J Mol Evol (2003) 57:731–736
DOI: 10.1007/s00239-003-2530-6

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2003

Selection for Highly Biased Amino Acid Frequency in the TolA Cell Envelope Protein of Proteobacteria

Alejandro P. Rooney

Microbial Genomics and Bioprocessing Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, Peoria, IL 61604, USA

Received: 19 March 2002 / Accepted: 21 July 2003

Abstract. The bacterial cell envelope protein TolA functions to maintain the integrity of the cell membrane. This protein contains high levels of alanine and lysine that are used in the formation of alpha helices, which are required for normal protein function. The neutral model of molecular evolution predicts that amino acid composition and nucleotide composition are driven by the underlying GC content, as a result of mutation bias. However, this study shows that selection has acted to maintain high levels of alanine and lysine in the TolA protein of Proteobacteria, which in turn has biased nucleotide composition in the corresponding *tolA* gene.

Key words: TolA — Cell membrane — Selection — Amino acid composition

Introduction

Bacteria produce a range of antimicrobial compounds including some, such as the bacteriocins, that are effective against other bacteria (reviewed by Riley and Wertz 2002). Colicins are a class of bacteriocins produced by and effective against *Escherichia coli* and related gram-negative bacteria. They are divided into two groups on the basis of their killing activities: group A comprises proteins that possess DNase, RNase, or pore-forming activities, while group B comprises proteins that possess protein synthesis inhibition, peptidoglycan synthesis inhibition, or pore-forming activities. Because they provide a means of defense and the ability to reduce competition for the same ecological resources (Chao and Levin 1981; Riley and Gordon 1999), colicins offer a selective advantage to gram-negative bacteria. In addition, these proteins represent potentially valuable resources for the control of microbial contaminants in the agricultural and food science industries. For example, nisin, a bacteriocin that is distantly related to colicins and produced by *Lactobacillus* species, is used as a preservative of dairy products (Ross et al. 2002).

In order to exert their activity, colicins must be taken up by the target cell. The Tol–Pal translocation system mediates the uptake of group A colicins, as well as bacteriophages, across the inner and outer membranes of gram-negative bacteria (Sun and Webster 1987; Levengood and Webster 1989; Heilpern and Waldor 2000). In *E. coli*, the TolA protein is anchored in the inner membrane, spans the periplasmic space, and interacts with porin receptors that span the outer membrane (Fig. 1). TolA possesses three domains: a hydrophobic N-terminal domain that spans the inner membrane, a helix-rich central domain, and a C-terminal domain that interacts with protein receptors embedded in the outer membrane

Correspondence to: Alejandro P. Rooney, U.S. Department of Agriculture, Agricultural Research Service, National Center for Agricultural Utilization Research, 1815 North University Street, Peoria, IL 61604, USA; *email:* rooney@ncaur.usda.gov

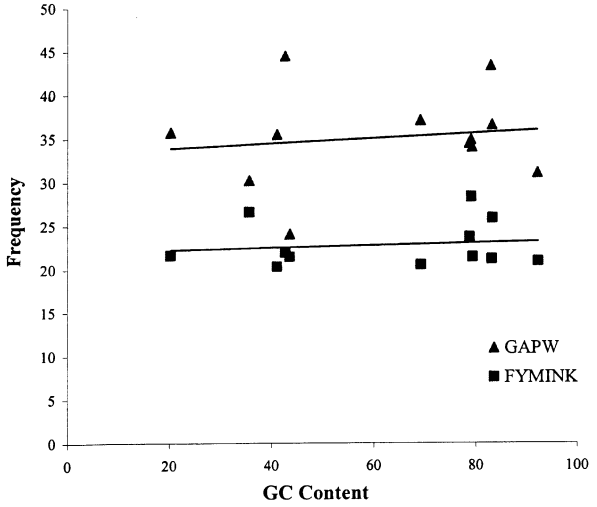

Fig. 1. Relationship between GC content and the frequencies of GC-rich (GAPW) and GC-poor (FYMINK) amino acid classes in TolA.

