

Codon usage and lateral gene transfer in *Bacillus subtilis*

Ivan Moszer*, Eduardo PC Rocha*† and Antoine Danchin*§

Bacillus subtilis possesses three classes of genes, differing by their codon preference. One class corresponds to prophages or prophage-like elements, indicative of the existence of systematic lateral gene transfer in this organism. The nature of the selection pressure that operates on codon bias is beginning to be understood.

Addresses

*Unité de Régulation de l'Expression Génétique, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France
§e-mail: adanchin@pasteur.fr

†Atelier de Bio-Informatique, Université Paris 6, 12 rue Cuvier, 75005 Paris, France

Current Opinion in Microbiology 1999, 2:524–528

1369-5274/99/\$ – see front matter © 1999 Elsevier Science Ltd. All rights reserved.

Introduction

It has long been known that codon preference is related to gene expression [1–4]. It was assumed that codon usage resulted from an equilibrium between the selective optimisation of codon frequency to the available tRNA pool and the drift towards no codon preference. Kurland and co-workers [5,6] showed that the relative abundance, as well as the proportion of isoaccepting tRNAs, varied as a function of the growth rate, and that the tRNA pool is permanently adjusted in order to fit the average amino acid composition of the proteins. These observations suggested that the actual codon preference in a gene is correlated to the composition of the tRNA pool when the gene is expressed at a high level, even under circumstances which may differ significantly from the exponential growth phase. Therefore, genes highly expressed under a specific physiological state should display a similar codon usage bias. The fitness between codon frequency and tRNA availability modulates the expression level, although one cannot tell simply from the sequence whether the codon usage bias is a cause or a consequence of the bias in tRNA abundance. We review this hypothesis and show that codon preference is linked to control of translation and evolution in *Bacillus subtilis* genes, and that this has implications for horizontal transfer of genes.

Genome structure and codon usage

The genome of *B. subtilis* is 4215 kb long. Its G+C content is 43%. Some regions in the chromosome are more A+T rich [7]. A major trend of the genome organisation is that 75% of the genes are transcribed in the same orientation as the movement of the replicating fork. This is superimposed on a general bias in bacterial gene organisation, which clearly separates genes transcribed from the leading strand compared to genes transcribed from the lagging strand [8,9••].

tRNAs and codon usage

The *B. subtilis* genome contains 86 genes coding for 33 tRNAs (or 33 anticodons), distributed in 11 transcription units (comprising up to 21 tRNA genes) and eight single genes spread along the chromosome. One to six copies of each tRNA gene exist. The number of genes coding for isoacceptor tRNAs varies between one and eight, as a function of the cognate amino acid (Table 1). *B. subtilis* possesses ten rRNA operons; this is by far the largest number of all fully sequenced genomes (e.g. seven rRNA operons in *Escherichia coli*). These rRNA operons are all located on the leading strand and near the origin of replication, thereby optimising the coupling between replication and transcription and the gene dosage effect [7].

Overall, as shown in the presented table of codon usage, there is no strong bias in codon preference. However, some codons are rarely used, such as CUA (leucine) or AUA (isoleucine). AGG (arginine) is rare, presumably because it matches the core sequence of the ribosome-binding site (AAGGAGGU) and in some contexts this may induce abnormal translation initiation. This should be qualified, however, since GGA (glycine) is frequently used.

Codon usage classification and horizontal transfer

Grantham and his colleagues were the first to analyse codon usage using correspondence analysis in *E. coli* [10]. Factorial correspondence analysis (FCA) is a basic statistical technique that uses distances between objects in the space of a set of properties (here, the frequency of each codon used in a gene, normalised for the length of the protein and the average number of codons for each amino acid). It allows calculation of the projection of the cloud of points yielding maximum scattering. To analyse this graphical representation in terms of individual classes it is necessary to use a second method that automatically clusters the objects which are located close to one another [4]. Later, when a large number of genes became available, it was observed that the best simultaneous two-dimensional representation of the genes and codon usage had a 'rabbit head' trimodal shape [4]. This shape could be explained by the existence of three major classes of genes differing in their codon bias.

