The map of the cell is in the chromosome

Commentary

Antoine Danchin* and Alain Hénaut†

Addresses
*Régulation de l'Expression Génétique, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France;
e-mail: adanchin@pasteur.fr
†Université de Versailles Saint Quentin, Laboratoire Génome et Informatique, 45 avenue des Etats Unis, 78000 Versailles, France

Current Opinion in Genetics & Development 1997, 7:852–854

http://biomednet.com/elecref/0959437X00700852

© Current Biology Ltd ISSN 0959-437X

Introduction

Completely sequenced bacterial genomes are accumulating at the rate of one per month, allowing analysis of the distribution of genes along the chromosome. We suggest here that a consistent picture exists demonstrating that the map of the cell is in the chromosome and that translation is the driving force in this process.

Knowledge of whole genome sequences is a unique opportunity for studying the relationships between genes and their products at the global cell architecture level [1,2]. Part of the difficulty of this study comes from the fact that — contrary to a generally accepted and intuitive idea — there is often no predictable link between structure and function in biological objects. Being influenced by natural selection pressure, however, there must exist some fitness between gene, gene products and the survival of the organism. This indicates that observing biases in features which would conceptually be thought of as being unbiased would be the hallmark of some selection pressure [3]. As the genetic code is redundant, coding sequences can be studied by analysing their codon usage. If there were no bias, all codons for a given amino acid should be used more or less equally. This is not what is generally observed [4–6]. In fact, it has long been observed in *Escherichia coli* that genes could be split into three classes according to the way that codons were arranged [7] and the same is true for *Bacillus subtilis* [8]. One might think, however, that random mutations should somehow smooth out such differences but this is not the case: indeed, in the case of leucine (where six codons are used) we find that the CUG codon is used >70% of the cases in genes that are expressed at a high level during exponential growth conditions, whereas CUA is expressed in <2% of the cases [7].

Sources of the codon usage bias

What is the source of the strong codon usage biases observed in various genes? There could be a systematic effect of context, some DNA sequences being favoured or selected against. Although this could be true for some codons, this cannot be generalised [9]. We know that translation of mRNA into proteins requires the action of transfer RNA molecules. As there are fewer tRNAs specific for a given amino acid than the number of codons, this means that some tRNAs must read several codons. A bias in the concentration of some tRNAs might thus result in a bias in codon usage. Therefore we are led to analyse selection pressure occurring at the level of tRNA synthesis [10]. This is the generally accepted reason for the bias that is observed in codon usage [11,12].

Unfortunately, two facts contradict this interpretation. First, in much the same way that one might expect a smoothing out of bias in codon usage, similar constraints would smooth out biases in tRNA synthesis. For example, if a tRNA gene had a strong promoter, spontaneous mutations would tend to lower its efficiency, making transcription of this particular tRNA similar to its other counterparts. This is true *unless* there is selection pressure for the converse. Second, whereas explanation for the strong bias in a given class of genes could be explained in this way, the same explanation cannot hold for a strong bias in another class of genes. We know, however, both from the study of the *E. coli* and *B. subtilis* genomes, that two classes of genes (class II and class III) display extremely strong but *different* biases; a tRNA molecule cannot be simultaneously expressed at a high level and not at a high level.

The architecture and dynamics of the cell

The issue of codon-usage bias requires examination from a different perspective. In order to do so, we investigated the nature of translation in the physics of the cell. The cytoplasm of a cell cannot simply be considered as a tiny test tube. One of the most puzzling features of the cell cytoplasm organisation is that it must accommodate a hugely long thread of DNA and that this molecule must be transcribed as a multitude of RNA threads that are usually as long as the entire cell. Anyone who has tried fly-fishing knows what usually happens in this situation: knots are formed everywhere. This, however, does not seem to be the case in the cell cytoplasm, proving that there must be some kind of organisation of transcription, translation, and replication such that mRNA molecules and DNA are not always inextricably entangled [13].

Translation

Let us first consider the ribosomes. The volume of a ribosome is ~200 Å^3. In an *E. coli* cell growing exponentially in a rich medium, there are at least 15,000 ribosomes and therefore the fraction of the cell occupied by ribosomes is at least 12%. The remaining ribosome-free

fraction of the cell is in fact significantly smaller if one takes into account the volume occupied by the chromosome (which is folded 1000-fold in the cell, like a Peano curve) and by the transcription and the replication machineries. If one now considers also that the translation machinery requires an appropriate pool of elongation factors, tRNA synthetases, and tRNAs, it becomes clear that the cytoplasm is very far from a mere test-tube solution but behaves more like a gel.

