

***Searching for Molecular Solutions* – Additional Material**

CHAPTER 4

These Files contain additional material most relevant to **Chapter 4** of *Searching for Molecular Solutions* (but also relevant to Chapter 9, as below). The page numbers of the book pertaining to each section are shown in the Table below, the corresponding page number for this file, and the title of each relevant section.

Contents:

Most Relevant Book Page Numbers	Page Number in this File	Section	
		No.	Title
103-105; 337	2	A5	Ebg System

Section A5: *Ebg System*

This material is relevant to pp. 103-105 of *Searching for Molecular Solutions*, as an example of a system spanning the 'classical' period of directed strain modifications into the modern era of Directed Evolution. In its description of the engineering of an integrated genetic system, this topic is also relevant to the brief discussion of systems biology in Chapter 9 (from p. 337).

The Ebg System as a Model for Directed Evolution

Directed evolutionary experiments with fungi and bacteria have been conducted for decades, exploiting natural diversity arising within very large populations of cells^{1,2}. Also, by their natures it is not difficult to treat microorganisms with chemical mutagens to globally increase their mutation rates, often an advantage. A long-standing industrial aim has been the betterment of fungal strains producing useful compounds (especially antibiotics³). Directed evolution approaches in such organisms prior to the advent of recombinant DNA technology constitute one area in the general field of 'classical strain improvement'⁴. Let's consider in some more detail a case of directed evolution in the laboratory bacterium *E. coli* which reaches back into the 'classical' era but has continued beyond it, and has a number of interesting ramifications for the present time. This example involves enzymes with β -galactosidase activity, the ability to hydrolyze the sugar lactose.

E. coli strains which have irreversibly lost β -galactosidase capability through large deletions in the gene which encodes it (*lacZ*) should theoretically be completely unable to grow on a medium whose only carbon source is lactose (a substrate of β -galactosidase). Since a sizable chunk of the *lacZ* coding sequence is gone in such cases, reversion to normal β -galactosidase activity is not possible by *lacZ*

reconstitution. Nevertheless, it was reported in 1973 that *E. coli* strains with such *lacZ* deletions could acquire the ability to use lactose after repeated rounds of selection on lactose media ⁵. Moreover, such cells were shown to possess β -galactosidase activity, but it was clearly not the same β -galactosidase as that encoded by *lacZ*. This enzyme within the selected cells, dubbed 'Ebg' for 'evolved β -galactosidase', had different enzymatic and antigenic properties to the *lacZ* protein, and mapped to a different region of the *E. coli* chromosome ⁵. The Ebg system was subsequently studied by others (especially Barry Hall and colleagues) over a thirty-year period. It was shown that a single mutation in the Ebg enzyme (encoded by the *ebgA* and *ebgC* genes [♥]) allowed sufficient β -galactosidase activity for cells to grow (slowly) on lactose as a sole carbon source ⁸. (Sequencing studies showed that the *ebgA* and *lacZ* genes had a definable, albeit ancient, evolutionary relationship ⁸). This information alone does not convey the full picture, since it is necessary to understand the genetic circuitry involved. To do this, we must first revisit the 'regular' lactose-processing genetic arrangement in this bacterium.

The ordinary metabolism of lactose by *E. coli* involves a well-studied genetic control unit called the lactose (*lac*) operon (noted briefly in Chapter 6 of *Searching for Molecular Solutions*), which consists of three functional coding sequences (the *lacZ*, *lacY*, and *lacA*) genes. As we have seen, the *lacZ* gene encodes β -galactosidase, and *lacY* encodes lac permease which enables lactose to enter the cell. (The remaining gene, *lacA*, encodes a transacetylase enzyme which need not concern us further in this context). Each of these *lac* genes is controlled by a regulatory gene (*lacI*) which encodes the lactose repressor. This repressor protein binds to a specific DNA sequence termed the *operator*, which

[♥] Unlike the β -galactosidase encoded by the *lacZ* gene, the Ebg enzyme is composed of two separate types of subunits, EbgA and EbgC. The *ebgC* gene overlaps with the 3' end of the *ebgA* gene ⁶. LacZ β -galactosidase is a homotetramer (consisting of four identical subunits, each the product of the *lacZ* gene ⁷), while the Ebg enzyme is a hetero-octamer of four α subunits (encoded by *ebgA*) and four β subunits (encoded by *ebgC*) ⁸.

