



Just so genome stories: what does my neighbor tell me?

Stanislas Noria, Antoine Danchin*

HKU-Pasteur Research Centre, Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong, China

Abstract

Discovery supposes either chance or inductive reasoning. We document here an approach to explore genome texts, where the neighborhoods of genes and gene products (in the genome, in phylogeny, in metabolic pathways, in codon usage bias, “in biblio”...) were used to make inferences about the function of unknown genes. Examples are provided in the case of pyrimidine biosynthesis, showing that de novo DNA synthesis requires either mRNA or membrane turnover (or both). We further show that this approach indicated that uridylyate kinase must be compartmentalized, and that this compartmentalization was likely to be associated to transcription termination at Rho-independent terminators. Finally, this led us to explore links between secretion, translation and the architecture of the cell, pointing to a role of sulfur metabolism, methionine recycling and salvage in particular.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Functional analysis; In silico; Drug target; Predictive genomics; Sulfur metabolism

1. Introduction

In May 1991, when at Elounda in Crete the sequence of the first chromosome of yeast and the first 100 kb of the *Bacillus subtilis* genome were publicly disclosed, a remarkable surprise was placed in the limelight. As many as 60% of the putative genes did not look like anything known. This prompted Piotr Slonimski to name them EEC genes since the sequencing had been financed by a program of the European Union (then named EEC) and

Abbreviations: AdoHcy, S-adenosyl-homocysteine; NDK, nucleoside diphosphokinase; PNPase, polynucleotide phosphorylase; UMK, uridylyate kinase.

* Corresponding author. Tel.: +852-2816-8402; fax: +852-2872-5782.

E-mail address: adanchin@hkucc.hku.hk (A. Danchin).

that they had clearly been Elusive, were Esoteric and were certainly Conspicuous! Since then, the number of genes without function has remained somewhat of a mystery. Of course now many have counterparts in different organisms, but still, what does “unknown, similar to unknown” mean? It has been urgent ever since to find ways to propose functions for those genes, and to construct appropriate experimental methods to explore these functions. While it is easy to understand how hypothetico-deductive methods (that some tend to assimilate—wrongly—with tautology) work, not much is known about ways to make inductive guesses. In this presentation, we shall show that looking for neighbors is a way to approach unexpected functions, in a self-consistent inductive method.

2. Results and discussion

2.1. What types of neighborhoods?

When the National Library of Medicine in the United States started to place Medline Abstracts on the World Wide Web, not many scientists noticed that its core Entrez software had, associated to each abstract, an unobtrusive, little cited link: “Related articles” (<http://www.ncbi.nlm.nih.gov:80/entrez/>). Clicking on the link directly led to articles with a content strongly reminiscent of that of the article of interest, so much that this is now a convenient way to identify plagiarism! However, this also paved the way to make inductive reasoning: when scientists associate two gene names in a same article, this is because they made some connection between them, and this is not by chance, this uses the still extremely poorly understood human intelligence. We must use this approach, and this “in biblio” complement of the in silico study of genes is not only useful because it makes connections between objects, but it shows us the way, indicating that many more neighborhoods should be used.

The first and intuitive relationship between two genes is their proximity in the chromosome. Although the concept of operon, or in a broader sense, of pathogenicity island, is clearly related to such proximity, this kind of relationship is far from sufficient to explain *functional* relationships between genes. We must therefore consider many other kinds of gene neighborhoods when exploring their function. The classical comparison method, which is used in the ubiquitous Blast searches, relates genes or gene products because they evolved from a common ancestor. This constitutes families of paralogues (within a genome) or orthologues (between different genomes). Paralogous or orthologous genes constitute clusters of genes with similar sequences [1]. In fact, one should only speak of orthologues when not only the sequence, but also the function, is identical in the organisms. The implicit unqualified extension of the use of the word is quite misleading because sequence orthology does not always mean function orthology [2]. Of course, there is no reason to reduce possible neighborhoods to these two kinds of relationships. For example, we can consider that genes coding for proteins involved in the same metabolic pathways, or using the same substrate, are neighbors. This constitutes the metabolic neighborhoods. Compartmentalization is an obvious architectural neighborhood: genes may correspond to proteins that belong to a common complex. The length of proteins, their amino acid composition and their electric charge are also typical neighborhoods.

