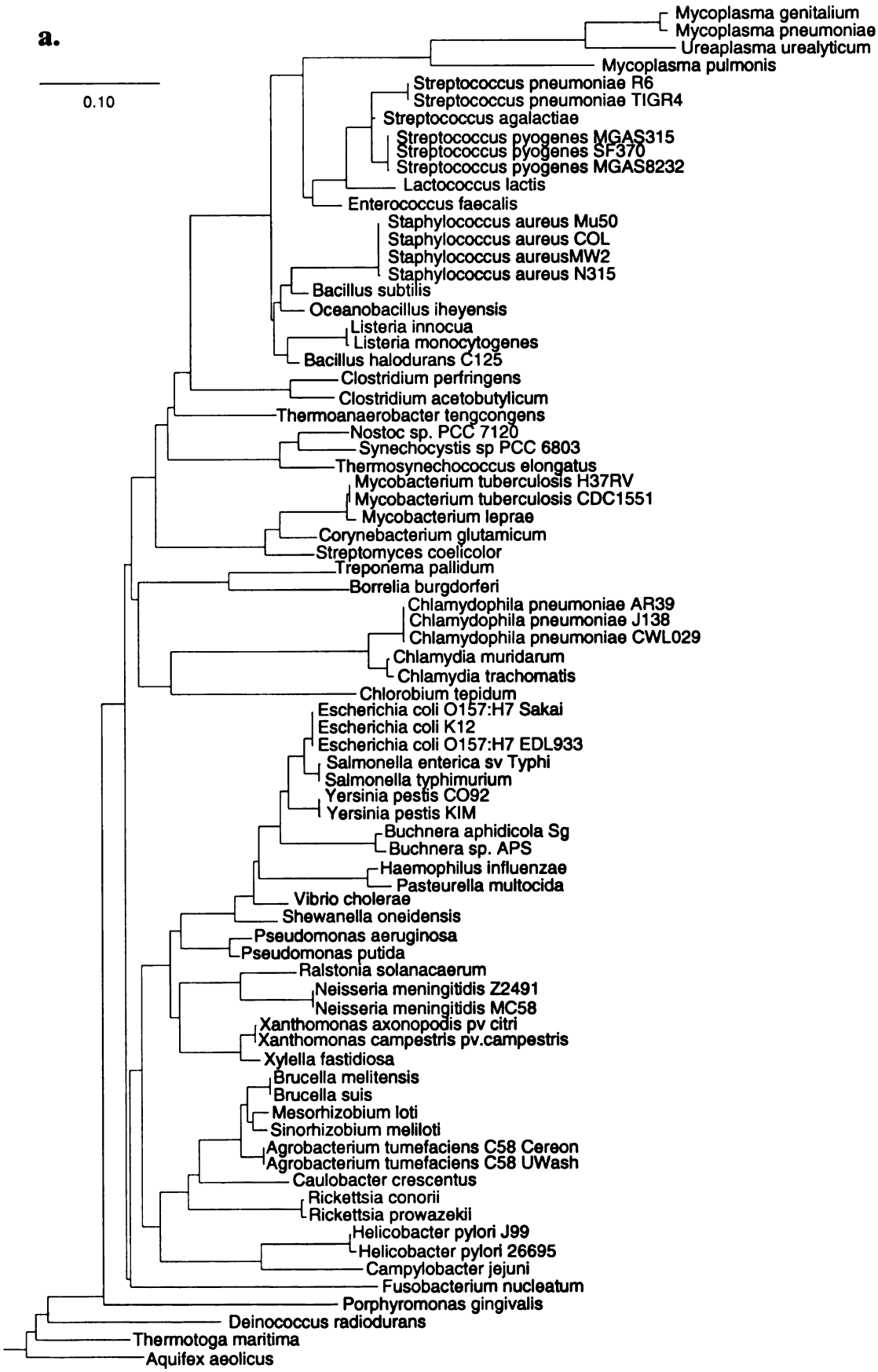
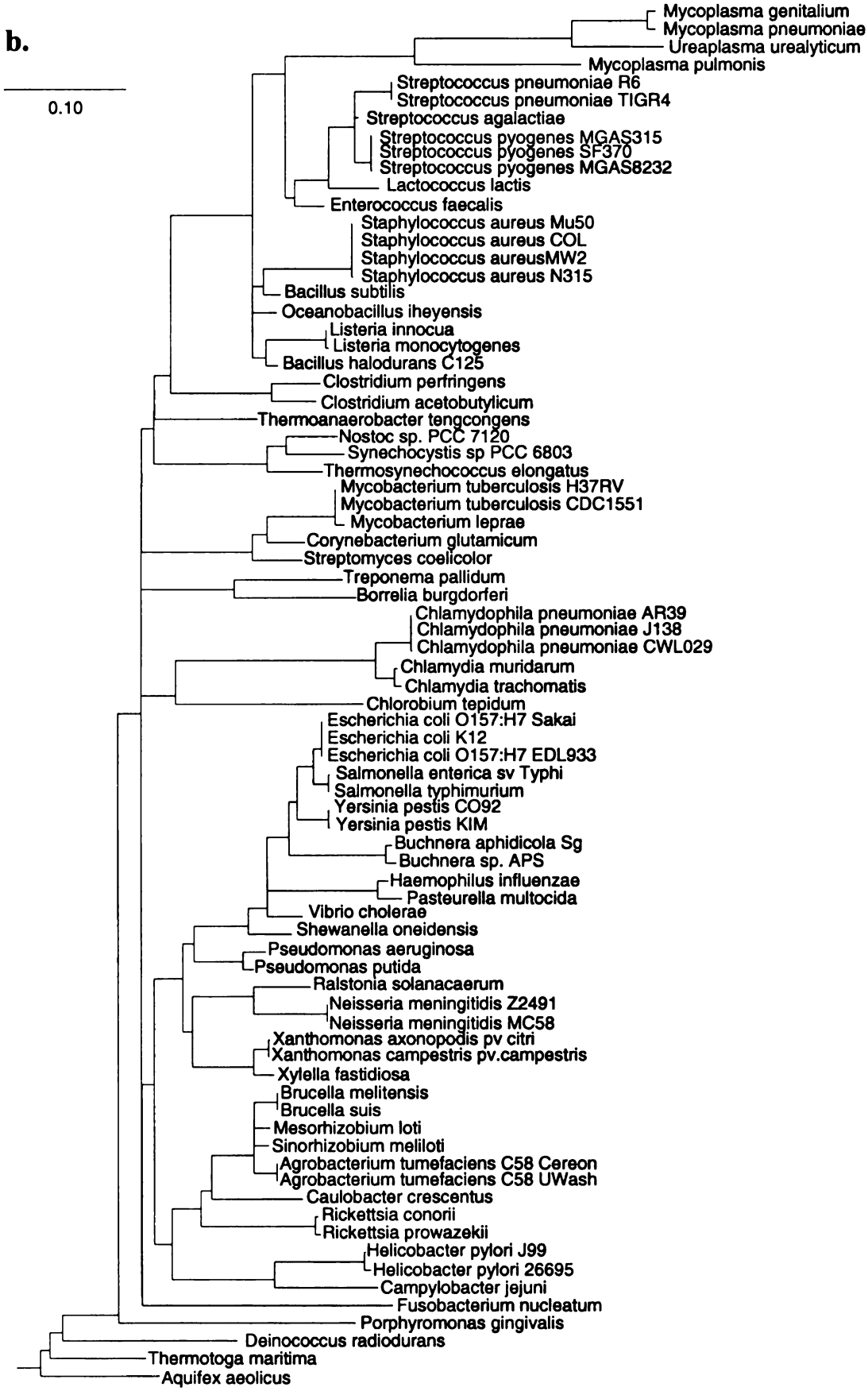