overlaps the promoter sequences which enable the operon genes to be transcribed by RNA polymerase[▼]. We have already noted that the lactose-hydrolyzing Ebg mutant enzyme was isolated in a strain with a *lacZ* deletion, so what is the relevance of the *lac* operon to the Ebg system? It is through the *lacY* gene product, the *lac* permease enzyme, which mediates transport of lactose into the bacterial cell itself. Without this, intracellular lactose levels are too low to allow growth even if a β -galactosidase enzyme is present, and the Ebg system has no permease-equivalent.

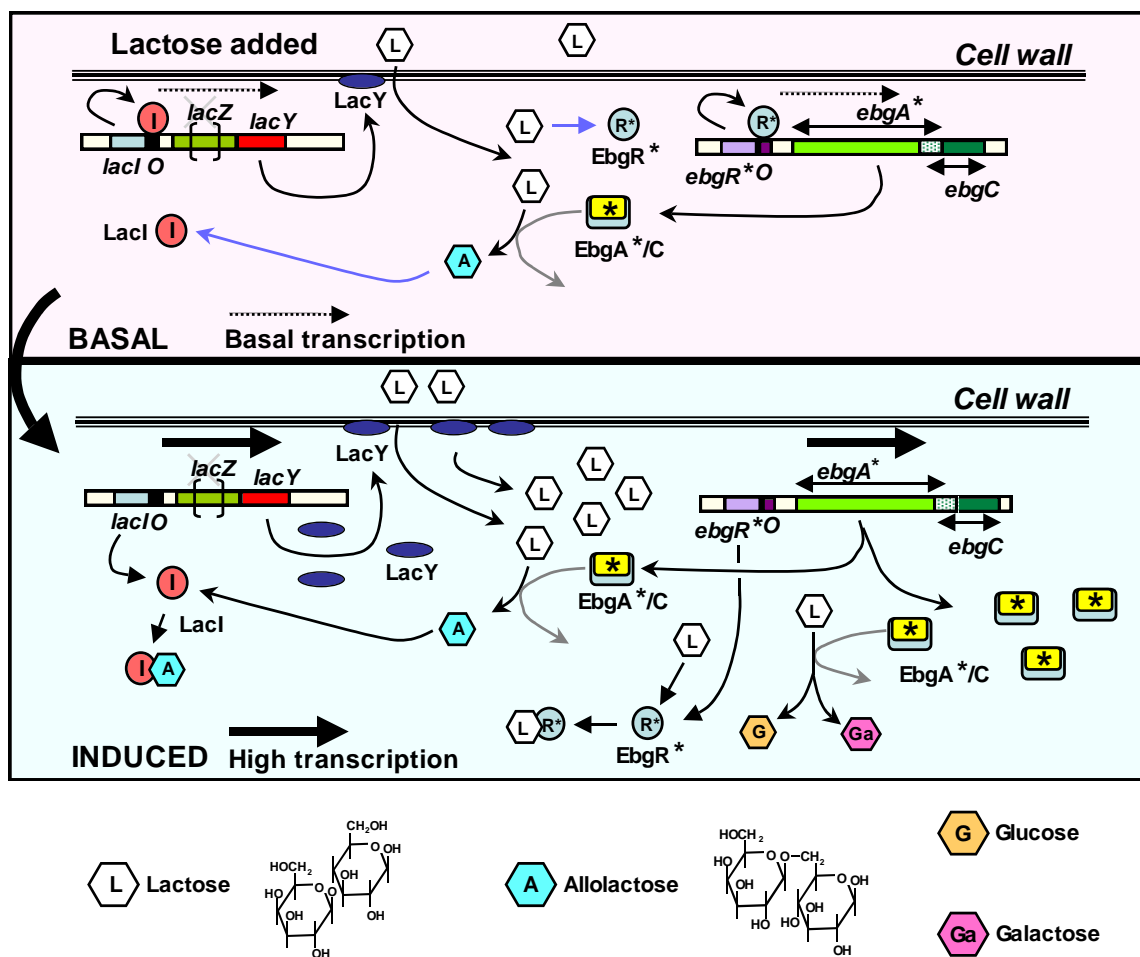


Fig. 4A5.1

[▼] Two additional operator sites are also present in the *lac* operon which increase the effectiveness of repressor action⁹.