More complex relationships have been described, such as relationships based on the genetic codon usage bias: two genes can be related because they use synonymous codons with the same frequency [3]. In what follows, we shall see a few examples of the results obtained when one combines the study of a variety of neighborhoods to explore gene functions.

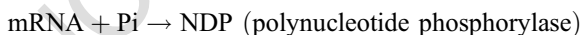
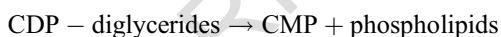
3. A few examples of neighborhood searches

3.1. The nucleoside diphosphate paradox

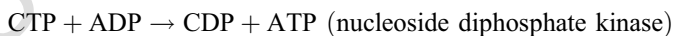
Seen from the eye of a chemist, metabolic pathways look like the work of a very bizarre apprentice. For example, biosynthesis of nucleotides has quite a few surprising features, and this has unexpected biological consequences. The recent discovery that de novo pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii* is probably much more general than this particular example [4]. Starting from orotidine monophosphate, pyrimidine synthesis ends up with CTP according to the widely spread sequence (metabolic neighborhood):



The absence of CMP and CDP is prominent in this de novo pathway (naturally scavenging and salvage pathways exist, where these latter intermediates are present, but under exponential growth conditions, for example, the de novo pathway has to operate in its full). Remarkably, both these nucleotides, which have counterparts in the de novo pathway of the three other nucleoside triphosphates, can only be produced in the case of cytidine nucleotides by the turnover of molecules consuming CTP in their biosynthesis, such as (other pathways exist):



and not

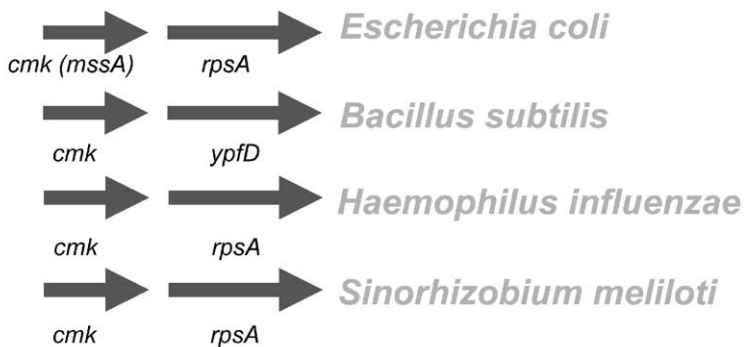


because the equilibrium between CDP and CTP is normally driven by the ADP/ATP ratio in the direction of the triphosphate synthesis.

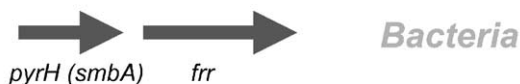
This original feature of cytidine nucleotides biosynthesis would not have important significance had not DNA synthesis required nucleoside *diphosphates* and not *triphosphates* as precursors for deoxyribonucleotides synthesis [5]. This requirement, which is (almost) universally true, creates an interesting metabolic bottleneck on cytosine deoxyribonucleotides availability (for both synthesis of dC and dT nucleotides).

It was therefore of interest to explore the question further using genome data and other types of neighborhoods. Analysis of the chromosome neighborhood of the cytidylate kinase gene in a variety of bacterial genomes demonstrated that the cognate gene, *cmk*, is generally expressed in an operon comprising an RNA binding protein gene, highly similar to (or identical to) the gene coding for ribosomal protein S1, *rpsA* (Fig. 1A). This proximity was puzzling. It drove us to analyze another neighborhood, that of the sequence of the protein S1 itself (phylogenetic neighborhood). Comparison with protein data libraries showed that this protein contains a motif, the S1-box, present in many other

