doi:10.1016/j.jmb.2007.04.041 J. Mol. Biol. (2007) 369, 1200–1213

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Modelling Transcriptional Interference and DNA Looping in Gene Regulation

Ian B. Dodd[1,2]*, Keith E. Shearwin[2] and Kim Sneppen[1]

[1]*Centre for Models of Life Niels Bohr Institute Blegdamsvej 17, 2100 Copenhagen Ø, Denmark*

[2]*Molecular and Biomedical Sciences (Biochemistry) University of Adelaide SA 5005, Australia*

**Corresponding author*

We describe a hybrid statistical mechanical and dynamical approach for modelling the formation of closed, open and elongating complexes of RNA polymerase, the interactions of these polymerases to produce transcriptional interference, and the regulation of these processes by a DNA-binding and DNA-looping regulatory protein. As a model system, we have used bacteriophage 186, for which genetic, biochemical and structural studies have suggested that the CI repressor binds as a 14-mer to form alternative DNA-looped complexes, and activates lysogenic transcription indirectly by relieving transcriptional interference caused by the convergent lytic promoter. The modelling showed that the original mechanisms proposed to explain this relief of transcriptional interference are not consistent with the available *in vivo* reporter data. However, a good fit to the reporter data was given by a revised model that incorporates a novel predicted regulatory mechanism: that RNA polymerase bound at the lysogenic promoter protects itself from transcriptional interference by recruiting CI to the lytic promoter. This mechanism and various estimates of *in vivo* biochemical parameters for the 186 CI system should be testable. Our results demonstrate the power of mathematical modelling for the extraction of detailed biochemical information from *in vivo* data.

Keywords: repressor; recruitment; RNA polymerase; transcriptional interference; DNA looping

Introduction

Biological systems confront researchers with an astounding depth and breadth of detail, an overwhelming amount of data that could be collected by observation and experiment. Thus, to direct biological research efficiently, it is important to know what kind and how much information about a system is necessary in order to obtain the desired level of understanding of that system.

Bacteriophage λ has been subjected to decades of intensive study, making it one of the best understood gene regulatory systems.[1] *In vitro* experiments have measured a large fraction of the interaction strengths and reaction rates that parameterize the regulation of the lysogenic and lytic promoters by the CI and Cro proteins, such that detailed, quantitative physico-chemical models of gene regulation in this system can be constructed.[2–5] However, a λ-style approach has limitations for investigating other gene regulation systems that we wish to understand in depth. First, it is usually not practical to measure the large numbers of biochemical parameters that comprise such systems. Second, it is difficult to know whether reaction parameters measured by *in vitro* experiments reflect those *in vivo*.

Bacteriophage 186 provides a counterpoint to λ. The two phages have virtually identical lifecycles; both infect *Escherichia coli*, are capable of utilizing lytic and lysogenic reproductive pathways, and form integrated, SOS-inducible prophages. Like λ, 186 has a lysogenic repressor protein, also called CI (but not sequence-related to λ CI), that regulates the 186 lytic and lysogenic promoters. Reporter studies show that the effect of 186 CI on the 186 lytic and

Abbreviations used: CTD, C-terminal domain; NTD, N-terminal domain; TI, transcriptional interference; TEC, transcription elongation complex; RNAP, RNA polymerase; WLU, wild-type lysogenic unit.

E-mail address of the corresponding author: ian.dodd@adelaide.edu.au

lysogenic promoters,*pR* and *pL*, respectively, is very similar to the effect of λ CI on the analogous λ promoters, P_R and P_{RM}. 186 *pR* is repressed strongly by CI, while 186 *pL* activity is increased at low to intermediate concentrations of CI but reduced again at high concentrations of CI.[6] However, there is much less information available about the strengths of the interactions between the components in the 186 CI system than there is for λ. Here, we develop a mathematical model that allows us to extract estimates for *in vivo* biochemical parameters and to make mechanistic predictions for the 186 CI system from analysis of *in vivo* reporter data.

The 186 promoters and CI binding sites are arranged quite differently from those in λ (Figure 1(a)). The *pR* and *pL* promoters lie face-to-face, separated by 62 bp. Also, there are additional "flanking" distal CI binding sites, *FL* and *FR*, located ~300 bp on either side of the promoters that are important in their regulation by CI. *In vitro*, 186 CI binds to its three major operator sites at *pR* (Figure 1(a)) in an all-or-none manner.[7] CI bound at *pR* also occupies neighbouring DNA, including part of the *pL* promoter, though less strongly. Recent crystal structures of the 186 CI repressor have provided insights into how CI might bind DNA.[8] The

