Project no. 043338

Project acronym: EMERGENCE

Project title: A foundation for Synthetic Biology in Europe

Instrument: NEST Pathfinder

Thematic Priority: Synthetic Biology

Deliverable 4.1:
Database on quantitative prokaryotic promoter performance

Due date of deliverable: Jan. 2007
Actual submission date: March 2008

Start date of project: 1.12.2006 Duration: 36 months

Victor de Lorenzo
Centro Nacional de Biotecnologia, CSIC, Campus de Cantoblanco, 28049 Madrid, Spain

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Database on quantitative prokaryotic promoter performance[1]

Introduction. One of the trademarks of Synthetic Biology is the rational combination of regulatory modules in artificial circuits for performing non-natural tasks, including complex binary computation operations based on logic gates [1,2]. The basis of such an endeavour is the implicit adoption of the metaphor of the cell as a sort of Turing machine. In this way, physicochemical environmental signals (the *inputs*) activate an existing gene expression *program* (encoded in the DNA), which is ultimately executed by transcriptional regulators on promoters and then by the downstream protein expression machinery [3]. This results e.g. in changes of the cell metabolism through the increase or decrease of the production rate of specific proteins (the *output*). Under this conceptual frame, the program behind any biological function could in principle be de-constructed into minimal operative units, called by many *biological parts* (see http://parts.mit.edu [1]). Such units can then ideally be re-assembled following a rational blueprint to perform a different program, resulting in altogether new properties and behaviours. In this respect, *Synthetic Biology* clearly takes off from what since the late 70s was called *Genetic Engineering*, as it brings into Biology robust engineering principles such as abstraction, hierarchical design, modularization and definition of systems boundaries -rather than vague analogies to cutting and pasting DNA sequences. In this mini-review, we briefly assess what is actually available for designing genetic circuits, how to upgrade natural modules to meet the requirements of robust engineering, and where to find the pieces that are still missing. Furthermore, we raise the questions of connectivity and evolvability of biological modules as two of the major bottlenecks that hinder the development of synthetic biological circuitry.

De-constructing naturally-occurring genetic circuits into usable regulatory elements. The principal actors of the biological input/output functions are the *cis*-(promoters) and the *trans*-regulatory elements (transcriptional regulators). Prokaryotic transcriptional factors (TFs) drive the activity of their cognate promoter(s) in response

[1] Note that much of this deliverable has been published as review in FEBS Letters: **Silva-Rocha, R. and de Lorenzo** (2008) Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS Lett* **582:** 1237-44.

to one or more environmental stimuli. TFs can generally be *activators* by enhancing the binding or the activity of the RNA polymerase (RNAP) in the cognate promoters, or *repressors* by blocking this binding, or both [4]. Most known prokaryotic activators bind the upstream region of a promoter in response to a signal (for example, a substrate of the metabolic pathway regulated by the TF) and enhance the recruitment of the RNAP to the site. Alternatively, they may promote the escape and further progression of the transcription machinery from the promoter into the transcribed DNA sequence [5]. In contrast, transcriptional repressors typically interfere with the binding of RNAP to the -35 and -10 DNA hexamers of bacterial promoters. In this case, environmental stimuli decrease the affinity of the TF for its binding site, thereby allowing the RNAP to access the promoter and proceed with transcription [6,7]. One question relevant to circuit design emerges now: why activators and repressors instead of just one mechanism or the other? Sometimes the very same biological function (for instance, the *ara* systems for arabinose consumption) is positively regulated in one bacterium (*E. coli*, activated by AraC [8]) and negatively controlled in another (*B. subtillis*, repressed by AraR [9]). There is not an easy answer to this. It seems that activators generally produce more transcriptional output than repressors [10]. It is also likely that positive regulation allows a higher connectivity of the corresponding promoter to physiological co-regulation [11].