(Levengood et al. 1991). The helix-rich central domain contains high levels of alanine, lysine, and glutamic acid, and stable alpha-helices are known to form as a result of short stretches of alanine or through salt-bridge formation between lysine and glutamic acid (Marqusee and Baldwin 1987; Marqusee et al. 1989; Levengood et al. 1991). These alpha helices act as coils that give TolA a spring-like quality, allowing the bacterial membrane to shrink and swell in response to osmotic pressure. In this way, TolA functions to help maintain the integrity of the outer membrane (Levengood et al. 1991; Rodríguez-Herva et al. 1996; Lazzaroni et al. 1999; Llamas et al. 2000; Lloubès et al. 2001).

A direct relationship exists between protein sequence conservation and structural conservation resulting in the preservation of protein function (Nei 1987; Li 1997; Nei and Kumar 2000). Thus, protein domains or functional motifs are typically conserved even among highly divergent species. For some proteins, highly divergent species are observed to display highly biased amino acid frequencies at similar levels, suggesting that amino acid content (and not just the specific sequence of amino acids) is selectively important (Rooney et al. 2000). This runs contrary to expectations under the neutral model of molecular evolution, which predicts that amino acid frequencies will be largely determined by GC content, which is believed to evolve neutrally (Sueoka 1961, 1962, 1988; Kimura 1983; Jukes and Bhushan 1986; Nei 1987; Li 1997; Nei and Kumar 2000). In this study, it is shown that biased amino acid composition has shaped the underlying nucleotide composition of the *tolA* gene as a result of selection for high frequencies of functionally important amino acids.

Materials and Methods

Sequences were extracted from published bacterial genomes archived in GenBank. To identify *tolA* homologs, BLASTP (Altschul et al. 1997) searches were conducted separately with the deduced amino acid sequences of the TolA proteins from the *E. coli* K12, *Brucella mellitensis*, *Vibrio cholerae*, and *Pseudomonas aeruginosa* genomes, as these sequences have been previously identified on the basis of both sequence similarity and experimental studies (Levengood et al. 1991; Tibor et al. 1994; Dennis et al. 1996; Rodriguez-Herva et al. 1996; Heilpern and Waldor 2000). A minimum e-value of 5×10^{-4} was used at the criterion to define *tolA* homologs. Once the homologs were identified, the deduced amino acid sequences were aligned using Clustal X (Thompson et al. 1997). A nucleotide sequence alignment was then constructed on the basis of the deduced amino acid sequence alignment.

Nucleotide and amino acid compositions were calculated using the computer program MEGA2 (Kumar et al. 2001). To test for deviations from neutrality, two different approaches were used. In both cases, the GC content at fourfold degenerate sites was assumed to represent the genomic GC content, because the latter has already been shown to be a good approximation of the former (Nei 1987; Li 1997). The GC content at fourfold degenerate sites was also used as an approximation to the neutral expectation. The first test was used to determine if selection influences certain amino acids. For this test, the correlation between the genomic GC content and the proportion of GC-rich (GAPW) and GC-poor (FYMINK) amino acids was computed. Under the neutral model (Sueoka 1961, 1962, 1988; Jukes and Bhushan 1986), GC-rich amino acids will be positively correlated with genomic GC content, whereas GC-poor amino acids will be negatively correlated with genomic GC content. Amino acid frequency and GC content will show no correlation, however, if they are influenced by selection. Correlations were computed using the Spearman rank method and statistical significance was assessed using a standard regression analysis.

