

# Cells need safety valves

Antoine Danchin\*

Institut Pasteur, Genetics of Bacterial Genomes, CNRS, rue du Docteur Roux, Paris Cedex, Paris, France

**In *Escherichia coli*, the role of *lacA*, the third gene of the lactose operon, has remained an enigma. I suggest that its role is the consequence of the need for cells to have safety valves that protect them from the osmotic effect created by their permeases. Safety valves allow them to cope with the buildup of osmotic pressure under accidental transient conditions. Multidrug resistance (MDR) efflux, thus named because of our anthropocentrism, is ubiquitous. Yet, the formation of simple leaks would result in futile influx/efflux cycles. Versatile modification enzymes with low sensitivity solve the problem if the modified metabolite is the one exported by MDR permeases. This may account for the pervasive presence of acetyltransferases, such as LacA, associated to acetyl-metabolite exporters. This scenario of constraints imposed by efficient influx of metabolites provides us with a model that should be followed when constructing synthetic cells.**

**Keywords:** bistability; genetic addiction; operon; permease; synthetic biology

## Introduction

The aim of synthetic biology (SB) is to take an engineering stance to (re)construct cells in a rational and reproducible manner.<sup>(1)</sup> The baseline of this engineering philosophy is the combinatorial association of building blocks based on the abstract rules which have been established over the years governing the control of transcription. Engineering combines an abstract reflection on the properties of the organisation of building blocks with explicit implementation of material structures.<sup>(2)</sup> Curiously, some explicit engineering constraints raised by such material implementation have not yet been considered, while they may already be quite visible in what we know from our previous studies of model bacteria. This apparent ignorance may well be the consequence of a blind spot in our description of the paradigm of gene expression, the *E. coli* lactose operon. With its three genes and its repressor, the lactose operon is presented in all university textbooks.<sup>(3)</sup> It

is also a paradigm for the design of standard biobricks of the SB effort, seen as the paradigm of a logical system whose expression is driven by a Boolean function: lactose AND NOT glucose.

Remarkably, nobody appeared to be motivated to uncover the explicit function of the third gene of the operon, *lacA*. This is in contrast with the many works discussing and placing in relevant perspective (including the most complex idiosyncratic features) the function of the repressor, the catabolite activator protein and the associated cyclic AMP regulatory molecule, and the first two genes of the operon, coding for a  $\beta$ -galactosidase (LacZ) and a proton-driven  $\beta$ -galactoside permease (LacY).<sup>(4)</sup> An early example of this lack of interest is the review of the interactions mediating the control of cellular metabolism by Jacques Monod himself (one of the inventors of the operon model), where he described the lactose operon, omitting its *lacA* gene.<sup>(5)</sup> This lack of interest was substantiated experimentally when it was observed that in some natural plasmids coding for a lactose operon, *lacA* was absent.<sup>(6)</sup> Strangely, this avowal of ignorance remained enough to satisfy the curiosity of investigators, educators and laypersons alike.