Prokaryotic promoters as Boolean logic gates. The participation of one or more TFs in the regulation of a given promoter confers the system the ability of integrating different input signals in a fashion not unlike those described by the gates of Boolean logic. Such gates perform operations on one or more inputs and produce each time a single logic output. Since the output is also a logic-level value, an output of one logic gate can connect to the input of one or more other logic gates. The logic thereby performed is thus adequate for the functioning of digital circuits. Logic gates are typically implemented electronically using diodes or transistors but, as discussed below, can they also be constructed using *inter alia* promoters and regulators. An archetypical example in this context is the *lac* operon of *E. coli*, where expression of the genes for lactose metabolism is controlled by the *lacI* repressor and by the cyclic AMP receptor protein (CRP) activator. The LacI repressor binds to the *lac* promoter (*Plac*) as a tetramer and inhibits gene expression both through the physical occupation of the RNAP binding site and through the formation of a DNA loop [12]. The binding of the inducer (lactose or

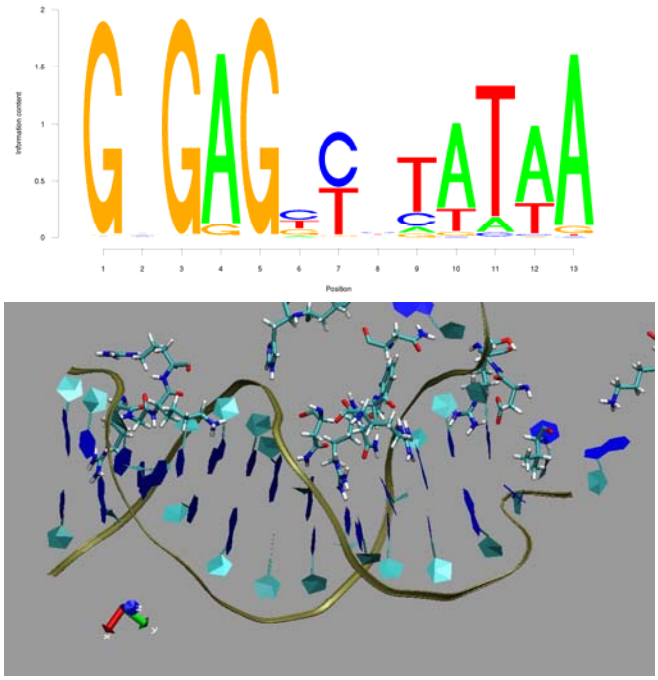
IPTG) to LacI triggers a conformational switch in the tetramer that decreases the affinity to the operator sequences and thus allows transcription initiation from the *Plac* [12,13]. The behaviour of the *lac* regulatory system has been described to be an intermediate between AND-gate and OR-gate logic function (see below; [14]).

Although binary logic circuits are based on functions with just two possible states (0 or 1), existing biological systems typically display continuous values for the input/output functions [15]. In addition, such values are submitted to noise and cell-to-cell stochastic variations due to the nature of the molecular interactions involved [16]. This has important consequences for the construction of artificial genetic circuits based in the naturally occurring transcriptional modules and its applicability in synthetic networks [17]. For example, an artificial system with oscillatory properties constructed by the combination of the repressor properties of three well characterized TFs (LacI, TetR and the repressor), lost its periodicity after a few rounds of oscillation [18]. Although promoters destined for building artificial circuits should ideally behave as bi-stable switches resembling a digital response, this is not the case in most available instances. Whether or not naturally occurring promoters can be artificially re-designed to achieve permanently such a binary performance remains an open question, as Darwinian selection may eventually press against such a conduct.