Furthermore, simply counting the number of tRNA molecules dotted around a ribosome, it appears that one cannot speak about the concentration of such molecules but only about a small and finite number. Compartmentalisation has been demonstrated to be important even for small molecules, despite the fact that they could diffuse quickly [14]. Even in eukaryotes, channelling of aminoacyl-tRNA for protein synthesis has indeed been shown to occur *in vivo* [15,16]. As a consequence, one should consider a translating ribosome as an attractor of a certain pool of tRNA molecules. In such a case, diffusion should only be considered as occurring locally. The cytoplasm thus becomes a ribosome lattice, probably undergoing relatively slow movements (with respect to local diffusion of small molecules as well as macromolecules). This scenario therefore constitutes an efficient source of selection pressure leading to the adaptation of codon usage of the translated message as a function of its position in the cell's cytoplasm. If the codon usage changes from one mRNA molecule to another, then these different molecules probably do not encounter the same ribosomes in the usual life cycle of the organism; in particular, if two genes have very different codon usage, the indication is 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).

Transcription

In the case of transcription, several experimentally valid models account for a process in which the transcribed regions are present at the surface of the chromoid in such a way that RNA polymerase does not have to travel around the double helix it is unwinding and transcribing [17]. Thus mRNA threads, usually structured at their 5′ end, are pulled off DNA by the lattice of ribosomes, going from one ribosome to the next one, as does a thread in a wire-drawing machine (this is exactly the opposite view of translation as shown in textbooks, where ribosomes are depicted travelling along fixed mRNA molecules). In this process, a nascent protein is synthesised on each ribosome, spread throughout the cytoplasm by the linear diffusion of the mRNA molecule from one ribosome to the next one, obviating the requirement for the much slower three-dimensional diffusion of the protein.

There is strong experimental evidence supporting this process. If mRNA is indeed pulled off DNA, it must sometimes break but broken mRNA is a potentially dangerous molecule because, if translated, it will produce a truncated protein and such proteins are often toxic because they can disrupt the architecture of multisubunit complexes (this is why many nonsense mutants are negative dominant, rather than recessive). If this is real then there must exist a process for coping with such frequent accidents. The 10Sa RNA molecule fits this role nicely. When a truncated mRNA molecule reaches its end, the ribosome stops translating and 'waits'. 10Sa RNA—folded like tRNA and charged with alanine—comes in, places its alanine at the caraboxyl terminus of the nascent polypeptide then takes the place of the mRNA inside the ribosome, being translated as a short sequence: AANDENYALAA. This tail is a motif that is used to direct it to the proteasome, where it is degraded [18,19]. It seems likely that the organisation of the ribosome lattice, coupled to the organisation of the transcribing surface of the chromoid ensures that mRNA molecules are translated parallel to each other in such a way that they do not become knotted. In addition, it is important that transcription does not conflict with ongoing replication, which requires specific organisation of both processes [20,21].

Polycistronic operons ensure that proteins having related functions are co-expressed locally, permitting channelling of the corresponding substrates and products. It also seems likely that the structure of mRNA molecules is coupled to their fate in the cell and to their function in compartmentalisation.

If we consider genes translated sequentially in operons as being physiologically and structurally relevant, we should also analyse mRNAs that are translated in parallel to each other. Indeed, if there is correlation of function and/or localisation in one dimension, there should also exist a similar constraint in the orthogonal directions. How would this be manifested? This is where codon usage comes in again. Indeed, if ribosomes act as attractors of tRNA molecules, the implication is of a local coupling between these molecules and the codons they can use in the message they read. This obviously requires that the same ribosome mostly translates mRNAs with similar codon usage but the consequence is that, as one goes away from a strongly biased ribosome, there is increasingly less availability of the most biased tRNAs. In turn, the suggestion is that there would be selection pressure for a gradient of codon usage as one goes away from the most biased messages and ribosomes. If this reasoning is well founded, then the transcripts are nested around central core(s), formed of transcripts for highly biased genes. This fits with what is seen of the general organisation of genes in the chromosome, however, correlating in particular with the observation that the distance between *E. coli* genes oriented in the same direction on the chromosome is positively related to the expression level of the downstream gene [22].

Replication

A further consideration should be taken into account. The chromosome strands must separate from each other and migrate to each of the daughter cells. This means that there must exist some kind of repulsive force which pushes newly synthesized chromosomes away from each other. Although there probably are gene products involved directly in this process [23], ribosome synthesis—in particular from regions near the origin of replication—performs exactly what is required by continuously creating new ribosomes. Continuous synthesis of ribosomes in between the replicating forks would also provide a mechanical stress on the bacterial wall in the middle of the cell. Koch [24] has argued convincingly that the bacterial wall is indeed a stress-bearing structure. If ribosome sources are organisers of the cell, mRNA for genes highly expressed under exponential growth conditions should be located near the center of these organisers, whereas other mRNAs should be translated in nested layers all the way to the ribosomes that are located near the cytoplasmic membrane and which would be involved in cotranslational membrane protein localisation. Organisation of the genes in the chromosome should therefore show regularities that are linked to this architecture. Our work on the origin of the outer membrane of *E. coli* [25] fits well with this picture, and strongly suggests to us that genes along the chromosome specify the map of the cell—a kind of 'celluloculus'.