A similar analysis of *B. subtilis* coding sequences also revealed that the hypothesis requiring the minimum of assumptions supported a clustering of genes into three well separated classes [7,11•]. As seen by the absence of a category of very rare codons in this class, the codon bias is weak in class I, very strong in class II, and follows an A+U enrichment trend in class III (Table 1). The proof that this partition was significant came from the observation that these classes of genes could also be distinguished by their

Table 1

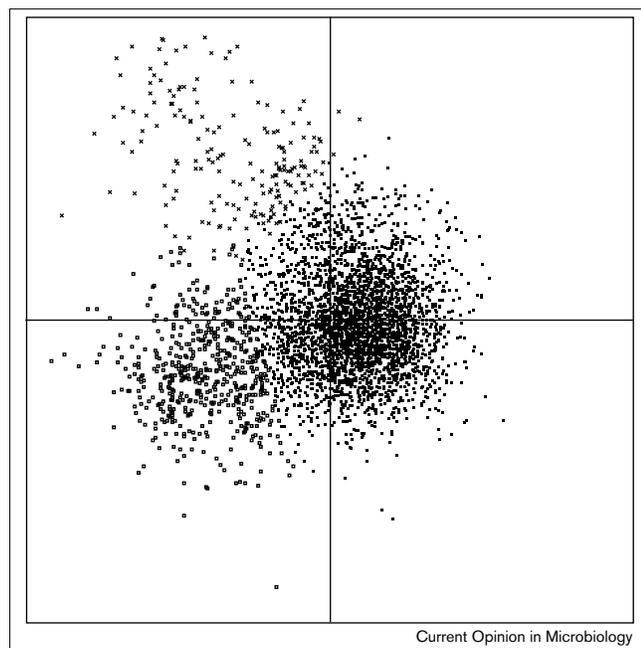
Codon usage bias in *B. subtilis* genes.

AA*	Codon† tRNA genes	Gene class§				AA*	Codon† tRNA genes	Gene class§			
		All	I	II	III			All	I	II	III
Phe	UUU	68.7	69.5	36.1	74.9	Leu	CUU	23.9	23.6	40.6	19.5
	UUC	31.3	30.5	63.9	25.1		CUC	10.7	11.7	5.6	6.5
Leu	B,D,Y UUA	21.0	18.4	24.7	36.2	S2 CUA	5.6	4.7	6.5	11.2	
	B2,D1,J2 UUG	16.1	16.1	11.6	17.8	J1,S1 CUG	22.7	25.5	11.0	8.8	
	D2					B1					
Ser	UCU	20.8	19.2	43.1	23.6	Pro	CCU	29.5	27.5	40.6	38.0
	UCC	12.4	13.5	5.6	7.7		CCC	10.1	10.4	2.6	10.8
	D UCA	23.8	23.5	21.3	26.2		CCA	20.4	16.9	34.3	37.7
	B1,SL1 UCG	9.9	10.8	2.4	6.6		B,I,J CCG	40.0	45.3	22.6	13.5
Tyr	UAU	65.0	65.4	39.0	71.7	His	CAU	66.8	67.1	42.4	73.6
	UAC	35.0	34.6	61.0	28.3		CAC	33.2	32.9	57.6	26.4
OCH	D,SL1 UAA	-	-	-	-	Gln	B,D CAA	53.5	49.4	75.6	70.8
AMB	UAG	-	-	-	-		D,S,SL1/2 CAG	46.6	50.6	24.4	29.2
Cys	UGU	46.7	45.0	43.7	58.3	Arg	CGU	17.8	16.7	48.1	14.4
	UGC	53.3	55.0	56.3	41.8		B,E,I,J CGC	19.2	20.3	33.4	7.8
OPA	D UGA	-	-	-	-	CGA	10.4	10.1	2.5	14.5	
Trp	UGG	100	100	100	100	CGG	14.9	16.8	1.0	8.2	
Ile	AUU	50.4	50.5	42.4	52.5	Val	GUU	28.9	26.3	41.6	40.4
	AUC	35.7	37.0	54.9	20.9		GUC	24.3	26.8	12.9	12.5
	A,B2,O AUA	14.0	12.6	2.7	26.6		SL2 GUA	20.7	18.6	30.6	30.4
Met	AUG	100	100	100	100	GUG	B,D,J,SL1 GUG	26.2	28.3	14.9	16.7
	B1/2/3,D,E,SL1										
Thr	ACU	17.3	13.8	35.0	33.4	Ala	GCU	25.9	23.3	41.1	36.5
	ACC	15.8	17.0	4.0	12.6		GCC	20.2	22.0	8.1	12.8
	I ACA	41.1	40.7	46.5	41.6		SL1 GCA	29.2	27.6	30.0	39.3
	B,D,J,SL1 ACG	25.8	28.5	14.5	12.4		A,B,I,J,O GCG	24.7	27.0	20.7	11.4
Asn	AAU	57.1	56.7	29.5	69.7	Asp	GAU	64.0	63.0	52.5	74.6
	AAC	42.9	43.3	70.6	30.3		GAC	36.0	37.1	47.6	25.4
Lys	B,D,I,S AAA	70.4	69.3	83.2	72.8	Glu	B,D,E,Y GAA	68.2	67.6	77.0	69.3
	B,J,S,Y AAG	29.6	30.7	16.8	27.2		B,D,S,SL1/2,Y GAG	31.8	32.4	23.0	30.7
Ser	AGU	11.0	9.4	8.3	22.1	Gly	GGU	18.8	16.5	33.0	28.4
	AGC	22.1	23.6	19.3	13.8		GGC	32.2	34.6	33.4	16.9
Arg	B2,S AGA	27.8	26.5	14.4	40.4	B1,D,I,J GGA	32.3	31.5	29.3	38.5	
	SL2 AGG	9.9	9.7	0.6	14.8	B2,E,SL1 GGG	16.7	17.5	4.2	16.3	
	SL1										