Fig. 4A5.1. Genetic circuitry between the lac operon (with a *lacZ* gene deletion) and mutated Ebg operon of the 'Class IV' type ⁸, shown in the transition from the basal state (uninduced, top panel) to fully induced state (bottom panel), through the addition of lactose. The Ebg enzyme is encoded by the *ebgA* and *ebgC* genes as a heteromeric protein, and as a β -galactosidase, efficiently generates galactose and glucose from lactose substrate when fully induced (bottom panel). The repressor of Ebg (encoded by the *ebgR* gene) is normally poorly induced by lactose, but a mutant (shown as R*) is much more efficiently lactose-induced. Class IV mutant Ebg enzyme (shown as EbgA*/C) can produce allolactose from lactose, which binds to the lac repressor (shown as lacI) and induces the lac operon. (The key events during the induction phase of repressor inactivation (through lactose binding to the EbgR* repressor protein, and allolactose binding to the lacI repressor) are shown with blue arrows in the top panel. Although in this case the lac operon has an extensive *lacZ* deletion, the *lacY* gene product (lac permease) is intact and expressible, allowing lactose to efficiently enter the cell. Basal (low-level) expression of lacY protein and EbgA*/C proteins (top panel) allow the induction process to commence.

The key feature of the lac regulatory system is that the lac repressor binds to the operator sequence (and prevents operon transcription) only in the *absence* of lactose; this ensures that the lac genes are only expressed when needed. Thus in the presence of lactose, the repressor undergoes a change in conformation and no longer binds its operator, allowing the lac operon to be transcribed and expressed. (This process is an example of *allostery*, noted in *Searching for Molecular Solutions* especially in the context of functional nucleic acids). This is logically satisfying, but one additional factor serves to complicate the arrangement. Molecules which bind repressors and change their DNA-binding properties are *inducers*, but it eventuates that lactose itself does not directly bind the lac repressor. The direct inducer turns out to be an isomer of lactose, *allolactose*, which differs in the sites of chemical bonding between the glucose and galactose subunits constituting the disaccharide (Fig. 4A5.1). Allolactose is formed by the action of β -galactosidase on lactose itself (this enzyme can

perform this action as well as break down lactose into monosaccharides ¹⁰). It might seem paradoxical that this is so: if the lac repressor was 100% efficient in blocking lac operon expression in the absence of lactose, then theoretically no β -galactosidase should be present to convert lactose into allolactose when the former sugar appears. And the lac operon would accordingly be frozen in an 'off' state, and uninducible. The answer, unsurprisingly, is that the repression is less than perfect, and low-level 'leakage' of β -galactosidase expression occurs which enables switching to the 'on' state to occur when lactose becomes available [♥]. It has been known for many years that induction of the lac operon by allolactose can be bypassed experimentally with 'gratuitous' inducers which bind the lac repressor. The gratuitous inducer isopropyl- β -D-thiogalactopyranoside (abbreviated to IPTG for obvious space-saving reasons) has been widely used in many experimental and biotechnological systems using the lac repressor to control gene expression.

Now, the Ebg system itself is under the control of a repressor, encoded by the *ebgR* gene. Since lactose is not the natural substrate of the Ebg system ^{*}, it is

[♥] There is a message here in the shaping of biological systems by evolution. The optimum solution for a complex system as a whole may not necessarily involve 'optimized' interactions for each of its components. Consider that non-natural symmetric lac operators have been identified which are bound with higher affinity than normal operators by wild-type lac repressor ^{11,12}. Conversely, mutant lac repressors have been found with much tighter binding to the wild-type lac operator than normal ¹³. Though these are 'improvements' by one definition, the wild-type lac operon as a whole is presumably 'tuned' by a balance-sheet of factors which in combination determine optimal fitness. One such factor is the need to allow low-level 'leakage' expression of β -galactosidase. Experimental evidence for this view has been obtained ¹⁴. These considerations are reminiscent of the evolutionary optimization of receptor-ligand interactions, which will not necessarily move to the highest attainable affinity (See Chapter 3 of *Searching for Molecular Solutions* for more on this theme in the context of immune receptors).