Conclusion: a question of eukaryotes

This short description of chromosomal organisation holds for bacteria but it is clear that the situation must be more complex in eukaryotes, not only because of the existence of a nucleus but also because the cell volume–gene number ratio, has increased enormously. In this case, we think that the split gene structure, as well as the large membrane structure fraction (rough endoplasmic reticulum, Golgi, mitochondria, etc.) must participate in the translation of the genomes' map of the cell into the construction of the cell structure. This somehow requires a complex interaction between transcription and translation in a cell that separates these processes by the nuclear envelope. It is interesting, in this respect, to see that there exists a splicing-dependent process that detects translation signals [26].

Acknowledgements

Our work was supported by grants from the Groupement de Recherche et d'Étude des Génomes, from Centre National de la Recherche Scientifique (GDR 1029), from the *Bacillus subtilis* Industrial Platform (BACIP), from Institut Pasteur and from Université de Versailles.

References

1. Magpie Genome Sequencing Project List on World Wide Web URL: http://www.mcs.anl.gov/home/gaasterl/genomes.html
2. The Institute for Genomic Research (TIGR) on World Wide Web URL: http://www.tigr.org
3. Grantham R, Gautier C, Gouy M, Mercier R, Pavé A: **Codon catalog usage and the genome hypothesis.** *Nucleic Acids Res* 1980, **8**:49-61.
4. 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**:R43-R74.
5. Ikemura T: **Correlation between the abundance of *Escherichia coli* transfer RNAs and occurence of the respective codons in its protein genes.** *J Mol Biol* 1981, **146**:1-21.
6. Ikemura T: **Correlation between the abundance of *Escherichia coli* transfer RNA and the occurrence of the respective codons in its protein gene: a proposal for a synonymous codon choice that is optimal for *E. coli* translation system.** *J Mol Biol* 1981, **151**:389-409.
7. 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.
8. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni V, Azevedo V, Bertero G, Bessières P, Bolotin A, Borchert S *et al.*: **The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.** *Nature* 1997, **390**:249-256.
9. Hénaut A, Danchin A: **Analysis and predictions from *Escherichia coli* sequences, or *E. coli in silico*.** In Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*. Edited by Neidhardt F *et al.* Washington DC: ASM Press; 1996:2047-2066.
10. Dong H, Nilsson L, Kurland C: **Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates.** *J Mol Biol* 1996, **260**:649-663.
11. Emilsson V, Nüslund AK, Kurland CR: **Growth-rate-dependent accumulation of twelve tRNA species in *Escherichia coli*.** *J Mol Biol* 1993, **230**:483-491.
12. Berg O, Kurland C: **Growth rate-optimised tRNA abundance and codon usage.** *J Mol Biol* 1997, **270**:544-550.
13. Danchin A: **By way of introduction: some constraints of the cell physics that are usually forgotten, but should be taken into account for *in silico* genome analysis.** *Biochimie* 1996, **78**:299-301.
14. Welch G, Easterby J: **Metabolic channeling versus free diffusion: transition-time analysis.** *Trends Biochem Sci* 1994, **19**:193-197.
15. Negrutskii BS, Deutscher MP: **Channeling of aminoacyl-tRNA for protein synthesis *in vivo*.** *Proc Natl Acad Sci USA* 1991, **88**:4991-4995.
16. Negrutskii BS, Deutscher MP: **A sequestered pool of aminoacyl-tRNA in mammalian cells.** *Proc Natl Acad Sci USA* 1992, **89**:3601-3604.
17. Felsenfeld G: **Chromatin as an essential part of the transcriptional mechanism.** *Nature* 1992, **355**:219-224.
18. Keiler K, Waller P, Sauer R: **Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA.** *Science* 1996, **271**:990-993.
19. Atkins J, Gesteland R: **A case for trans translation.** *Nature* 1996, **379**:769-771.
20. Heggeler-Bordier BT, Wahli W, Adrian M, Stasiak A, Dubochet J: **The apical localization of transcribing RNA polymerases on supercoiled DNA prevents their rotation around the template.** *EMBO J* 1992, **11**:667-672.
21. Liu B, Wong ML, Tinker RL, Geiduschek EP, Alberts BM: **The DNA replication fork can pass RNA polymerase without displacing the nascent transcript.** *Nature* 1993, **366**:33-39.
22. Eyre-Walker A: **The distance between *Escherichia coli* genes is related to gene expression levels.** *J Bacteriol* 1995, **177**:5368-5369.
23. Rothfield L: **Bacterial chromosome segregation.** *Cell* 1994, **77**:963-966.
24. Koch AL: **Biophysics of bacterial walls viewed as stress-bearing fabric.** *Microbiol Rev* 1988, **52**:337-353.
25. Guerdoux-Jamet P, Hénaut A, Nitschké P, Risler J-L, Danchin A: **Using codon usage to predict genes origin: is the *Escherichia coli* outer membrane a patchwork of products from different genomes?** *DNA Research* 1997, **4**:257-265.
26. Carter MS, Li S ,Wilkinson MF: **A splicing-dependent regulatory mechanism that detects translation signals.** *EMBO J* 1996, **15**:5965-5975.