A shared, recent phylogenetic history can bias correlation analyses if closely related species are used in the comparisons (Felsenstein 1985). However, this can be avoided by using species that are not closely related, such as those above the genus level (Harvey and Pagel 1991; Gu et al. 1998). Another approach to gauge the relative level of species divergence is to use the magnitude of the number of synonymous substitutions per synonymous site (d_s). In this case, one species from a pair showing $d_s < 1.0$ can be eliminated from correlation analyses. A value of 1.0 was chosen as the cutoff value, because this indicates that all potentially synonymous sites are variable. Note that values of $d_s > 1.0$ are possible as a result of correcting for multiple hits. The modified Nei and Gojobori method (Zhang et al. 1998) with Jukes and Cantor (1969) distances and $R = 0.76$ (the mean of all pairwise R estimates using Kimura [1980] distances) was used to estimate d_s.

The second test was used to gauge the significance of the difference between the nucleotide frequencies at the second codon position versus the frequencies at fourfold degenerate sites. This is a modification of the approach taken by Rooney et al. (2000) involving comparisons between second and third codon positions. Because all changes at second positions are nonsynonymous, whereas all changes at fourfold degenerate sites are synonymous, the contrasting effects of selection and mutation bias at the nucleotide level will be highly apparent in comparisons between these two types of sites. This is because the nucleotide frequencies at second codon positions should not be significantly different than the frequencies at fourfold degenerate sites under a neutral model (Rooney et al. 2000). However, if selection acts to bias the frequency of certain amino acids, the frequencies of nucleotides found at the second codon positions for these amino acids will be elevated with respect to the neutral expectation, represented by those same nucleotide frequencies at

Table 1. Amino acid composition and GC content in the TolA central domain

Species[a]	Fourfold GC	GAPW[b]	FYMINK	Alanine (GCN)[c]	Lysine (AAR)	Glutamic acid (GGR)
α-Proteobacteria						
A. tumefasciens	78.9	34.41	23.66	13.44	12.9	11.83
B. mellitensis	79.5	33.97	21.38	13.84	15.09	13.84
S. meliloti	92.5	31.01	20.89	13.29	14.56	17.72
γ-Proteobacteria						
A. vinelandii	83.5	36.48	25.82	36.48	24.18	20.08
E. chrysanthemi	20.2	35.67	21.62	31.35	17.84	17.84
E. coli O157	69.3	37	20.5	35.5	20.5	16.5
H. influenzae	42.7	44.39	21.94	43.88	20.41	16.33
P. aeruginosa	79.3	34.78	28.25	34.78	26.81	15.22
P. multocida	35.6	30.2	26.57	27.6	21.88	20.31
S. typhimurium	83.3	43.27	21.15	42.31	21.15	14.42
Y. pestis	41.1	35.42	20.31	34.38	18.23	14.06
V. cholerae	43.6	24.03	21.43	24.03	19.48	25.32

[a] The accession numbers for the TolA sequences are *A. tumefasciens* (NC_003062), *B. mellitensis* (AF35866), *S. meliloti* (NC_003047), *A. vinelandii* (NZ_AAAD01000068), *E. chrysanthemi* (AJ297885), *E. coli* O157 (NC_002695), *H. influenzae* (NC_000907), *P. aeruginosa* (NC_002516), *P. multocida* (NC_002663), *S. typhimurium* (NC_003197), *Y. pestis* (NC_003143), and *V. cholerae* (NC_002505).
[b] GAPW—GC-rich codons; FYMINK—GC-poor codons.
[c] Letters in parentheses denote codons for the given amino acid.

fourfold degenerate sites. In order to assess statistical significance, a χ^2 test with 1 df is used.