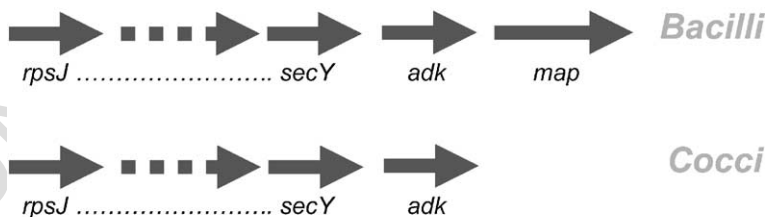
108
109
110
111
112
113
114
115



A



B



C

Fig. 1. Some surprising genome neighborhoods. (A) The *cmk rpsA* bicistron is conserved in many bacterial species. Its role is to supply the cell in CDP for DNA synthesis. (B) The *pyrH frr* bicistron is almost universally conserved in bacteria. This indicates a coupling between translation, transcription termination and pyrimidine biosynthesis. (C) The *rpsJ* ribosomal protein operon ends with *secY adk map* in *Bacilli*, while *map* disappears in *Cocci*.

proteins (<http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00575>). The S1-box is around 70 amino acids long, found in a large number of RNA-associated proteins (often nucleases) [6,7]. Ribosomal protein S1 itself binds RNA through its S1 domain with some degree of sequence specificity [8]. Among the S1-domain proteins is the once famous polynucleotide phosphorylase (PNPase, thought at one time to be RNA polymerase), which directly makes NDPs by degrading mRNAs from their 3'-end. Study of PNPase led us to another neighborhood, a biochemical one: it is a member of a protein complex. Indeed, this protein is part of the "degradosome", a complex that contains not only PNPase, but also enolase and polyphosphate kinase. Recently, new proteins have been found to belong to the complex [9]: S1 (which we now find directly involved here) and polyA polymerase (also known to be required for efficient processing of mRNA) (Fig. 2). Enolase, which is required for glycolysis and gluconeogenesis, is a puzzling member of the complex. It was therefore suspected to give clues about the functional explanation for the whole setup of this specific complex. This metabolic enzyme possesses a sequence region that does not appear to be indispensable for its catalytic function [10], but it may be missing in case of ribosomal hopping [11]. This would allow the cell to modulate the distribution of this important enzyme among different types of complexes, depending on the environment of the cell. Now, looking for the PNPase protein, we found that it degrades mRNA from their 3' end into NDPs and not NMPs as other RNases do, a feature that had not been readily explained until this new role was uncovered.

A further neighborhood allowed us to elaborate on this scenario (use of synonyms, this time): *cmk* was formerly named *mssA* because it was isolated as a suppressor of a suppressor of a chromosome segregation defect mediated by inactivation of the myosin-like protein MukB. Because MssA is Cmk, this involves directly cytidylate kinase in DNA metabolism, either through its catalytic activity or because it belongs to some complex needed for appropriate DNA segregation. To elaborate on this observation, we started from known RNA degradation processes and identified the cognate genes with similar codon usage bias [3]. In *Escherichia coli*, this yielded the following list of neighbors: *bla*, *cat*,

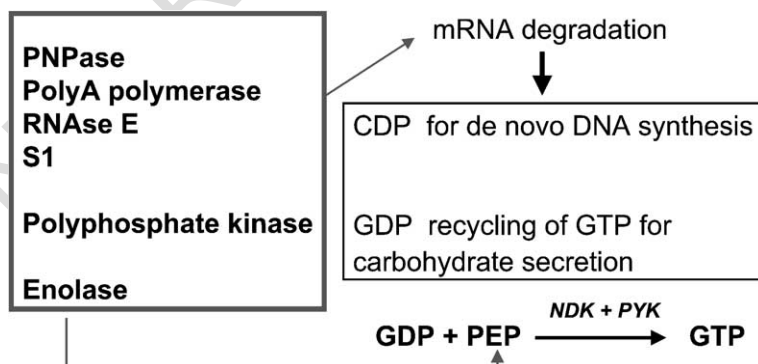


Fig. 2. The degradosome. The degradosome associate RNases and polynucleotide phosphorylase to provide the cell with ribonucleoside diphosphates. CDP is essential for de novo DNA synthesis, while GDP is phosphorylated at the membrane by a complex combining nucleoside diphosphokinase and pyruvate kinase, using PEP provided by the enolase present in the complex.