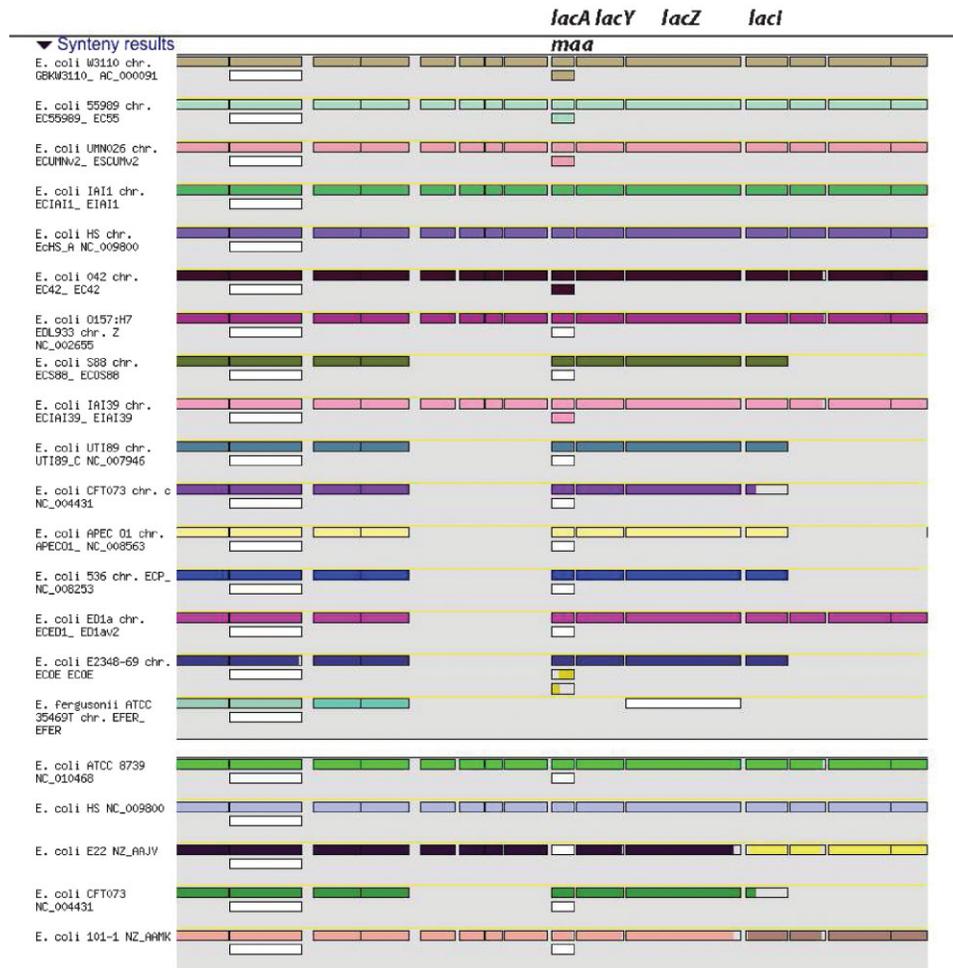
Yet, the complete lactose operon (including *lacA*) appears to be a marker of *E. coli* as forming an authentic bacterial species. Indeed a thorough analysis of environmental, commensal and pathogenic *E. coli* strains, as compared with their closest relative *E. fergusonii* (which is not distinguishable by its 16S RNA only), showed that the vast majority, including genomes where the local synteny is disrupted, comprised a complete lactose operon with its *lacA* gene (see as an exception *E. coli* E22 NZ\_AAJV)<sup>(7)</sup> (Fig. 1). Furthermore, in most instances the genome comprised a gene coding for another carbohydrate acetyltransferase, Maa, which is highly related to LacA (this is the case in strain E22 NZ\_AAJV), placed at another location in the genome sequence.<sup>(8)</sup>

It seems really remarkable that the blindness displayed by the community of investigators about the presence of a gene in an operon that plays the role of a paradigm in SB (advocated almost one century ago as a research program meant to understand what life is, by a somewhat unusual personality, Stéphane Leduc<sup>(9)</sup>) happens at a time when SB is spreading to biology as an extension of the principles of engineering.<sup>(2,10)</sup> Our knowledge about LacA is limited: based on *in vitro* biochemical experiments using a variety of galactosides as substrates, LacA is assumed to code for a

**Abbreviations:** MDR, multidrug resistance; SB, synthetic biology.

\*Correspondence to: A. Danchin, Institut Pasteur, Genetics of Bacterial Genomes, CNRS URA2171, 28, rue du Docteur Roux, 75724 Paris Cedex 15, Paris, France.

E-mail: antoine.danchin@normalesup.org



**Figure 1.** Comparative genomics of the lactose operon region in a variety of *E. coli* strains, with *E. fergusonii* as the closest outgroup, displayed using the MaGe platform.<sup>(37)</sup> W3110 is the *E. coli* K12 reference. Note that both upstream and downstream of the lactose operon there may be considerable variation between different strains.

lactose acetyltransferase displaying an unusually low affinity for its substrates,  $\beta$ -galactosides,<sup>(11)</sup> and is interpreted to be involved in detoxification.<sup>(12)</sup> Could we, by simple reasoning, predict an expected function for this gene? Can we go further in terms of engineering?