*The first column displays the three-letter code for each amino acid. †The second column displays each of the corresponding codon sequences, together with an identifier of the tRNA genes with the corresponding anticodon, abbreviated in the following way: the tRNA operon identification symbol (uppercase letter), followed by a number when several isoacceptor tRNA genes are present in the same operon. To retrieve the corresponding genes in Subtilist [32], append to 'trn' the

operon letter(s), a hyphen '-', the three-letter code of the cognate amino acid, and the number indicated after the operon letter, if any (e.g. 'B2' in the UUA codon cell means 'trnB-Leu2'). A black dot indicates that no tRNA gene has been identified with the corresponding anticodon. §Columns three to six show the average frequency of synonymous codons in all *B. subtilis* genes and in genes from classes I, II and III. AA, amino acid.

Figure 1



Factorial correspondence analysis of codon usage in *B. subtilis* genes. The two-dimensional projection of the cloud of genes in the codon space is presented on the plane formed by the two first axes displaying maximum inertia (i.e. spreading the projection onto a plane where it is maximally spread out). Genes from class I are represented by dots, genes from class II are represented by crosses, and genes from class III are represented by open squares.

biological properties. Class I contains most genes of intermediary metabolism, with the noticeable exception of genes involved in the core of carbon assimilation (i.e. glycolysis, TCA cycle and fatty acids synthesis). It also contains genes specifying gene regulation (activators and repressors), and genes responsible for DNA metabolism. Thus, class I comprises those genes that maintain a constitutive low or intermediary level of expression. In contrast, class II contains genes that are expressed at a high level under exponential growth conditions. Most of these genes are involved in the translation and transcription machinery, in the core of intermediary metabolism, and in the folding of proteins. The A+U-rich codon preference characterising class III is different from that of classes I and II. Genes of this class are in general clustered together in the chromosome, which, along with their functional classification and similarity to prophage genes or typical horizontally transferred genes (e.g. toxins), suggests a foreign origin [7,11•]. Several of these A+T-rich islands are located at the terminus of replication, a region sensitive to integration of foreign DNA due to the resolving of the knot-like structure necessary to the completion of replication [12]. It may even be that the machinery permitting termination of replication is, at least in part, resulting from a gene lateral transfer [13•].

Recently Lawrence and Ochman proposed a complementary approach to the inference of horizontal transfer from

genetic signatures [14,15••]. Their work explored the notion of genetic ‘amelioration’, that is a process by which a recently transferred gene gains gradually a genetic signature similar to its new host genome. This amelioration includes similarity of codon preference, but also G+C content and other variables [14]. Applied to a comparison between *E. coli* and *Salmonella typhimurium*, this argument led to the estimation that about 18% of the *E. coli* genome was recently acquired. This approach, though methodologically quite different from correspondence analysis, gave values close to the ones given by the latter analysis method (16% of horizontally transferred genes) [4]. Unfortunately the absence of a sequenced genome close to *B. subtilis* does not yet allow a similar study for this species.