pyrF, *dicB*, *hflB*, *ftsH*, *mrsACF*, *lpp*, *nusA*, *ompA*, *pcnB*, *pnp*, *rna*, *rnb*, *rnc*, *rnelams*, *rph*, *trxA* and *metY*. Several of these genes obviously correspond to the specific case of a particular mRNA degradation process, identified to be slow in this case (*bla*, *cat*, *dicB*, *lpp*, *ompA*). Among the other genes of the neighborhood control processes, several linked to RNA turnover: RNaseIII (*rnc*, *nusA*, *metY*) RNase H (*rnh*), housekeeping RNases (*rna*, *rnb*), *pnp* (for polynucleotide phosphorylase), *pcnB* (polyA polymerase) or *rne* (for RNase E). For the latter, we found a remarkable neighbor: *rng* (RNaseG, cytoplasmic axial filament protein gene). It has the following noteworthy function (SwissProt entry): “involved in processing of the 5' end of 16S rRNA. Could be involved in chromosome segregation and cell division. It may be one of the components of the cytoplasmic axial filament bundles or merely regulate the formation of this structure”. This relates the RNase activity to chromosome segregation mediated by MukB.

Taken together, these observations allowed us to draw the following picture: coupled to ribosomal protein S1, polynucleotide phosphorylase and polyA polymerase, cytidylate kinase is a key enzyme of intermediary metabolism in bacteria, providing fine coupling between translation, mRNA turnover and DNA synthesis, through control of chromosome separation mediated by MukB, and depending on the availability of dCTP [3,12]. The role of the *cmk rpsA* operon is to provide CDP via CMP for dCDP synthesis. The role of enolase in the complex is to use another dinucleotide produced by polynucleotide phosphorylase, GDP, for synthesis of secreted carbohydrate oligomers [3] (Fig. 2). A general consequence is that mRNA turnover is an important step in CDP availability in bacteria, needed for DNA synthesis [13]. In eukaryotes, the obvious counterpart pathway is the synthesis and turnover of membranes, which are highly abundant there. In this respect, it is interesting to see that rapidly multiplying eukaryotic cells exhibit rapid membrane turnover [14,15]. This suggests that interfering with membrane turnover might control de novo DNA synthesis, thus opening a new avenue for anticancer drugs. This would make the discovery of *T. gondii* attenuation of virulence a first step to further explore pyrimidine biosynthesis for curing many diseases (including viral infections).

3.2. Further pyrimidine metabolism enigma: compartments of the cell

Rather than look for other neighborhoods (this would ask to set up the stage entirely anew), let us stick to pyrimidines. The de novo pyrimidine synthesis pathway showed that, in addition to the apparent absence of CDP for DNA synthesis, the cell readily produces UDP, a dangerous metabolite, since it is recognized by nucleoside diphosphate reductase, making dUDP, which is readily transformed into dUTP by nucleoside diphosphokinase (NDK). Now, dUTP would be readily incorporated into DNA instead of dTTP! The cell had therefore to devise an appropriate means to deal with this dangerous situation, and an enzyme, dUTP pyrophosphohydrolase, cleaves this molecule to dUMP, which serves as a substrate to thymidylate synthetase, making dTMP. This cycle is, of course, energy costly. It may even lead to a futile cycle if uridylate kinase (UMK), coded by the *pyrH* gene, efficiently transforms dUMP into dUDP (and this is indeed a common feature of the enzyme, [16]). This observation points to UMK as a very important enzyme indeed: the question can be solved either by increasing the accuracy of the enzyme (that would discriminate against dUMP while recognizing UMP) or by segregating it at some place.

What do we know about *pyrH*? In bacteria, whether monoderms or diderms, one usually finds the gene cluster *pyrH frr*, where *frr* codes for the essential ribosome recycling factor (once again involving translation) and is cotranscribed with *pyrH* [17] (Fig. 1B). We noted that the codon usage bias of *frr* in *B. subtilis* is remarkable: in addition to genes involved in the construction of the ribosome (*rpmC*, *rpmD*, *rpmH*, *rplW*) and RNA polymerase (*rpoA*), or linked to membrane-associated processes (flagellar hook synthesis *flgE*, *yocH*, muraminidase), it is associated to nucleotide synthesis through *ndk* and, above all, *yosS* which most probably codes for dUTP pyrophosphohydrolase. With the chromosome vicinity of *pyrH*, this makes a strong connection of the role of uracil derivatives in preventing alteration of proper DNA synthesis. *pyrH* itself is connected to genes involved in oxygen stress monitoring and control (*dps*, *yhaH*, *mrgA*, and the neutral protease gene *nprE*) and redox properties (thioredoxins, *trxB* and *ytpP*). The neighbors in *E. coli* are different, but they point to similar processes: translation/transcription (*asnS*, *lysS*, *proS*, *leuS*, *pheST*, *rpoD*), DNA metabolism and membrane insertion (*secA*, *gidC*). *gidC*, which is a ubiquitous protein, was recently identified as a Sec-dependent or Sec-independent membrane protein translocase essential for maintaining viability in *E. coli* [18,19]. Taken together, these observations suggested some kind of relationship between translation, some aspect of membrane localization and DNA metabolism.