## Using the lactose operon as a reference set of biobricks

Forty years after the discovery of the lactose operon, many articles are still based on the underlying physics and logics of this remarkable genetic setup.<sup>(13–17)</sup> It is therefore worth analysing the explicit and implicit hypotheses made when the lactose operon is used as a theoretical or an experimental model. Interestingly, in a work combining *in silico* modelling with experiments,<sup>(15)</sup> an equation placed in the Supplementary Material of the corresponding article (which is the typical

place where one finds authentic engineering constraints) shows that a leak or a degradative process is essential to permit realistic modelling of the multistability process in the lactose utilisation network. However, no physical process is proposed to account for this leak, while it is generally accepted that bacterial membranes are fairly impermeable to hydrophilic compounds. Even more interesting is a more recent and deeper model which provided a detailed account of the well-investigated bistability of the lactose operon, showing that bistability is easy to account for when artificial inducers (non-metabolisable alkylated thiogalactosides) are used, while it is not expected to be a common feature when bacteria are grown on their natural substrate, lactose.<sup>(16)</sup>

After having given a short introduction to the lactose operon (including the *lacA* gene) the authors immediately start constructing a very detailed model in which *lacA* is purely and simply omitted. The model rests on a series of realistic assumptions, and on some utterly unrealistic ones. In

particular the model is based on *reversible* (my emphasis) lactose influx by permease. A further assumption states that there is indeed a significant passive lactose efflux, explicitly claimed to be independent of the presence of glucose (which itself is also permitted to leak out of the cell) and differing whether one considers lactose or its alkylated thiogalactoside analogues.<sup>(16)</sup> The overall description of bistability, emphasising differences between artificial and natural substrates, is entirely dependent on these premises.

A leak is certainly not the ultimate refinement in engineering. It is difficult to control and to stabilise and it is a waste. In most engineering setups where there is an influx system, appropriate ratchet-like processes prevent straightforward reversibility, even though reversibility is in principle physically permitted. Assuming that the LacY permease is passively reversible does not fit what is known of its function: a proton-driven pump that bases its activity on the considerable electrochemical potential of the cell. Actualisation of its reversibility would imply a considerable load on the availability of energy in the cell.<sup>(18)</sup>

As we shall see below, there exist energy-independent reversible permeases (often named facilitators<sup>(19)</sup>), but this is not the case for LacY, which, although being a member of the large class of facilitators, is alternating an open conformation at either face of the cytoplasmic membrane as a function of co-permeation of protons together with its galactoside substrates.<sup>(20)</sup> As a consequence, if lactose leaked out of *E. coli* cells in an uncontrolled manner, the presence of LacY would create an energy-consuming futile cycle, especially as with the permease being located in the cytoplasmic membrane, the leaking lactose would tend to remain fairly concentrated in the cell's periplasm. The same line of reasoning also holds for glucose, even though its permeation system differs from that of lactose (but it is also energy-driven, *via* the phosphoenolpyruvate phosphotransferase system). Taken together, these observations suggest that we miss a function. On the other hand, we have an extra gene, *lacA*, for which we do not have a really convincing function. Let us try to match these discrepancies.

## Predicting the authentic LacA function and discovery of safety valves

All kinds of neighbourhoods may be used to make functional inferences.<sup>(21,22)</sup> Much work about the gene neighbourhoods has suggested that when proximity is conserved in the genome, there is good chance that there is some kind of structural or functional relationship between the proteins coded in the corresponding genes.<sup>(23–25)</sup> Conservation of *lacA* in the vicinity of *lacZ* and *lacY* therefore suggests that, despite its current dispensability, it may have a function related to that of the lactose operon. Furthermore, it has been

established that there is a polarity in gene expression in the lactose operon, with *lacZ* expressed at a significantly higher level than *lacY* and *lacA*, which go together.<sup>(26)</sup> This permits us to infer that the function of *lacA* might be related to that of *lacY*.

Engineers are keen to make a mechanical and logical assessment of the machines they are constructing, trying not to forget unobtrusive defects in design. Thinking of a cell which would use lactose as a carbon and energy source in terms of engineering leads to the following paradox. In their usual environment, the mammalian gut, *E. coli* cells are prone to meet variable concentrations of lactose. A common situation met by the bacteria could be some lactose inducing the *lac* operon, an intermediate period with not much of the sugar, and suddenly a large concentration of it. According to the commonly accepted model, the cell's response to the final situation would be to trigger efficient influx of the sugar, because the operon had previously been induced to place lactose permease in the membrane. Because LacY allows the intracellular lactose concentration to climb up considerable gradients, it might sometimes reach levels that would build up an osmotic pressure which would lead to the bursting of cells or at least interfere with the building up of murein (Gram-negative organisms do not have a thick peptidoglycan envelope), unless some protective mechanism exists, typically the safety valve that engineers place on steam engines. Some type of leak is indeed expected to cope with these deleterious situations.