^{*} The natural substrate of the Ebg enzyme is still uncharacterized. Although *E. coli* cells with deletions of Ebg ⁸ are fully viable, the long evolutionary persistence of Ebg indicates that it must confer a fitness benefit under some circumstances not seen under laboratory culture conditions.

then not surprising that lactose is a poor inducer of the Ebg operon. Thus in order to grow effectively on lactose even when the Ebg enzyme had undergone mutation(s) which permitted lactose utilization, it is also necessary to circumvent the action of the *ebgR* repressor product. One way for this to occur is simply for mutational inactivation of the *ebgR* product, such that its repression is removed and Ebg enzyme production becomes constitutive. A combination of repressor inactivation and Ebg mutation is therefore one pathway towards lactose utilization in the absence of normal *lacZ*-encoded β -galactosidase. But this alone is not enough, because some mechanism for transporting lactose into the bacterial cell is still required. As we have seen above, this is provided by the *lacY* permease enzyme, but in the absence of the *lacZ* β -galactosidase lactose will not be converted into allolactose and induce the lac operon. This can be side-stepped by adding the gratuitous inducer IPTG to Ebg-mutant and *ebgR*-null cells, whereupon *lacY* is induced and lactose enters the cells, enabling the use of lactose as a carbon source through the agency of the mutant Ebg enzyme.

The need to add the artificial IPTG inducer shows that this system alone is not a fully independent evolved operon, but other mutations were characterized which allowed this limitation to be overcome. Specific single codon changes in the *ebgR* gene resulted in repressors which could be readily induced by lactose⁶, and a minimum of two mutations produced an Ebg enzyme capable of synthesizing allolactose from lactose substrate (a 'Class IV' enzyme, which is also improved in its ability to hydrolyze lactose to galactose and glucose^{6,8}). These alterations then enabled a fully regulatable system, where lactose induced the *ebgR* repressor and thus permitted the expression of the *ebgA* and *ebgC* genes whose products constitute the (mutant) Ebg enzyme. This enzyme in turn produces allolactose which induces the lac operon and allows the express of the permease enzyme (*lacY* product) and thereby lactose entry into the cell (Fig. 4A5.1). This coupling between Ebg and the *lacZ*-deleted lac operon in effect is the birth of a new operon through laboratory evolution, which enables fully

controlled lactose utilization to resume, overcoming the original loss of *lacZ* expression (Fig. 4A5.1).

Simple enough for you? In fact, this sketch of an *E. coli* regulatory system as indicated thus far is in reality considerably simpler than the full picture in reality. One such additional process to note is a positive control mechanism centering on glucose, mediated through a molecular 'second messenger', cyclic adenosine monophosphate (cAMP), and in turn a protein which binds cAMP (cAMP Activator Protein; CAP). The activated CAP protein (which results from low levels of glucose) binds to a site near the *lac* promoter and activates transcription of the *lac* genes. In the presence of glucose, cAMP levels drop, CAP is de-activated, and *lac* expression falls. A putative CAP-binding site is also present near the *Ebg* promoter, suggesting it is under the same control mechanism⁶. If glucose is available in sufficient quantities for *E. coli* growth, it is unnecessary for the cell to invest energy in producing catabolic enzymes such as those of the *lac* operon, and to avoid this, the CAP / cAMP signaling system has evolved¹⁵. While this and other complications need not detract us from the overall regulatory picture (Fig. 4A5.1), it is well to keep in mind that gene regulation in general usually has multiple levels of operation, many of which have only recently been determined.

Although a key component of a functional system, the efficiency of the best evolved *Ebg* enzyme as a β -galactosidase falls far short of the natural *lacZ* gene product⁸. This shortfall highlights an important limitation of directed laboratory evolution where selection is based on organismal growth rather than a direct property of the enzyme. A point is reached where it becomes difficult to distinguish a superior growth rate in itself, even if the evolved β -galactosidase activity required for lactose utilization is still markedly inferior to the original

enzyme[▼]. It was estimated that this difficulty would be encountered when Ebg enzyme activity was only 7% that of the wild-type *lacZ*-encoded β -galactosidase⁸. So this foray into bacterial regulatory systems demonstrates that despite the great advantages of microbial systems over multicellular organisms for directed evolution studies, selection for a whole-organism phenotype still presents limitations. In turn, it highlights the need for a selection or screening process which more directly targets the efficiency of the enzyme as the experimental read-out. (This point also is at the heart of the ‘First Law’ of directed evolution noted in Chapter 4 of *Searching for Molecular Solutions*; ‘You get what you screen / select for’.)