Results

A total of 19 *tolA* sequences was identified in GenBank searches (excluding identical sequences produced by different strains of the same species, such as *E. coli* K12 and O157:H7). Four of these species were from the α-proteobacteria, while all others were from the γ-proteobacteria. The *tolA* genes of *Shigella flexneri* and *Mesorhizobium loti* gave $d_s < 1.0$ in comparisons with certain taxa (*S. flexneri* and *E. coli*, $d_s = 0.07$; *M. loti* and *S. meliloti*, $d_s = 0.70$; *M. loti* and *A. tumefasciens*, $d_s = 0.83$). When these species were eliminated, values of $d_s > 1.0$ resulted in all pairwise comparisons. Therefore, they were not used in the subsequent correlation analyses. In addition, because *tolA* genes from four species of *Pseudomonas* (*P. aeruginosa*, *P. fluorescens*, *P. putida*, and *P. syringae*) and two species of *Salmonella* (*S. typhi* and *S. typhimurium*) were found, only one was used for each genus. In this way, two different methods were used to exclude closely related taxa. Because the use of different species to represent *Pseudomonas* or *Salmonella* gave highly similar results, only the results from one representative species are reported. The species analyzed are listed in Tables 1 and 2.

The correlations between GC content and the frequency of GC-rich amino acids ($r_S = 0.021$; Fig. 1) and GC-poor amino acids ($r_S = 0.044$; Fig. 1) were not significant. The GC-rich class is composed primarily of alanine, and the GC-poor class is composed primarily of lysine (Table 1). When the correlations between GC content and the frequencies of these amino acids were computed, they were not found to be significant ($r_S = -0.077$ and $r_S = 0.021$ for alanine and lysine, respectively; Fig. 2). Because alanine and lysine are found at very high frequencies and are GC-poor and GC-rich, respectively, one would predict an ensuing bias in nucleotide composition. Codons for alanine possess the C nucleotide at second positions, whereas codons for lysine possess A at second positions. The mean C + A content at second positions (87.9 ± 6.1%) is significantly greater than the mean C + A content (45.0 ± 15.0%) at fourfold degenerate sites ($p = 1.43 \times 10^{-10}$, χ^2 test). In fact, the two are significantly different in all species analyzed (Table 2). In addition, the C + A contents at second positions and fourfold degenerate sites are not significantly correlated ($r_S = 0.026$). These results indicate that nucleotide composition is influenced by selection to maintain high frequencies of alanine and lysine in TolA.

It should be pointed out that glutamic acid is another amino acid found at moderate to high frequencies in TolA (Table 1). Testing for selection on GC-intermediate amino acids through the types of analyses presented here is not practical, because GC-rich and GC-poor codons of such amino acids will cancel each other out by and large in interspecific comparisons, resulting in a lack of correlation regardless of the cause. Therefore, it is difficult to determine whether selection or mutation bias influences the frequency of glutamic acid in TolA. However, it should be noted that glutamic acid codons possess A at the second position, as do lysine codons. So, it is possible that the elevated CA content at second positions is also due to selection for glutamic acid in addition to alanine and lysine.

Interestingly, there are substantial differences in proline and alanine usage between the γ-proteobac-

Table 2. Results of comparisons between the frequencies of C and A nucleotides at coding and noncoding sites within the *tolA* central domain

Species	2nd pos. C + A	Fourfold C + A	χ^2 value	*p* value[a]
α-Proteobacteria				
A. tumefasciens	86.6	47.3	32.65307	1.10×10^{-8}
B. mellitensis	87.4	39.8	56.92864	4.52×10^{-14}
S. meliloti	85.5	50.6	24.07134	9.28×10^{-7}
γ-Proteobacteria				
A. vinelandii	93	59.2	19.29797	1.12×10^{-5}
E. chrysanthemi	91	34	95.55882	1.44×10^{-22}
E. coli O157	91.9	48.9	37.81186	7.79×10^{-10}
H. influenzae	86	43	43	5.47×10^{-11}
P. aeruginosa	92.7	63.7	13.20251	0.0003
P. multocida	84.4	56.2	14.15018	0.0002
S. typhimurium	92.3	22.6	214.9597	1.14×10^{-48}
Y. pestis	87.5	31.7	98.22208	3.74×10^{-23}
V. cholerae	79.2	48.4	19.6	9.55×10^{-6}

[a] All *p* values are significant after application of the Bonferroni correction.