In keeping with the SB view of biology, a natural way to engineer the process in a way that would not interfere with the primary goal of the cell – to accumulate enough lactose to use it as a carbon source – would be to modify lactose into another component, progressively preventing it from reaching a deleterious concentration, and then to excrete the modified compound, which would not go back into the cell. The first part of the process fits exactly the function of LacA with its surprisingly poor  $K_M$ ,<sup>(11)</sup> which is now explained as nicely contributing to fitness (it is in the molar range) by making acetyl-galactosides. What, then, would be the fate of acetyl-lactose? Do we know whether it would be expelled from the cell? While this possibility was not explored and remained unknown for a long time, three permeases, SetA, SetB and SetC, which belong the multidrug resistance (MDR) permeases family, perform this role in *E. coli*. They can also export other related acetylated sugars.<sup>(27)</sup> This function nicely fits the requirement for a safety valve.

Let us also note that, in another domain, the existence of these efflux permeases implies that all the past work on *in vivo* efflux of galactosides from the cell, performed prior to this important work, should be analysed retrospectively. Indeed the efflux of acetylated carbohydrates discovered accounts fairly well for the puzzling passive efflux of modified galactosides (alkylated thiogalactosides) when the two main energy-yielding pathways are poisoned<sup>(28)</sup>: they would not be

expelled from the cell by LacY, but by one or several of SetA, SetB or SetC.

Do we have further arguments suggesting that there is indeed, under some circumstances, a steady-state or an overflow problem driven by the efficiency of LacY? While there is as yet no direct experimental evidence proving the conjecture presented here (it would require inactivation not only of the three *setA*, *setB* and *setC* genes, but the maltose acetyltransferase gene *maa* as well), several pieces of work provide strong support to the idea.

- (1) Overflow: A considerable proportion of *E. coli* cells (80–98%) taken from a lactose-limited chemostat die when plated on standard lactose minimal media. This effect has been explored in detail and it was shown that the action of rapid transport across the membrane must be the cause of the phenomenon, while the toxic effect was not due to derivatives of lactose.<sup>(29)</sup>
- (2) In the presence of cyclic AMP and in a particular genetic background (namely, constitutively expressing the lactose operon and lacking the phosphoenolpyruvate-dependent glucose phosphotransferase system), cells grown in rich medium were lysed upon addition of lactose, and this was due to lactose influx.<sup>(30)</sup>
- (3) Steady-state conditions: Authors have long looked for ways to isolate mutants of the *lacY* gene, in order to explore the lactose channel and proton co-transport in the permease. At the time of the discovery of the lactose operon, many lactose analogues were discovered with different properties in terms of influx, efflux or control of the lactose operon expression.  $\beta$ -thio-*o*-nitrophenylgalactoside was found to be toxic when the cells were grown in a medium containing succinate as the carbon source, permitting selection of a variety of LacY mutants.<sup>(31)</sup> Taken together, all these experiments fit well in the scenario proposed here, where the lactose permease is so efficient that, under a variety of circumstances, it leads to a toxic accumulation of lactose, requiring some sort of safety valve.

## Why so many acetyltransferase genes in genomes?

This fairly simple line of reasoning has many consequences. A straightforward consequence is that one should study in this new light the considerable number of metabolite acetyltransferases that are present in genomes. They may be used to modify foreign or unusual molecules to prevent them from becoming unwanted substrates of major metabolic processes.<sup>(32)</sup> More generally, scenarios such as the one discussed here should be investigated again with this idea of the need for safety valves in mind. This type of modification/

efflux system should then be included in the basic biobricks of SB. In fact, all models of metabolism involving permeases that are highly efficient should include similar processes as well, and a reflection on those involving ATP-driven transport should probably also take this process into account (or at least make a model of the physical relationships between osmotic pressure, proton-driven ATP synthesis and transport). In this respect it may be significant that the Maa acetyltransferase not only modifies maltose, but acetylates glucose as well.<sup>(8)</sup> It is therefore expected that, for each efficient permeation process, there exists an acetyltransferase associated to an efflux system, accounting for at least some of the large number of acetyltransferase genes found in genomes.