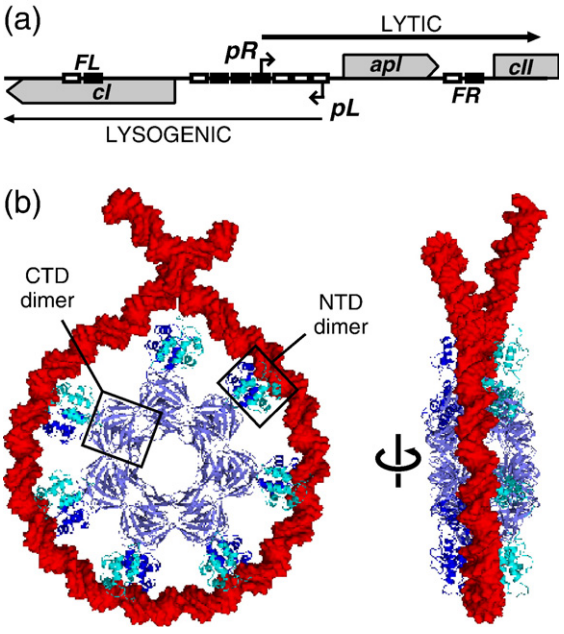

Figure 1. CI DNA binding. (a) A map (not to scale) showing the CI regulatory region in phage 186. Boxes denote CI binding sites based on footprinting data and sequence analysis.[6,37] Filled boxes represent strong, well-defined binding sites; open boxes represent more weakly protected regions. The sites are centred two DNA turns apart. (b) A hypothetical structural model of a CI-*pR* complex based on crystal structures of the CI C-terminal domain (CTD) 14-mer and an intact CI dimer.[8] The CI N-terminal domains (NTDs) have been positioned with the helix-turn-helix domains of CI in the major groove of the DNA using PyMol [http://pymol.sourceforge.net/]. The model of curved DNA was provided by Mitchell Lewis (University of Pennsylvania).

C-terminal domain (CTD) of 186 CI crystallized as a 14-mer, a 7-mer of dimers, arranged in a ring (Figure 1(b)). The *in vivo* relevance of the 14-mer, or a similar oligomer, is supported by the finding that the monomer–monomer and dimer–dimer interfaces seen in the crystal are consistent with the locations of mutations that disrupt CI-CI interactions *in vivo*.[8] The crystal structure of a full-length dimer showed that the DNA-binding N-terminal domains[9] (NTDs) would be located on the outside of the 14-mer. Figure 1(b) shows a hypothetical structural model with DNA wrapped around a CI "wheel", in which the helix-turn-helix units of the NTDs are positioned to contact adjacent major grooves of operators located two DNA turns apart, as found at *pR*. Sedimentation equilibrium measurements indicate that, like λ CI, high order 186 CI multimers are unlikely to form in solution at physiological concentrations.[10] Rather, we envisage that formation of a CI 14-mer *in vivo* would be facilitated by DNA binding.

The structure itself and experiments in which λ P_R and P_{RM} were regulated by a chimeric repressor carrying the λ CI NTD fused to the 186 CI CTD, led to the proposal that steric clashes between DNA segments would prevent non-adjacent operators binding to adjacent NTDs on the CI wheel.[8] For example, in the model in Figure 1(b), if one of the two DNA segments shown "exiting" the wheel were to bind to the seventh, unoccupied CI dimer, then the two exiting DNA segments would not be able to avoid each other unless there was a large distortion in the DNA or the protein. Thus, the model in Figure 1(b) depicts how CI might occupy up to six consecutive binding sites at *pR*-*pL*.

The response of *pR* and *pL* to controlled expression of CI has been measured using chromosomally integrated *lacZ* reporters.[6] Various combinations of mutations that either eliminated CI binding to *FL* (FL^-) or *FR* (FR^-), or inactivated *pR* without affecting CI binding (pR^-), were used and concentrations of CI were measured by quantitative Western blotting. These data are plotted in Figure 2(a). The reporter, DNA binding and structural information has been incorporated into an unusual mechanistic model of CI regulation of *pR* and *pL*,[6,8] which is depicted in Figure 2(b) and described below.

In the absence of CI (Figure 2(b), species 1), the strong *pR* promoter inhibits *pL* by transcriptional interference (TI).[11,12] Inactivation of *pR* by mutation increases *pL* activity ~6-fold *in vivo* (Figure 2(a)).[6,11] Experimental and modelling analyses of the TI support a "sitting duck" mechanism, where transcription elongation complexes (TECs) from *pR* cross *pL* and inactivate pre-clearance RNA polymerase (RNAP) complexes (sitting ducks) bound at *pL*.[11,13] Smaller contributions to TI are expected from collisions of *pR* and *pL* TECs (since the inter-promoter distance is only 62 bp) and occlusion,[14] in which passing *pR* TECs prevent RNAP binding to *pL*. There is no significant inhibition of *pR* by *pL*.[11]