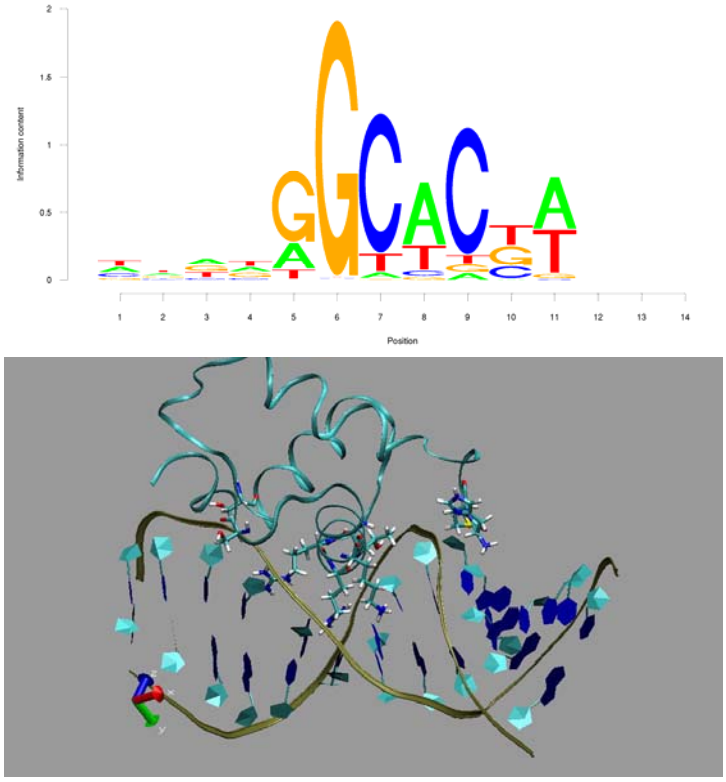
Simple logic gates shape the bulk of transcriptional regulation circuits. Despite the constrains mentioned above, representing the reactions and interactions involved in gene expression control using circuit diagrams and Boolean logic operators is still an extremely useful abstraction. As the biological reactions adopt somewhat continuous values, the 0/1 states are generally agreed to reflect low/high states for the *input* status and off/on for *output* promoter activity. Silva-Rocha and de Lorenzo [19] have reported most posible combinations of prokaryotic promoters as logic gates.

Towards a database on quantitative (orthogonal) promoter performance. We have been improving the FoldX software to be able to deal with DNA and to mutate it. A test set of 99 protein mutants binding DNA were created (including 46 conservative mutations) taken from the ProNIT database. For each of these mutants, the changes in affinity for the nucleic acid were experimentally determined and the correlation with

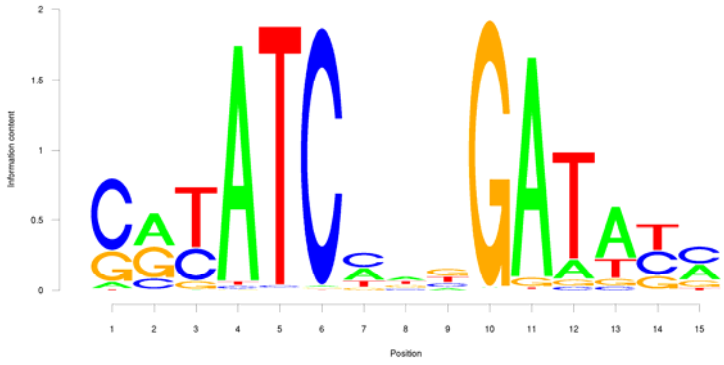
FoldX predictions is 0.56 for all available mutantions and 0.71 considering just the conservative ones. In order to generate an orthologous system, we selected a couple of proteins to study in detail with the aim of engineer them to have graduated transcription regulation. We decided to study T7 polymerase, TetR and Sigma 54 proteins. In all cases we can predict reasonably well its binding profiles which encourage us to the following steps. In the case of Sigma 54 we are trying to generate mutants that change its binding specificity to a new binding site that could be unique in bacteria. On the other hand, for T7 pol and TetR we plan to design variants with different affinities and overlaping binding sites that combined could provide a wide lanscape for transcription regulation. The following figures show the structure-based prediction of the binding profiles for T7 Polymerase (PDBID: 1CEZ), Sigma 54 (bound to -24 region; PDBID: 2O8K) and TetR (PDBID: 1QPI) proteins, expresed as SequenceLogos. Each logo is presented together with a picture of the interface region of each complex. Only residues contacting DNA are depicted.