Links between codon usage and cell general metabolism tRNA abundance in the cell

Because some preferred or avoided codons matched tRNA abundance in *E. coli*, it was quickly accepted that there is a match between codon usage and tRNA availability [16•]. However, this could not be the whole story because some genes expressed at a high level do not display a highly biased codon preference. As a case in point, expression of a colicin gene in *E. coli* was very high: this was unexpected, considering the codon preference of the gene (containing for instance several ‘rare’ AUA isoleucine codons [17]). At that time, such facts were seen as exceptions to the rule, but now that class III genes have been found to constitute a large fraction of the *B. subtilis* (and *E. coli*) genome (13%), it is accepted as a major genomic trend. Moreover, experiments demonstrated that the relative availability of tRNA varied as a function of the growth rate. Bacteria are only rarely growing exponentially, and the other states of growth are certainly as important for survival and evolution of the population. Analysis of different codon usage biases may provide an important hint to the function of a gene by classifying it together with genes expressed under similar physiological conditions. However, more accurate statistical techniques of classification and discrimination need to be developed to improve our understanding.

Other translation biases

In *B. subtilis* there are three major start codons (AUG, GUG and UUG). AUG is, by far, the most frequently used. The ambiguity is on the first rather than on the third letter of the codon. This is an indication that the codon/anticodon interaction at the start site is in a structure that differs from that of the usual codon/anticodon interaction at the A-site of the ribosome. Indeed, the initiator tRNA is at the P-site, brought in by a specific factor, IF2: during the messenger RNA movement a conformational change is triggered at the anticodon of the tRNA when it changes from an aminoacylated form to a peptidylated form. There seems to be an effect in the translation efficiency linked to the nature of the translation start codon: AUG is more

efficient than its counterparts [18]. However, genes expressed constitutively at high rates do not show a more important preference for AUG. Moreover, the start codon is not conserved between paralogues, apparently contradicting the higher efficiency of AUG [19].

Three codons terminate translation: UAA, UGA and UAG. In *B. subtilis* these codons are present with the relative abundance UAA > UGA > UAG. For class II genes, one finds an almost exclusive use of UAA as the stop codon [19,20]. This is probably an indication of a higher efficiency similar to that observed in *E. coli* [21].

Codon usage is different in the neighbourhood of the start codon in several enterobacteria [22]. Such bias also occurs in *B. subtilis*. It can be explained either on the basis of some amino acid bias in this region, or by the A-richness necessary to avoid the formation of mRNA secondary structures, which would interfere with translation initiation [9•,18]. Interestingly a similar bias, though weaker, was observed in the vicinity of the stop codon, probably indicating that such mRNA structure avoidance is also important for a correct interaction between the release factor and the mRNA [19]. At the translation stop it was also recently found that amino acid context effects can determine the efficiency of termination in *E. coli* and *B. subtilis* [23,24•]. In particular, this seems to be related to the van der Waals volume of the last amino acid and to the hydrophobicity of the penultimate residue [25].

Codon context, rate and accuracy of protein synthesis

During translation, peptidyl-tRNA and the incoming charged tRNA are adjacent to each other. It is, therefore, expected that some interaction exists between these molecules, implying constraints on the relative frequency of adjacent anticodons. This should be reflected in the context of individual codons, and might result in specific biases in the local nucleotide composition of strings longer than three letters [6,26]. There exists an important literature devoted to context effects both *in vivo* and *in vitro* [6]. The context has particularly important consequences in the case of rare codons. For example, the AGG codon is slowly translated [27], and two such codons in a row are discriminated against because it results in abnormal translation (frameshifting or downstream initiation of unwanted translation) [28]. The effect of context is also seen in the case of suppression of translation termination: suppressor tRNAs can have widely variable efficiency, according to the nature of the codon following the non-sense suppressed codon [29•].

Synonymous codons can be chosen for various reasons, including, as discussed, preference for a given tRNA. But one should also consider that ribosomes could differ in their discriminating power according to the nature of the codon, resulting in variations in accuracy of translation. It is known that the average translation error rate is rather

high, between 10^{-3} and 10^{-4} per nucleotide (i.e. an expected error in one protein among ten of length 1000 amino acids). This is not very important in general because neighbour codons in the genetic code table code for amino acids having similar properties, and because proteins are robust for most amino acid changes. In specific cases, however, there are drastic constraints on the conservation of a given residue (e.g. necessary for proper folding or catalysis) and this will require, locally, to select for high accuracy of translation. Dix and Thompson [30] have shown that indeed accuracy can heavily depend on a given codon. In particular, they have shown that the efficient UUC codon is less accurate than UUU, but preferred in highly expressed genes (class II genes), indicating that there has not been a strong selection for accuracy at the expense of translation speed [30]. This means that most proteins are robust to amino acid misincorporation, in particular those that are important for the core of the cell machinery, which are probably selected to avoid 'error catastrophes'. However, the role of codon context might be an efficient means to counter-select genes incorporated by horizontal transfer.