It was therefore of interest to analyze the localization of UMK in the cell. Although the enzyme looked like a typical cytoplasmic protein, work developed in the laboratory of Barzu unambiguously demonstrated that, in bacteria, UMK is found in close association with the bacterial envelope [16,20], thus solving the necessary compartmentalization of UDP. A puzzle remained: why is this activity related to ribosome recycling? We could expect that the place where ribosomes have to recycle is the end of the last gene of every operon. This is generally located upstream of a 3' region of the mRNA that forms a so-called 'Rho-independent' stem and loop structure, and regulation of pyrimidine synthesis involves explicitly such a structure [21]. These terminators are uracil-rich and they must therefore consume a significant amount of UTP, yielding UDP. Availability of pyrimidine nucleotides regulates their synthesis by attenuation, suggesting that indeed they may be limiting, and it has explicitly been demonstrated that ribosomes are involved in this essential regulatory process [22]. Is that a sufficient selection pressure for what could be an association of UMK to the ribosomes? Does this mean that terminating ribosomes are frequently associated to the cell membrane? This is the conclusion that this investigation of neighborhoods would propose.

3.3. How do genome parasites evolve?

The conclusion we reached is that there is a strong pressure in all living organisms for availability of cytosine containing nucleotides, at least with the present data we possess on pyrimidine biosynthesis (the conclusion may be that G+C-rich organisms have a special pathway for this). In terms of phylogenetic neighborhoods, these constraints are so strong that they are likely to be reflected in the G+C content of genomes [23]. The G+C content of bacterial genomes varies from 25% to 75%, but the reason for this variation is still unclear. A way to evaluate the contribution of intermediate metabolism to this variation was to analyze the behavior of genetic elements reproducing inside the cell (plasmids,

phages, insertion sequences). Using the cell's replicating machinery, they were expected to reflect the G+C content of the host. Exceptions to this rule were thought to indicate recent horizontal transfer, and detecting unusual G+C composition bias could be used in the identification of pathogenicity islands in bacterial genomes. Naturally, such horizontally transferred elements are progressively assimilated into the average nucleotide composition of the host genome, so that we can only witness the most recent events [23]. This observation, however, gives credence to the idea that studying the neighborhood of horizontally transferred elements is worth exploring. All data tend to substantiate the observation that laterally transferred genetic elements, parasites and commensals do indeed have a more A+T-rich DNA composition than their hosts. An interesting consequence, because this imposes a different bias in the amino acid composition of the corresponding proteins, was that this very inescapable constraint could be used by the cell for constructing special compartments [24]. Here again we see that it becomes interesting to study what could be the driving forces of cell compartmentalization.

3.4. The sulfur enigma

Here is an example. A general way to explore the validity of the neighborhood approach is to start with cases where proximity in the chromosome is not readily understood. This is the case, in particular, with operons involved in the central replication, transcription or translation machineries, that also contain genes apparently involved in intermediary metabolism and/or other processes. A case in point is the trio of genes *secY* *adk* *map*, which is situated at the end of the very long *rpsJ* operon in A+T-rich bacilli (Lu et al., submitted for publication) (Fig. 1C). SecY is involved in secretion, which is frequently co-translational in bacteria; its position near ribosomes is therefore understandable. Adk is adenylate kinase, an enzyme essential for recovering ADP (then ATP) from AMP: translation consumes much ATP and loading of amino acids on tRNA synthetases yields AMP. Finally, Map, methionine aminopeptidase, cleaves about half of the N-terminal methionines of proteins (those whose second residue is a small amino acid). It could therefore belong to a large complex surrounding the ribosomes. The situation, however, is different in A+T-rich cocci: there, *secY* and *adk* genes are still distal to *rpsJ* in the cognate operon, but *map* has disappeared, to be replaced by the counterpart of another *map*-like gene present in *B. subtilis*, *yflG*.