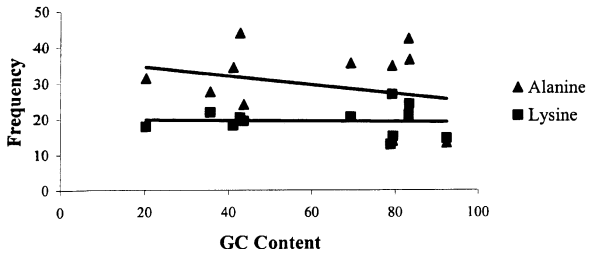

Fig. 2. Relationship between GC content and the frequencies of alanine and lysine in TolA.

teria and α-proteobacteria TolA protein sequences. Among the γ-proteobacteria, the mean frequencies of alanine and proline are 34.5 ± 6.4 and 0.53 ± 1.1%, respectively. In contrast, among the α-proteobacteria the mean frequencies of alanine and proline are 13.7 ± 0.5 and 20.0 ± 2.6%, respectively. It is possible that the above results showing evidence of selection pertain to the γ-proteobacteria but not to the α-proteobacteria, as only three species of the latter were available for this study. Clearly, further studies are required to resolve this question with respect to the α-proteobacteria.

Discussion

Few studies have shown that natural selection is more important than mutation bias in determining the amino acid composition of proteins (Rooney et al. 2000; Akashi and Gojobori 2002; Palacios and Wernegreen 2002). High levels of alanine and lysine are required for the normal functioning of TolA (Levengood and Webster 1991; Lloubès et al. 2001), and the results of this study indicate that their frequencies have been maintained at high levels in γ-proteobacteria due to selection. This finding stands in contrast to the neutral model, which predicts that a protein's amino acid content is influenced by the nucleotide composition of its corresponding gene (Sueoka 1961, 1962, 1988; Kimura 1983; Jukes and Bhushan 1986; Nei 1987; Osawa et al. 1992; Li 1997). Accordingly, if a gene primarily consists of GC nucleotides, the protein that it encodes will consist primarily of amino acids that are encoded by GC-rich codons as a result of mutation bias or some other mechanism apart from selection.

However, the action of natural selection at the protein level could alter nucleotide composition bias if it were strong enough. This was shown to occur in a mammalian sperm DNA-binding protein, protamine-1, in which the composition of arginine is maintained at roughly 50% between species that diverged as much as a few hundred million years ago (Rooney et al. 2000). High levels of arginine are needed for the DNA-binding activity of protamine-1. Therefore, maintenance of this amino acid at high levels can be related to the primary function of the protein, thereby attesting to its importance (Rooney et al. 2000). A similar effect of selection for high frequencies of an amino acid occurs in *tolA*, in which a bias in the C + A nucleotide composition of the gene has resulted due to selection for high levels of alanine and lysine (Tables 1 and 2). The strictly neutral model of molecular evolution predicts that the four nucleotides comprising DNA (adenine, guanine, thymine, and cytosine) should occur with equal frequencies if mutations are random and have an equal probability of being fixed (Sueoka 1961, 1962; King and Jukes 1969; Kimura 1968, 1983; Nei 1987). While the large majority of studies over the past two decades discount selection as the explanation for biased nucleotide composition, certain studies have shown that natural selection could play some part in shaping nucleotide composition (Bernardi and Bernardi 1986; Clark et al. 1999; Bernardi 2000; Wang and Hickey 2002). In that sense, selection for highly biased amino acid

composition will influence nucleotide composition, as shown here in the case of TolA and also by Rooney et al. (2000) in the case of protamine.