Figure A.1: continued



Appendix References

1. Peterson, J.D., L.A. Umayam, T. Dickinson, E.K. Hickey and O. White, 2001. *The Comprehensive Microbial Resource.* Nucleic Acids Research, **29**(1):123-125.

2. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A.W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode and K.-H. Schleifer, 2004. *ARB: a software environment for sequence data.* Nucl. Acids. Res., **32**(4):1363-1371.

3. Cole, J.R., B. Chai, T.L. Marsh, R.J. Farris, Q. Wang, S.A. Kulam, S. Chandra, D.M. McGarrell, T.M. Schmidt, G.M. Garrity and J.M. Tiedje, 2003. *The Ribosomal Database Project (RDP-II): Previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy.* Nucleic Acids Research, **31**(1):442-443.

4. Cannone, J., S. Subramanian, M. Schnare, J. Collett, L. D'Souza, Y. Du, B. Feng, N. Lin, L. Madabusi, K. Muller, N. Pande, Z. Shang, N. Yu and R. Gutell, 2002. *The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs.* BMC Bioinformatics, **3**(1):2.

5. Ludwig, O. and H.P. Klenk, 2001. *Overview: A phylogenetic backbone and taxonomic framework for prokaryotic systematics*; pp. 49-65 in *Bergey's Manual of Systematic Bacteriology*; D.R. Boone and R.W. Castenholz, eds. Bergey's Manual Trust, East Lansing, Michigan.