BioSystems 112 (2013) 139–144

Contents lists available at SciVerse ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Modeling bacterial immune systems: Strategies for expression of toxic – but useful – molecules

Marko Djordjevic

Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia

A B S T R A C T

Protection of bacterial cells against virus infection requires expression of molecules that are able to destroy the incoming foreign DNA. However, these molecules can also be toxic for the host cell. In both restriction–modification (R–M), and the recently discovered CRISPR/Cas systems, the toxicity is (in part) avoided through rapid transition of the expression of the toxic molecules from "OFF" to "ON" state. In restriction–modification systems the rapid transition is achieved through a large binding cooperativity, and low translation rate of the control protein. On the other hand, CRISPR array expression in CRISPR/Cas systems involves a mechanism where a small decrease of unprocessed RNAs leads to a rapid increase of processed small RNAs. Surprisingly, this rapid amplification crucially depends on fast non-specific degradation of the unprocessed molecules by an unidentified nuclease, rather than on large cooperativity in protein binding. Furthermore, the major control elements that are responsible for fast transition of R–M and CRISPR/Cas systems from "OFF" to "ON" state, are also directly involved in increased stability of the steady states of these systems. We here discuss mechanisms that allow rapid transition of toxic molecules from the unproductive to the productive state in R–M and CRISPR/Cas systems. The main purpose of this discussion is to put relevant theoretical and experimental work in a perspective that points to general similarities in otherwise mechanistically very different bacterial immune systems.

1. Introduction

Bacterial immune systems defend host cell against infection by bacteriophages (bacterial viruses). Two most prominent examples of such systems are restriction–modification systems, and the recently discovered CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR associated sequences) systems. In order to defend the host bacteria against the incoming infection, these systems have to express molecules that can destroy the genome of the incoming virus. While these molecules are evidently useful, they can also be toxic, due to autoimmunity problems. That is, the same mechanism that is responsible for destruction of the foreign DNA, can also, in principle, lead to the destruction of the host genome.

An example of the balance between toxicity and usefulness is provided by the restriction enzyme within a type II restriction modification system (R–M system) (Kobayashi, 2001). Since the restriction enzyme makes cuts in specific DNA sequences, it can, in principle, cut both the DNA of the incoming virus and the host DNA. Destruction of the host DNA is prevented by methylase, which protects the same DNA sequences that are cut by the restriction enzyme. Consequently, unmethylated DNA sequences of the incoming virus will be cut by the restriction enzyme, while the host genome is protected by its methylation.

A quite different type of bacterial immune system is provided by a recently discovered CRISPR/Cas system (Barrangou et al., 2007; Makarova et al., 2006). The system consists of CRISPR array and associated *cas* genes (Al-Attar et al., 2011), and is represented by Fig. 1. CRISPR cassettes consist of identical direct repeats of about 30 bp in length, interspaced with variable spacers of similar length. CRISPR presents an adaptive prokaryotic immune system, which is responsible for defending prokaryotic cell against invaders, so that a match between a CRISPR spacer and an invading phage (bacterial virus) sequence provides immunity to infection. In addition to the match between a CRISPR spacer and the invading phage, CRISPR-associated (*cas*) genes are also required for this immunity. Specifically, one of the Cas proteins (CasE in *Escherichia coli*) is responsible for processing of the long transcripts that correspond to the entire CRISPR locus (called pre-crRNAs), to small interfering RNAs (called crRNAs) (Brouns et al., 2008; Pougach et al., 2010; Pul et al., 2010). Furthermore, a large complex of Cas proteins is, together with crRNAs, responsible for the recognition and inactivation of invading viruses (Al-Attar et al., 2011).

While CRISPR/Cas system has to efficiently recognize foreign DNA, it also has to prevent autoimmunity (Al-Attar et al., 2011). Regarding this, note that crRNAs are complementary to the spacers on chromosomal CRISPR array from which they are transcribed. Furthermore, it is frequently observed that CRISPR spacers are homologous to host chromosomal sequences (Cui et al., 2008). This implies a possibility of spurious recognition, and subsequent destruction, of the host DNA (Al-Attar et al., 2011) by CRISPR/Cas, though the exact process ("antidote") which prevents autoimmunity is still unclear.

E-mail address: dmarko@bio.bg.ac.rs

0303-2647/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biosystems.2013.03.004

Fig. 1. A scheme of CRISPR/Cas genomic arrangement. The genomic arrangement of different *cas* genes and CRISPR array elements is indicated. R and S within the CRISPR array correspond, respectively to the repeats and spacers; note that the spacer sequences differ from each other, and are labeled by the consecutive numbers (1,2,3,. . .). IGLB and L in figure correspond to the intergenic regions where promoters for, respectively, *cas* genes and CRISPR array are located. Different *cas* genes are labeled by cas1-3 and casABCDE.

From the above discussion, it is evident that bacterial immune systems can employ a quite different mechanisms for expression of toxic molecules. Despite these differences, it may also be useful to think in terms of more general principles that govern expression of toxic molecules inside bacterial cell. For example, expression of a toxic molecule should generally be accompanied by expression of an antidote (e.g. methylation in the case of R–M systems). Furthermore, it seems plausible that generation of a toxic molecule should involve a rapid transition from "OFF" to "ON" state, so that toxic molecules are present in small amounts when they are not needed, but are then rapidly generated upon infection by invasive DNA. Finally, additional, more subtle, principles may also be relevant: e.g. fluctuations of the toxic molecule in its steady state might need to be small, in order to evade that a large fluctuation of the toxic molecule is unmatched by the antidote amount. We will below discuss relevant theoretical and experimental results on bacterial immune systems, with the purpose of pointing to some possible strategies for expression of toxic molecules inside cell.