We made use of the in biblio neighborhood to further explore this situation. Combination of "SecY" and "translation" yielded 28 references in PubMed: the first one was highly relevant to our study [25]. Indeed, the corresponding work demonstrated that SecY is involved in protein translocation concomitant with protein synthesis in a way that is essential for the cell's architecture. This provided a concrete link between translation and secretion, and suggested that the cell architecture plays a role. Now, YflG is the counterpart of the bacilli Map in cocci, which also suggested a link with the architecture of the cell. One therefore expected to find a link between some feature of methionine aminopeptidase, translation and the cell's shape. But what about Adk? ATP synthesis is required for all cell processes, so why is the *adk* gene conserved in this ribosomal protein operon? Once again, an in biblio neighborhood provided a challenging tentative answer. Using "sensing" "ATP" and "promoter" yielded 13 references, one of which was highly relevant to our

topic [26]. Gourse et al. indeed proposed that ATP is the sensor of the cell's state that controls ribosome synthesis. The presence of adenylate kinase in a ribosomal protein operon is an extremely interesting way to create a feedback loop in the system. We were therefore led to make the hypothesis that this explained the selective pressure that maintains *adk* in the *rpsJ* operon. Appropriate mathematical models should help to clarify whether this is a reasonable interpretation of this observation. The unavoidable conclusion of this exploration of neighborhoods, remembering that the bulk of the ATP production is membrane bound, was, once again, that there is a strong link between some feature of translation and association to the membrane. We saw the role of pyrimidines, and now this implies a role of methionine. What do we know about sulfur metabolism?

3.5. Sulfur islands and the architecture of the cell

Sulfur is ubiquitous in life. It has been important since its beginnings, and it is present in all proteins as their first amino acid residue. However, sulfur is extremely reactive since its oxidation state varies from -2 to $+6$. Sulfur will therefore be extremely sensitive to the local oxido-reduction potential of the cell, and in particular, to reactive oxygen species. Hence, we can expect that sulfur metabolism genes will have a tendency to be compartmentalized. As a consequence, if there is a relationship between the genome and the architecture of the cell, sulfur metabolism-related genes should be clustered, as indeed found in the case of *E. coli* [27]. Was this observation general? At the moment, many sulfur metabolism genes are still unknown and this precluded a justified statistical analysis, except in *E. coli*. However, knowledge is building up fast in other bacteria, and especially in *B. subtilis*. There, sulfur anabolic genes are often organized in the so-called S-box operons, making a first cluster [28]. Extending beyond these operons can usually lead to other genes involved in sulfur metabolism, such as in the case of the methionine salvage pathway [29] or the enigmatic operon *yrrT mtnA yrhA yrhB*. The latter is linked to the methionine salvage pathway through nucleosidase MtnA. However, this enzyme is also an *S*-adenosyl-homocysteine (AdoHcy) nucleosidase, leading to release of homocysteine, which can be cleaved into H_2S by YrhA. All this asks for an important source of AdoHcy. Because the biotope of *B. subtilis* is the surface of leaves, it must be exposed to high concentrations of oxygen. This implies that its quinone respiration cofactors must be stabilized. Ubiquinone being the stable methylated form of menaquinone, one expects that methylases play the appropriate role, while producing AdoHcy: the prediction of this metabolic neighborhood would therefore be that because YrrT is similar to methylases (phylogenetic neighborhood), it is the expected enzyme. Protein YrhH, in the vicinity of YrrT (genome neighborhood), is also a putative menaquinone methyltransferase (metabolic neighborhood: two methyl groups have to be transferred to make ubiquinone from menaquinone), and this gives credence to the hypothesis. Experimental work will be the only way to explore this hypothesis further. The ultimate interpretation of the data would be that complexes comprising enzymes able to protect sulfur metabolic complexes from oxygen and radical oxygen species have to be involved, perhaps making some kind of protecting shield. The observation that the final stage of the methionine salvage pathway is ensured by a dioxygenase (MtnZ) [29], or that gene YrhJ codes for a protein similar to cytochrome *P450*/NADPH-cytochrome *P450* reductase, substantiates this hypothesis.