One question that remains to be answered regards the driving force behind high frequency amino acid selection on TolA. Interestingly, not all species of Proteobacteria possess *tolA* genes. For example, *tolA* genes were not found in searches of the complete genomes of *Rickettsia prowazeckii* (α-proteobacteria), *R. connori* (α-proteobacteria), *Caulobacter crescentus* (α-proteobacteria), *Neisseria meningtidis* Types A and B (β-proteobacteria), *Buchnera* sp. (γ-proteobacteria), *Campylobacter jejuni* (ε-proteobacteria), and *Helicobacter pylori* (δ/ε-proteobacteria). The reasons why these species lack *tolA* are unclear. However, assuming that an analogous protein has not evolved to take the place of TolA, it may be that the protein provides a selective advantage allowing species to inhabit osmotically unstable environments, as it has been shown to function in maintaining the integrity of the cell membrane (Rodríguez et al. 1996; Lazzaroni et al. 1999; Llamas et al. 2000; Lloubès et al. 2001). In that context, species that lack *tolA* might be restricted to an osmotically stable environment. It is interesting to note that all of the species listed above inhabit such environments, as they are either obligate intracellular pathogens (*Rickettsia*), intracellular endosymbionts (*Buchnera*), inhabitants of aquatic environments (*Caulobacter*), or gastrointestinal pathogens or commensals (*Campylobacter* and *Helicobacter*). Similarly, the α-proteobacteria included in this study are either endosymbionts (*Sinorhizobium*) or intracellular pathogens (*Agrobacterium* and *Brucella*). This might explain the observed differences (Table 1) between these species and the γ-proteobacteria with respect to selection for high levels of alanine, which was not observed to occur in the former. In that regard, it would be interesting to conduct further studies comparing the TolA proteins of organisms from osmotically stable versus unstable environments across various groups of Proteobacteria.

In conclusion, this study shows that selection has occurred for maintaining high levels of functionally important amino acids in TolA and that this has subsequently altered nucleotide composition bias within the *tolA* gene. To date, most studies do not support selection as being important in the determination of amino acid frequency at the genome level (e.g., Karlin et al. 1992; Foster et al. 1997; Gu et al. 1998; Nishizawa and Nishizawa 1998; Singer and Hickey 2000). Therefore, it is reasonable to presume that, in general, amino acid frequency is driven by mutation bias in most proteins. Nevertheless, in cases where the bias in favor of a particular amino acid is exceptional and is maintained over long periods of evolutionary time, selection is likely to be the underlying driving force. It should also be noted that studies aimed at detecting whole-genome patterns overlook deviations on a local scale. Thus, if we focus our attention on local (i.e., individual gene, protein, or genomic region) levels, we may come to find that selection is more important than currently believed in shaping nucleotide and/or amino acid composition, an idea supported by this study and a previous one (Rooney et al. 2000). The question that remains is whether these represent isolated cases of an unusual form of selection or examples of a more widespread phenomenon that is only now being uncovered.

Acknowledgment. Helpful comments were provided by C.P. Kurtzman, D.P. Labeda, T.J. Ward, J. Zhang, and two anonymous reviewers.

References

Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc Natl Acad Sci USA 99:3695–3700

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bernardi G (2000) The compositional evolution of vertebrate genomes. Gene 259:31–43

Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. J Mol Evol 24:1–11

Chao L, Levin BR (1981) Structured habitats and the evolution of anticompetitor toxins in bacteria. Proc Natl Acad Sci USA 78:6324–6328

Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol Biol Evol 16:1586–1598

Dennis JJ, Lafontaine ER, Sokol PA (1996) Identification and characterization of the *tolQRA* genes of *Pseudomonas aeruginosa*. J Bacteriol 178:7059–7068

Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15

Foster PG, Jermiin LS, Hickey DA (1997) Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J Mol Evol 44:282–288

Gu X, Hewett-Emmett D, Li WH (1998) Directional mutational pressure affects the amino acid composition and hydrophobicity of proteins in bacteria. Genetica 102–103:383–391

Harvey PH, Pagel MD (1991) The comparative methods in evolutionary biology. Oxford University Press, Oxford

Heilpern AJ, Waldor MK (2000) CTXϕ infection of *Vibrio cholerae* requires the *tolQRA* gene products. J Bacteriol 182:1739–1747