Finally, this reflection places the so-called MDR efflux systems in an entirely new light. While they have initially been found to be involved in acquisition of resistance to a variety of antibiotics, Saier and co-workers have already provided us with convincing evidence that the primary functions of MDRs at their origin are unrelated to resistance to antibiotics or drugs.<sup>(33)</sup> If one takes the present hypothesis seriously, many such proteins must be normally present in most cells, where they play the role of safety valves for normal metabolised substrates. As a consequence, they will also permit to export any structurally related non-metabolised substrate, precisely because the absence of metabolism will result in rapid increase in its intracellular concentration. It therefore becomes obvious how MDRs could rapidly be diverted to play the role of drug resistance devices. This remark also fits well with the observation that, besides MDRs, another large class of antidrug proteins is indeed made by acetyltransferases (*e.g.* chloramphenicol acetyltransferases): acetylation would be a general group modification process that would modify unwanted substrates to divert them to novel metabolic pathways or to excrete them into the external medium. This central function would then have been recruited for another role,<sup>(34)</sup> *i.e.* the protection against harmful molecules such as antibiotics or anticancer drugs.

## A word of conclusion

The lactose operon, with its repressor gene *lacI* and its three genes *lacZYA*, is the paradigm of molecular biology. Surprisingly, it appears that nobody has been curious enough to explore the enigmatic function of the *lacA* gene. An acetyltransferase with a very poor  $K_M$  is hardly a convincing answer for a curious mind. I have suggested here that cells with efficient permeases, also need safety valves, in order to prevent the buildup of excessive osmotic pressure. This would be particularly important in cells lacking a rigid envelope, such as eukaryotic cells. Yet this scenario might result in energy-costly leaks, or provide selection pressure favouring poorly active permeases. Preventing the futile cycles that leaks

would create, seems remarkably illustrated in the organisation of the lactose operon, whose function is to efficiently scavenge a sporadically abundant carbon source. Naturally, any function must have a cost. In the present situation the trade-off could be that the influx/modification/efflux system we have described produces a form of addiction: indeed, it has some abstract properties similar to those of the restriction/modification systems, or the toxin/antitoxin systems.<sup>(35,36)</sup> It will be interesting to explore whether this is used in situations where cooperation between cells is essential, such as in formation of communities, symbiosis, parasitism or pathogenicity.

**Acknowledgments:** This work was supported by the PRO-BACTYS programme, grant CT-2006-029104, and the TAR-POL programme, grant KBBE-2007-212894, in an effort to define genes essential for the construction of a synthetic cell. I would like to thank an anonymous reviewer for triggering me to explore further published data on the toxicity of lactose in *E. coli*.