2. Regulation of R–M systems

We will first discuss regulation of R–M systems. To understand regulation of these systems, it is important to note that they are often mobile and can spread from one bacterial host to the other (Jeltsch and Pingoud, 1996; Kobayashi, 2001). When a R–M system enters a naive bacterial host, the host genome is initially unmethylated, and can consequently be cut by the restriction enzyme. It is, therefore, evident that expression of the restriction enzyme and methylase must be tightly regulated in order to ensure that the bacterial genome is protected by the methylase, before it is cut by the restriction enzyme. This tight regulation is often achieved through a dedicated control protein (C protein) (Tao et al., 1991), and the mechanism which ensures such regulation will be further discussed below.

2.1. Regulation by control protein

A typical gene arrangement in a R–M system is such that the restriction enzyme (R) and the control protein (C) are transcribed together. For definiteness, we will below concentrate on AhdI type II restriction–modification system (which we will further, for simplicity, denote as R–M system). Transcription of both of these proteins is regulated by the control protein C, which binds to the upstream operator sequence (i.e. C protein regulates both its own expression and expression of R gene) (Bart et al., 1999). The main property of transcriptional control by C protein is a large binding cooperativity (McGeehan et al., 2006; Streeter et al., 2004), which is further discussed below.

Basal rate of transcription of C and R genes is very low, due to a low binding affinity of RNA polymerase (RNAP) to the core promoter in the operator sequence (see Fig. 2A) (Bogdanova et al., 2008). In order to activate transcription of these genes, it is necessary to have C proteins. In solution, C protein exists as a monomer, but in order to bind to DNA, it must first form a dimer in solution. In the operator sequence, there are two dimer binding sites. The first dimer binding site is located immediately upstream of the core promoter; binding of C protein to this position leads to transcription activation (see Fig. 2B). The second dimer binding site directly overlaps with the core promoter, so that binding of C protein to this position leads to transcription repression (see Fig. 2C).

Due to a very large binding affinity, as soon as one dimer is bound to DNA, it immediately leads to binding of the second dimer. Due to this, in the absence of RNAP only tetramer can be observed to be bound to DNA (Bogdanova et al., 2008; McGeehan et al., 2006). However, when RNAP is added to the solution, it can displace one of the two C protein dimers in order to form the activation complex (see Fig. 2B). On the other hand, when concentration of C protein is increased, it becomes increasingly entropically favorable to have the other dimer bound to DNA; consequently, at higher C protein concentrations, RNAP is displaced from the core promoter, which leads to formation of the repressor (tetramer complex) (Bogdanova et al., 2008).

In Fig. 2, configurations that correspond to the different arrangements of RNAP and C protein are schematically shown. With each configuration, the appropriate interaction energies are indicated (for the definition of the interaction energies, see the legend of Fig. 2). The weights that correspond to the basal complex (RNAP bound to the promoter) (Fig. 2A), the activator complex (Fig. 2B) and the repressor complex (Fig. 2C) are denoted, respectively, as Z_{RNAP}, Z_{D-RNAP} and Z_T, and given by the following expressions:

$$Z_{RNAP} = K\,[RNAP]\,\exp(-\Delta G_{RNAP}) \tag{1.1}$$

$$Z_{D-RNAP} = K^3\,[C]^2\,[RNAP]\exp(-\Delta G_D - \Delta G_L - \Delta G_{D-RNAP} - \Delta G_{RNAP}) \tag{1.2}$$

$$Z_T = K^4\,[C]^4\exp(-\Delta G_L - \Delta G_R - \Delta G_T - 2\Delta G_D) \tag{1.3}$$

where K is a multiplicative constant (with units of the inverse protein concentration), $[C]$ is concentration of C protein monomers,

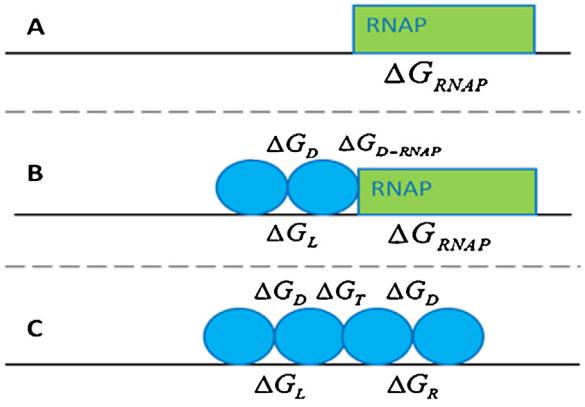

Fig. 2. A scheme of the possible configurations of C protein and RNAP in the promoter region. (A) The basal transcription configuration. (B) The activator configuration. (C) The repressor (tetramer) configuration. ΔG_{RNAP} is the binding energy of RNAP to the promoter; ΔG_D is the dimerization free energy; ΔG_{D-RNAP} is the interaction energy of the dimer with RNAP; ΔG_L and ΔG_R are the interaction energy of C protein dimer with, respectively, the upstream ("left") and the downstream ("right") binding site; ΔG_T is the tetramerization free energy.