Conclusions

Remarkably, the same details in the translation mechanisms seem to hold true for both *B. subtilis* and *E. coli*, indicating an important conservation of the eubacterial translation machinery through 1.2 billion years of divergent evolution. This allows efficient incorporation of horizontally transferred genes, even in bacteria such as *B. subtilis*, which seems to avoid processes and structures such as insertion sequences that permit efficient lateral transfer. Maintenance of a codon preference asks either for a systematically low expression level or for some sort of compartmentalisation of the translation machinery linked to the organisation of the chromosome [31•]. This can be accounted for as follows [31•]. When considering diffusion of molecules and rates of biochemical reactions, the cytoplasm of bacteria is a slowly moving ribosome lattice. As a function of the messenger RNA molecules that are read by a given ribosome, this ribosome will behave as an attractor of those tRNA molecules that are adapted to the specific codon bias of the mRNA. This creates an efficient selection pressure leading to adaptation of the codon usage of the translated message as a function of its position in the cell's cytoplasm. Because the codon bias changes from mRNA to mRNA this shows that these different molecules are not translated in the same ribosome environment in the usual life cycle of the organism. In particular, if two genes have very different codon preferences this indicates that the corresponding mRNAs are not made from the same part of the cell (it is indeed difficult to see how ribosomes sitting next to each other could attract different tRNA molecules). This would suggest that stable insertion of foreign genes is not random. Comparison with genomes of closely related species will allow us to explore further this hypothesis.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R: **Codon catalog usage is a genome strategy modulated for gene expressivity.** *Nucleic Acids Res* 1981, **9**:43-74.
 2. Ikemura T: **Codon usage and tRNA content in unicellular and multicellular organisms.** *Mol Biol Evol* 1985, **2**:13-34.
 3. Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F: **Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within species diversity.** *Nucleic Acids Res* 1988, **16**:8207-8211.
 4. Médigue C, Rouxel T, Vigier P, Hénaut A, Danchin A: **Evidence for horizontal gene transfer in *Escherichia coli* speciation.** *J Mol Biol* 1991, **222**:851-856.
 5. Dong H, Nilsson L, Kurland CG: **Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates.** *J Mol Biol* 1996, **260**:649-663.
 6. Berg OG, Silva PJ: **Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection.** *Nucleic Acids Res* 1997, **25**:1397-1404.
 7. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessières P, Bolotin A, Borchert S *et al.*: **The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.** *Nature* 1997, **390**:249-256.
 8. Perrière G, Lobry JR, Thioulouse J: **Correspondence discriminant analysis: a multivariate method for comparing classes of protein and nucleic acid sequences.** *Comput Appl Biosci* 1996, **12**:519-524.
 9. Rocha EPC, Danchin A, Viari A: **Universal replication biases in bacteria.** *Mol Microbiol* 1999, **32**:11-16.
- The term 'universal' emphasises the fact that, when there exists a bias between the lagging and leading strand, it has always the same form in terms of nucleotides, codons and amino acids. When present, this demonstrates a remarkable stability in a genome: if genes were to shuffle too frequently, this would have smoothed out the bias.
10. Grantham R, Gautier C, Gouy M: **Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type.** *Nucleic Acids Res* 1980, **8**:1893-1912.
 11. Moszer I: **The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis.** *FEBS Lett* 1998, **430**:28-36.
- A model for the organisation of bacterial genome data and gene annotation.
12. Corre J, Cornet F, Patte J, Louarn JM: **Unraveling a region-specific hyper-recombination phenomenon: genetic control and modalities of terminal recombination in *Escherichia coli*.** *Genetics* 1997, **147**:979-989.
 13. Griffiths AA, Andersen PA, Wake RG: **Replication terminator protein-based replication fork-arrest systems in various *Bacillus* species.** *J Bacteriol* 1998, **180**:3360-3367.
- Is the control of replication termination transferred by lateral gene exchange? The nine 30 bp terminator sequences of *B. subtilis* 168 are located within the terminus region, outside of protein coding sequences and adjacent to an AT-rich region. The sequence is conserved in several related *Bacillus* species, but cannot be detected by Southern hybridisation in *B. licheniformis*. The homolog of the replication terminator protein has not been discovered either in this organism. The authors speculate that this might be the sign of a recent lateral gene transfer event.
14. Lawrence JG, Ochman H: **Amelioration of bacterial genomes: rates of change and exchange.** *J Mol Evol* 1997, **44**:383-397.
 15. Lawrence JG, Ochman H: **Molecular archaeology of the *Escherichia coli* genome.** *Proc Natl Acad Sci USA* 1998, **95**:9413-9417.