## References

1. Goodman, C., Engineering ingenuity at iGEM. *Nat Chem Biol* 2008. **4**: 13.
2. Endy, D., Foundations for engineering biology. *Nature* 2005. **438**: 449–453.
3. Hediger, M. A., Johnson, D. F., Nierlich, D. P. and Zabin, I., DNA sequence of the lactose operon: The *lacA* gene and the transcriptional termination region. *Proc Natl Acad Sci U S A* 1985. **82**: 6414–6418.
4. Reznikoff, W. S., The lactose operon-controlling elements: A complex paradigm. *Mol Microbiol* 1992. **6**: 2419–2422.
5. Monod, J., On the mechanism of molecular interactions in the control of cellular metabolism. *Endocrinology* 1966. **78**: 412–425.
6. Guiso, N. and Ullmann, A., Expression and regulation of lactose genes carried by plasmids. *J Bacteriol* 1976. **127**: 691–697.
7. Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 2009. **5**: e1000344.
8. Lo Leggio, L., Dal Degan, F., Poulsen, P., Andersen, S. M. and Larsen, S., The structure and specificity of *Escherichia coli* maltose acetyltransferase give new insight into the LacA family of acyltransferases. *Biochemistry* 2003. **42**: 5225–5235.
9. Leduc, S., La Biologie Synthétique, Paris, A. Poinat, 1912.
10. de Lorenzo, V. and Danchin, A., Synthetic biology: Discovering new worlds and new words. *EMBO Rep* 2008. **9**: 822–827.
11. Musso, R. E. and Zabin, I., Substrate specificity and kinetic studies on thiogalactoside transacetylase. *Biochemistry* 1973. **12**: 553–557.
12. Andrews, K. J. and Lin, E. C., Thiogalactoside transacetylase of the lactose operon as an enzyme for detoxification. *J Bacteriol* 1976. **128**: 510–513.
13. Bronson, J. E., Mazur, W. W. and Cornish, V. W., Transcription factor logic using chemical complementation. *Mol Biosyst* 2008. **4**: 56–58.
14. Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A. and Alon, U., Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* 2006. **4**: e45.
15. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. and Van Oudenaarden, A., Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 2004. **427**: 737–740.
16. van Hoek, M. J. and Hogeweg, P., In silico evolved lac operons exhibit bistability for artificial inducers, but not for lactose. *Biophys J* 2006. **91**: 2833–2843.
17. Veening, J. W., Smits, W. K. and Kuipers, O. P., Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* 2008. **62**: 193–210.
18. Fields, K. L. and Luria, S. E., Effects of colicins E1 and K on transport systems. *J Bacteriol* 1969. **97**: 57–63.
19. Berman-Kurtz, M., Lin, E. C. and Richey, D. P., Promoter-like mutant with increased expression of the glycerol kinase operon of *Escherichia coli*. *J Bacteriol* 1971. **106**: 724–731.
20. Nie, Y., Zhou, Y. and Kaback, H. R., Clogging the periplasmic pathway in LacY. *Biochemistry* 2009. **48**: 738–743.
21. Moszer, I., Glaser, P. and Danchin, A., SubtiList: A relational database for the *Bacillus subtilis* genome. *Microbiology* 1995. **141**: 261–268.
22. Schuler, G. D., Epstein, J. A., Ohkawa, H. and Kans, J. A., Entrez: Molecular biology database and retrieval system. *Methods Enzymol* 1996. **266**: 141–162.
23. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. and Eisenberg, D., A combined algorithm for genome-wide prediction of protein function. *Nature* 1999. **402**: 83–86.
24. Nitschké, P., Guerdoux-Jamet, P., Chiappello, H., Faroux, G., Hénaut, C., et al. Indigo: A World-Wide-Web review of genomes and gene functions. *FEMS Microbiol Rev* 1998. **22**: 207–227.
25. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. and Maltsev, N., The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* 1999. **96**: 2896–2901.
26. Li, Y. and Altman, S., Polarity effects in the lactose operon of *Escherichia coli*. *J Mol Biol* 2004. **339**: 31–39.
27. Liu, J. Y., Miller, P. F., Willard, J. and Olson, E. R., Functional and biochemical characterization of *Escherichia coli* sugar efflux transporters. *J Biol Chem* 1999. **274**: 22977–22984.
28. Bentaboulet, M. and Kepes, A., Counter-transport mediated by the lactose permease of *Escherichia coli*. *Biochim Biophys Acta* 1977. **471**: 125–134.
29. Dykhuizen, D. and Hartl, D., Transport by the lactose permease of *Escherichia coli* as the basis of lactose killing. *J Bacteriol* 1978. **135**: 876–882.
30. Alexander, J. K., Lysis of *Escherichia coli* mutants by lactose. *J Bacteriol* 1979. **140**: 643–648.
31. Varela, M. F., Wilson, T. H., Rodon-Rivera, V., Shepherd, S., Dehne, T. A., et al. Mutants of the lactose carrier of *Escherichia coli* which show altered sugar recognition plus a severe defect in sugar accumulation. *J Membr Biol* 2000. **174**: 199–205.
32. Sekowska, A., Robin, S., Daudin, J. J., Henaut, A. and Danchin, A., Extracting biological information from DNA arrays: An unexpected link between arginine and methionine metabolism in *Bacillus subtilis*. *Genome Biol* 2001. **2**: RESEARCH0019.
33. Saier, M. H., Jr., Paulsen, I. T., Sliwinski, M. K., Pao, S. S., Skurray, R. A., et al. Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *FASEB J* 1998. **12**: 265–274.
34. Caetano-Anolles, G., Kim, H. S. and Mittenthal, J. E., The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc Natl Acad Sci U S A* 2007. **104**: 9358–9363.
35. Fozo, E. M., Hemm, M. R. and Storz, G., Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev* 2008. **72**: 579–589.
36. Hoskisson, P. A. and Smith, M. C., Hypervariation and phase variation in the bacteriophage 'resistome'. *Curr Opin Microbiol* 2007. **10**: 396–400.
37. Vallet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., et al. MaGe: A microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 2006. **34**: 53–65.