Jukes TH, Bhushan V (1986) Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. J Mol Evol 24:39–44.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Karlin S, Blaisdell BE, Bucher P (1992) Quantile distributions of amino acid usage in protein classes. Protein Eng 5:729–738

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164:788–798

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Lazzaroni JC, Germon P, Ray MC, Vianney A (1999) The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. FEMS Microbiol Lett 177:191–197

Levengood SK, Webster RE (1989) Nucleotide sequences of the *tolA* and *tolB* genes and localization of their products, components of a multistep translocation system in *Escherichia coli*. J Bacteriol 171:6600–6609

Levengood SK, Beyer WF Jr, Webster RE (1991) TolA: A membrane protein involved in colicin uptake contains an extended helical region. Proc Natl Acad Sci USA 88:5939–5943

Li WH (1997) Molecular evolution. Sinauer, Sunderland, MA

Llamas MA, Ramos JL, Rodríguez-Herva JJ (2000) Mutations in each of the *tol* genes of *Pseudomonas putida* reveal that they are critical for maintenance of outer membrane stability. J Bacteriol 182:4764–4772

Lloubès R, Cascales E, Walburger A, Bouveret E, Lazdunski C, Bernadac A, Journet L (2001) The Tol-Pal proteins of the *Escherichia coli* envelope: An energized system required for outer membrane integrity? Res Microbiol 152:523–529

Marqusee S, Baldwin RL (1987) Helix stabilization by Glu^{-}...Lys^{+} salt bridges in short peptides of de novo design. Proc Natl Acad Sci USA 84:8898–8902

Marqusee S, Robbins VH, Baldwin RL (1989) Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci USA 86:5286–5290

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford

Nishizawa M, Nishizawa K (1998) Biased usage of arginines and lysines in proteins are correlated with local-scale fluctuations of the G + C content of DNA sequences. J Mol Evol 47:385–393

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Palacios C, Wernegreen JJ (2002) A strong effect of AT mutational bias on amino acid usage in *Buchnera* is mitigated at high-expression genes. Mol Biol Evol 19:1575–1584

Riley MA, Gordon DM (1999) The ecological role of bacteriocins in bacterial competition. Trends Microbiol 7:129–133

Riley MA, Wertz JE (2002) Bacteriocins: Evolution, ecology, and application. Annu Rev Microbiol 56:117–137

Rodríguez-Herva JJ, Ramos-González MI, Ramos JL (1996) The *Pseudomonas putida* peptidoglycan-associated outer membrane lipoprotein is involved in maintenance of the integrity of the cell envelope. J Bacteriol 178:1699–16706

Rooney AP, Zhang J, Nei M (2000) An unusual form of purifying selection in a sperm DNA-binding protein. Mol Biol Evol 17:278–283

Ross RP, Morgan S, Hill C (2002) Preservation and fermentation: Past, present and future. Int J Food Microbiol 79:3–16

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Singer GA, Hickey DA (2000) Nucleotide bias causes a genome-wide bias in the amino acid composition of proteins. Mol Biol Evol 17:1581–1588

Sueoka N (1961) Correlation between base composition of deoxyribonucleotic acid and amino acid composition of protein. Proc Natl Acad Sci USA 47:1141–1149

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Sun TP, Webster RE (1987) Nucleotide sequence of a gene cluster involved in entry of E colicins and single-stranded DNA of infecting filamentous bacteriophages into *Escherichia coli*. J Bacteriol 169:2667–2674

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Tibor A, Weynants V, Denoel P, Lichtfouse B, De Bolle X, Saman E, Limet JN, Letesson JJ (1994) Molecular cloning, nucleotide sequence, and occurrence of a 16.5-kilodalton outer membrane protein of *Brucella abortus* with similarity to *pal* lipoproteins. Infect Immun 62:3633–3639

Wang HC, Hickey DA (2002) Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. Nucleic Acids Res 30:2501–2507

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 95:3708–3713