This paper, along with [14], tries to delineate general approaches for the study of gene exchange, and exposes the consequences of horizontal gene

transfer for genome organisation. These issues will have an increasingly important impact on our ideas about the origin of the first cells.

16. Berg OG: **Synonymous nucleotide divergence and saturation: effects of site-specific variations in codon bias and mutation rates.** *J Mol Evol* 1999, **48**:398-407.
- Analysis of synonymous codon divergence in *E. coli* and *S. typhimurium*: a similar comparison is much awaited for *B. subtilis*!
17. Varenne S, Baty D, Verheij H, Shire D, Lazdunski C: **The maximum rate of gene expression is dependent on the downstream context of unfavourable codons.** *Biochimie* 1989, **71**:1221-1229.
 18. Vellanoweth RL, Rabinowitz JC: **The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* in vivo.** *Mol Microbiol* 1992, **6**:1105-1114.
 19. Rocha EPC, Viari A, Danchin A: **Oligonucleotide bias in *Bacillus subtilis*: general trends and taxonomic comparisons.** *Nucleic Acids Res* 1998, **26**:2971-2980.
 20. Sharp PM, Bulmer M: **Selective differences among translation termination codons.** *Gene* 1988, **63**:141-145.
 21. Poole ES, Brown, CM, Tate, WP: **The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*.** *EMBO J* 1995, **14**:151-158.
 22. Bulmer M: **Codon usage and intragenic position.** *J Theor Biol* 1988, **133**:67-71.
 23. Poole ES, Major LL, Mannering SA, Tate WP: **Translational termination in *Escherichia coli*: three bases following the stop codon crosslink to release factor 2 and affect the decoding efficiency of UGA-containing signals.** *Nucleic Acids Res* 1998, **26**:954-960.
 24. Mottagui-Tabar S, Isaksson LA: **The influence of the 5' codon context on translation termination in *Bacillus subtilis* and *Escherichia coli* is similar but different from *Salmonella typhimurium*.** *Gene* 1998, **212**:189-196.
- Codon context seems to be important for translation termination. At the present time we still do not possess laws explaining observed regularities. It seems necessary to reinterpret data using a partition of genes into various classes, in particular treating in a separate way genes that have been horizontally transferred.
25. Björnsson A, Mottagui-Tabar S, Isaksson LA: **Structure of the C-terminal end of the nascent peptide influences translation termination.** *EMBO J* 1996, **15**:1696-1704.
 26. Smith HO, Tomb JF, Dougherty BA, Fleischmann RD, Venter JC: **Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome.** *Science* 1995, **269**:538-540.
 27. Bonekamp F, Jensen KF: **The AGG codon is translated slowly in *E. coli* even at very low expression levels.** *Nucleic Acids Res* 1988, **16**:3013-3024.
 28. Chen GF, Inouye M: **Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes.** *Nucleic Acids Res* 1990, **18**:1465-1473.
 29. Mejlhede N, Atkins JF, Neuhaud J: **Ribosomal-1 frameshifting during decoding of *Bacillus subtilis* *cdd* occurs at the sequence CGA AAG.** *J Bacteriol* 1999, **181**:2930-2937.
- These authors demonstrate that programmed frameshifting exists in *B. subtilis*.
30. Dix, DB, Thompson, RC: **Codon choice and gene expression: synonymous codons differ in translational accuracy.** *Proc Natl Acad Sci USA* 1989, **86**:6888-6892.
 31. Danchin A, Guerdoux-Jamet P, Moszer I, Nitschké P: **Mapping the bacterial cell architecture in the chromosome.** *Philos Trans R Soc Lond B Biol Sci* 1999, in press.
- A general model of the organisation of the bacterial cytoplasm is discussed, with reference to the position of the genes in the genome.
32. The SubtiList World-Wide Web Server on URL: <http://bioweb.pasteur.fr/GenoList/SubtiList/>