4. Conclusion

This short exploration of genome texts demonstrated that the neighbors of gene and gene products tell much about their function. This allowed us to show that a question as essential as de novo DNA synthesis had not yet been explored in full. We further discovered that physical compartmentalization could be predicted from analysis of neighborhoods. We ended in finding out, as a concrete illustration of the approach, that methionine recycling and salvage was likely to have an important role in the cell, far more important than what would simply be expected from a rather ubiquitous and general purpose amino acid.

Acknowledgements

Science is never the work of a single person. The present situation of a world dominated by the role of mass media, including funding research, unfortunately tends to make one forget that, unlike artists, scientists should have no name. If Franklin, Wilkins, Watson and Crick had not discovered the structure of DNA, others would have done so within a few years, or perhaps even a few months time. This is the reason why Stanislas Noria, who is the summary of the ongoing work of many scientists spread throughout the world, is the first author of the present study, physically presented by its representative, AD.

References

- [1] R.L. Tatusov, D.A. Natale, I.V. Garkavtsev, T.A. Tatusova, U.T. Shankavaram, B.S. Rao, B. Kiryutin, M.Y. Galperin, N.D. Fedorova, E.V. Koonin, The COG database: new developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Res.* 29 (1) (2001) 22–28.
- [2] A. Danchin, From protein sequence to function, *Curr. Opin. Struct. Biol.* 9 (3) (1999) 363–367.
- [3] P. Nitschké, P. Guerdoux-Jamet, H. Chiapello, G. Faroux, C. Hénaut, A. Hénaut, A. Danchin, Indigo: a world-wide-web review of genomes and gene functions, *FEMS Microbiol. Rev.* 22 (4) (1998) 207–227.
- [4] B. Fox, D. Bzik, De novo pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii*, *Nature* 415 (2002) 926–929.
- [5] J. Stubbe, Ribonucleotide reductases: amazing and confusing, *J. Biol. Chem.* 265 (10) (1990) 5329–5332.
- [6] A. Shteiman-Kotler, G. Schuster, RNA-binding characteristics of the chloroplast S1-like ribosomal protein CS1, *Nucleic Acids Res.* 28 (17) (2000) 3310–3315.
- [7] C. Sugita, M. Sugiura, M. Sugita, A novel nucleic acid-binding protein in the cyanobacterium *Synechococcus* sp. PCC6301: a soluble 33-kDa polypeptide with high sequence similarity to ribosomal protein S1, *Mol. Gen. Genet.* 263 (4) (2000) 655–663.
- [8] I.V. Boni, V.S. Artamonova, N.V. Tzareva, M. Dreyfus, Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1, *EMBO J.* 20 (15) (2001) 4222–4232.
- [9] Y. Feng, H. Huang, J. Liao, S.N. Cohen, *Escherichia coli* poly(A)-binding proteins that interact with components of degradosomes or impede RNA decay mediated by polynucleotide phosphorylase and RNase E, *J. Biol. Chem.* 276 (34) (2001) 31651–31656.
- [10] A. Hénaut, F. Lisacek, P. Nitschké, I. Moszer, A. Danchin, Global analysis of genomic texts: the distribution of AGCT tetranucleotides in the *Escherichia coli* and *Bacillus subtilis* genomes predicts translational frameshifting and ribosomal hopping in several genes, *Electrophoresis* 19 (4) (1998) 515–527.