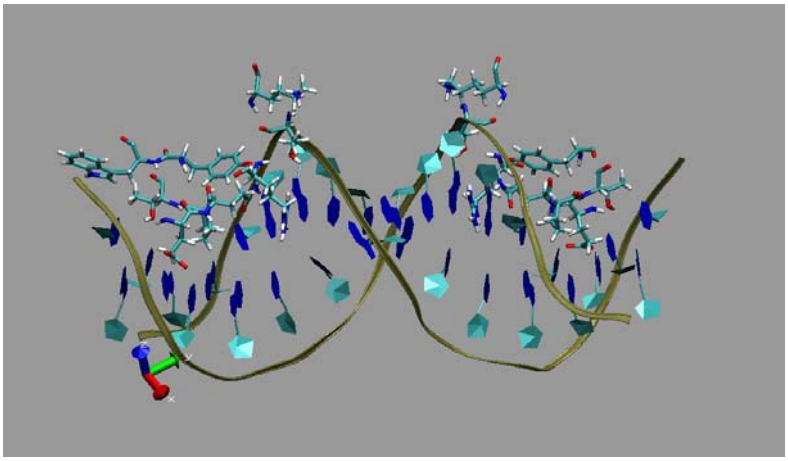


Crystal structure of a T7 RNA Polymerase-T7 promoter bound complex



NMR Structure of the Sigma-54 RpoN Domain Bound to the-24 Promoter Element





Crystal structure of a tetracyclin repressor/ operator complex

REFERENCES

[1] Endy, D. (2005) Foundations for engineering biology. Nature 438, 449-453.

[2] Hermsen, R., Tans, S. and ten Wolde, P.R. (2006) Transcriptional regulation by competing transcription factor modules. PLoS Comput Biol 2, e164.

[3] Danchin, A. (2003) The Delphic boat: what genomes tell us. Harvard University Press, Boston.

[4] Kaern, M., Blake, W.J. and Collins, J.J. (2003) The engineering of gene regulatory networks. Annu Rev Biomed Eng 5, 179-206.

[5] Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J. and Phillips, R. (2005) Transcriptional regulation by the numbers: models. Curr Opin Genet Dev 15, 116-124.

[6] Tropel, D. and Van Der Meer, J.R. (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. Microbiol Mol Biol Rev 68, 474-500.

[7] Zhou, D. and Yang, R. (2006) Global analysis of gene transcription regulation in prokaryotes. Cell Mol Life Sci 63, 2260-2290.

[8] Schleif, R. (2000) Regulation of the L-arabinose operon of *Escherichia coli*. Trends Genet 16, 559-565.

[9] Franco, I.S., Mota, L.J., Soares, C.M. and de Sa-Nogueira, I. (2007) Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. Nucl Acids Res 35, 4755-4766.

[10] Alon, U. (2006) An introduction to Systems Biology: design principles of biological circuits. Chapman and Hall/CRC, New York.

[11] Velazquez, F., de Lorenzo, V. and Valls, M. (2006) The *m*-xylene biodegradation capacity of Pseudomonas putida mt-2 is submitted to adaptation to abiotic stresses: evidence from expression profiling of *xyl* genes. Environ Microbiol 8, 591-602.

[12] Lewis, M. (2005) The *lac* repressor. C R Biol 328, 521-548.

[13] Wilson, C.J., Zhan, H., Swint-Kruse, L. and Matthews, K.S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. Cell Mol Life Sci 64, 3-16.

[14] Setty, Y., Mayo, A.E., Surette, M.G. and Alon, U. (2003) Detailed map of a *cis*-regulatory input function. Proc Natl Acad Sci U S A 100, 7702-7707.

[15] Arkin, A. and Ross, J. (1994) Computational functions in biochemical reaction networks. Biophys J 67, 560-578.

[16] Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. Science 297, 1183-1186.

[17] Voigt, C.A. (2006) Genetic parts to program bacteria. Curr Opin Biotechnol 17, 548-557.

[18] Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403, 335-338.

[19] Silva-Rocha, R. and de Lorenzo (2008) Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS Lett* **582:** 1237-44.