Although the Ebg case study focuses on a single enzymatic activity (β -galactosidase catalysis), it also highlights the need to design an entire pathway (the joint *lac*/Ebg operon) for its ultimate utility. For the practical needs of directed evolution in general, it is also increasingly the case that whole biosynthetic pathways need attention rather than focusing on a single enzyme target alone. Rational strategies for the most efficient implementation of evolutionary approaches to pathway design are under continued development¹⁶. Moreover, multiplexing strategies for pathway evolution promise to reveal novel solutions to improved biosynthetic routes in a relatively short period of time¹⁷.

[▼] It might then be wondered what selective pressures drove the high level of catalytic efficiency of natural *lacZ* β -galactosidase in the first place. Here the conditions of laboratory selection with pure cultures can be distinguished from the complex conditions of natural growth including competition from other organisms. Even a small selective advantage will be significant over time towards driving enzyme efficiencies.

References:

1. Francis, J. C. & Hansche, P. E. Directed evolution of metabolic pathways in microbial populations. I. Modification of the acid phosphatase pH optimum in *S. cerevisiae*. *Genetics* **70**, 59-73 (1972).
2. Ferenci, T. & Lee, K. S. Isolation, by affinity chromatography, of mutant *Escherichia coli* cells with novel regulation of *lamB* expression. *J Bacteriol* **154**, 984-7 (1983).
3. Vournakis, J. N. & Elander, R. P. Genetic manipulation of antibiotic-producing microorganisms. *Science* **219**, 703-9 (1983).
4. Lal, R. et al. Engineering antibiotic producers to overcome the limitations of classical strain improvement programs. *Crit Rev Microbiol* **22**, 201-55 (1996).
5. Campbell, J. H., Lengyel, J. A. & Langridge, J. Evolution of a second gene for beta-galactosidase in *Escherichia coli*. *Proc Natl Acad Sci U S A* **70**, 1841-5 (1973).
6. Hall, B. G., Betts, P. W. & Wootton, J. C. DNA sequence analysis of artificially evolved *ebg* enzyme and *ebg* repressor genes. *Genetics* **123**, 635-48 (1989).
7. Matthews, B. W. The structure of *E. coli* beta-galactosidase. *C R Biol* **328**, 549-56 (2005).
8. Hall, B. G. The EBG system of *E. coli*: origin and evolution of a novel beta-galactosidase for the metabolism of lactose. *Genetica* **118**, 143-56 (2003).
9. Oehler, S., Eismann, E. R., Kramer, H. & Muller-Hill, B. The three operators of the *lac* operon cooperate in repression. *Embo J* **9**, 973-9 (1990).
10. Huber, R. E., Kurz, G. & Wallenfels, K. A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (*E. coli*) on lactose. *Biochemistry* **15**, 1994-2001 (1976).
11. Sadler, J. R., Sasmor, H. & Betz, J. L. A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. *Proc Natl Acad Sci U S A* **80**, 6785-9 (1983).
12. Simons, A., Tils, D., von Wilcken-Bergmann, B. & Muller-Hill, B. Possible ideal *lac* operator: *Escherichia coli* *lac* operator-like sequences from eukaryotic genomes lack the central G X C pair. *Proc Natl Acad Sci U S A* **81**, 1624-8 (1984).
13. Kolkhof, P. Specificities of three tight-binding *Lac* repressors. *Nucleic Acids Res* **20**, 5035-9 (1992).
14. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588-92 (2005).
15. Lewin, B. *Genes IX* (Jones and Bartlett, 2007).
16. Haseltine, E. L. & Arnold, F. H. Synthetic gene circuits: design with directed evolution. *Annu Rev Biophys Biomol Struct* **36**, 1-19 (2007).

17. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-8 (2009).