- [11] A.J. Herr, C.C. Nelson, N.M. Wills, R.F. Gesteland, J.F. Atkins, Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene 60 bypassing signals during ribosome slippage on mRNA, *J. Mol. Biol.* 309 (5) (2001) 1029–1048. 361
362
363
- [12] A. Danchin, Comparison between the *Escherichia coli* and *Bacillus subtilis* genomes suggests that a major function of polynucleotide phosphorylase is to synthesize CDP, *DNA Res.* 4 (1) (1997) 9–18. 364
365
- [13] T.M. Wright, H.S. Shin, D.M. Raben, Sustained increase in 1,2-diaclyglycerol precedes DNA synthesis in epidermal-growth-factor-stimulated fibroblasts. Evidence for stimulated phosphatidylcholine hydrolysis, *Biochem. J.* 267 (2) (1990) 501–507. 366
367
368
- [14] R.B. Cornell, A.F. Horwitz, Apparent coordination of the biosynthesis of lipids in cultured cells: its relationship to the regulation of the membrane sterol:phospholipid ratio and cell cycling, *J. Cell Biol.* 86 (3) (1980) 810–819. 369
370
371
- [15] S. Jackowski, Coordination of membrane phospholipid synthesis with the cell cycle, *J. Biol. Chem.* 269 (1994) 3858–3867. 372
373
- [16] L. Serina, C. Blondin, E. Krin, O. Sismeiro, A. Danchin, H. Sakamoto, A.M. Gilles, O. Barzu, *Escherichia coli* UMP-kinase, a member of the aspartokinase family, is a hexamer regulated by guanine nucleotides and UTP, *Biochemistry* 34 (1995) 5066–5074. 374
375
376
- [17] S.L. Wadskov-Hansen, J. Martinussen, K. Hammer, The *pyrH* gene of *Lactococcus lactis* subsp. *cremoris* encoding UMP kinase is transcribed as part of an operon including the *frl1* gene encoding ribosomal recycling factor 1, *Gene* 241 (1) (2000) 157–166. 377
378
379
- [18] M.L. Urbanus, L. Froderberg, D. Drew, P. Bjork, J.W. de Gier, J. Brunner, B. Oudega, J. Luirink, Targeting, insertion, and localization of *Escherichia coli* YidC, *J. Biol. Chem.* 277 (15) (2002) 12718–12723. 380
381
- [19] M.O. Kebir, D.A. Kendall, SecA specificity for different signal peptides, *Biochemistry* 41 (17) (2002) 5573–5580. 382
383
- [20] S. Landais, P. Gounon, C. Laurent-Winter, J.C. Mazić, A. Danchin, O. Barzu, H. Sakamoto, Immunochemical analysis of UMP kinase from *Escherichia coli*, *J. Bacteriol.* 181 (3) (1999) 833–840. 384
385
- [21] E.R. Bonner, J.N. D’Elia, B.K. Billips, R.L. Switzer, Molecular recognition of *pyr* mRNA by the *Bacillus subtilis* attenuation regulatory protein PyrR, *Nucleic Acids Res.* 29 (23) (2001) 4851–4865. 386
387
- [22] K.L. Roland, C.G. Liu, C.L. Turnbough Jr., Role of the ribosome in suppressing transcriptional termination at the *pyrBI* attenuator of *Escherichia coli* K-12, *Proc. Natl. Acad. Sci. U. S. A.* 85 (19) (1988) 7149–7153. 388
389
- [23] E. Rocha, A. Danchin, Base composition bias in genomes might result from competition for scarce metabolic resources, *Trends Genet.* (in press). 390
391
- [24] P. Guerdoux-Jamet, A. Hénaut, P. Nitschké, J.L. Risler, A. Danchin, Using codon usage to predict genes origin: is the *Escherichia coli* outer membrane a patchwork of products from different genomes? *DNA Res.* 4 (4) (1997) 257–265. 392
393
394
- [25] M. van der Laan, E.N. Houben, N. Nouwen, J. Luirink, A.J. Driessen, Reconstitution of Sec-dependent membrane protein insertion: nascent FtsQ interacts with YidC in a SecYEG-dependent manner, *EMBO Rep.* 2 (6) (2001) 519–523. 395
396
397
- [26] M.M. Barker, R.L. Gourse, Regulation of rRNA transcription correlates with nucleoside triphosphate sensing, *J. Bacteriol.* 183 (21) (2001) 6315–6323. 398
399
- [27] E.P. Rocha, A. Sekowska, A. Danchin, Sulphur islands in the *Escherichia coli* genome: markers of the cell’s architecture? *FEBS Lett.* 476 (1–2) (2000) 8–11. 400
401
- [28] F.J. Grundy, T.M. Henkin, The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria, *Mol. Microbiol.* 30 (4) (1998) 737–749. 402
403
404
- [29] A. Sekowska, A. Danchin, The methionine salvage pathway in *Bacillus subtilis*, *BMC Microbiol.* 2 (2002) (in press). 405
406