FEBS Letters 563 (2004) 66–68 FEBS 28216

In vivo non-specific binding of λ CI and Cro repressors is significant

Audun Bakk*, Ralf Metzler

NORDITA, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

Received 16 January 2004; revised 12 February 2004; accepted 18 February 2004

First published online 12 March 2004

Edited by Thomas L. James

Abstract We propose a thermodynamic model that includes the non-specific binding of the λ phage regulatory proteins CI and Cro. By fitting the model to experimental in vivo data on activities of the two promoters P_{RM} and P_R versus concentration, we estimate the free energy upon non-specific binding to be -4.1 ± 0.9 kcal/mol for CI and -4.2 ± 0.8 kcal/mol for Cro. For concentrations >100 nM of CI or Cro, we find that $>50\%$ of these proteins are non-specifically bound. In particular, in a lysogen (~ 250 CI monomeric equivalents per cell) nearly 90% of CI is non-specifically bound.

Key words: Genetic regulation; Lambda phage; Non-specific binding; Thermodynamics

1. Introduction

The understanding of biological systems as integrated networks is a research area of great interest [1]. An example of a well-studied regulatory network is the system comprised of a λ phage-infected *Escherichia coli* bacterium [2]. The fate of the *E. coli* cell bifurcates into two pathways: lysogeny, during which the viral genome becomes integrated in the host genome and is subsequently silently replicated in the regular *E. coli* cell cycle for generations, or lysis, where the virus becomes massively replicated during a single *E. coli* cell cycle, whereupon the cell bursts and releases about 100 progeny phages. The decision-making for either pathway is tightly connected to the right operator (O_R), a sequence of 64 bp of the λ DNA (see Fig. 1). In particular, the stability of a lysogen is expected to crucially depend upon the binding of CI and Cro to the DNA binding sites O_R1, O_R2, and O_R3 [3]. In a lysogen, O_R1 and O_R2 are usually occupied by one CI dimer each, exhibiting a cooperative interaction, and P_{RM} is occupied by RNA polymerase (RNAP) such that CI is continuously expressed, thus maintaining repression of *cro*. Two genes are regulated to either side of O_R: *c*I and *cro* that encode the auto-regulating proteins CI and Cro, respectively. *c*I is transcribed from promoter P_{RM} and *cro* is transcribed from promoter P_R. The transcription activities of P_{RM} and P_R, and ultimately the CI and Cro concentrations, will have a strong impact on the destiny of the λ-infected *E. coli* cell.

There exist comprehensive thermodynamic protein–DNA binding data of O_R from in vitro experiments [4–7]. However, there is a lack of sufficient quantitative information about non-specific binding (NSB) of the repressors, i.e. protein–DNA binding outside O_R, and in particular numbers for the free energies for NSB in vivo are insufficient. NSB is important to quantify because it reduces the available (free) cellular concentration of a species for a given total concentration that experiments usually provide. The traditional view (also adopted in this work) is that it is only the free concentration of the regulatory proteins that contributes to the specific binding [8].

In the following, we quantify the free energies upon non-specific binding for the gene regulatory proteins CI and Cro of λ phage by comparing an expanded version of a well-established thermodynamic model to in vivo data on the P_{RM} and P_R activities. We estimate the free energy upon non-specific binding to be -4.1 ± 0.9 kcal/mol for CI and -4.2 ± 0.8 kcal/mol for Cro. From this we conclude that about 86% of the CI repressors in a lysogen are non-specifically bound.

2. Materials and methods

2.1. Energetics

Our model is based on monomer–dimer equilibrium constants (at 37°C) of 15 nM for CI [9] and 735 nM for Cro [10]. CI and Cro are supposed to bind both specifically and non-specifically to DNA as dimers. The Gibbs free energy differences (ΔGs) between the water-solvated (unbound) state and the DNA-bound state of CI, Cro, and RNAP in the different specific binding states are listed in table 4 in Darling et al. [7]. The ΔGs of CI and Cro are obtained in vitro at conditions that are supposed to resemble 'physiological' ones [11] while the ΔGs for RNAP are based on in vivo measurements. ΔGs for CI and RNAP are measured at 37°C, which is the temperature of the activity experiments we compare our model to, while ΔGs for Cro are measured at 20°C. Assuming that the affinities (ΔGs) for Cro do not change significantly from 20°C to 37°C, their statistical weight is evaluated through corresponding Boltzmann factors at 37°C.

2.2. Thermodynamic model

The model we apply in this work is based on the statistical-mechanical approach of Ackers et al. [8] that is expanded to include 40 different states [4]. Each state corresponds to a distinct DNA binding combination of CI, Cro, and RNAP. The associated probability f_s for finding the system in one of these 40 states s is

$$f_s = \frac{\exp(-\Delta G(s)/(RT))[\mathrm{CI_2}]^{i_s}[\mathrm{Cro_2}]^{j_s}[\mathrm{RNAP}]^{k_s}}{\sum_s \exp(-\Delta G(s)/(RT))[\mathrm{CI_2}]^{i_s}[\mathrm{Cro_2}]^{j_s}[\mathrm{RNAP}]^{k_s}} \quad (1)$$

where $R = 8.31$ J/(mol K) is the gas constant, $T = 310$ K is the absolute temperature corresponding to the physiological temperature 37°C, and $\Delta G(s)$ is the Gibbs free energy difference (binding energy) between state s and the unoccupied state ($s = 1$). From this model the probabilities for the different binding states may be quantified. In particular,

*Corresponding author. Fax: (45)-35389157.
E-mail addresses: audunba@nordita.dk (A. Bakk), metz@nordita.dk (R. Metzler).

Abbreviations: O_R, right operator; ΔG, Gibbs free energy difference; RNAP, RNA polymerase; NSB, non-specific binding; O_L, left operator

0014-5793/04/$30.00
doi:10.1016/S0014-5793(04)00249-2

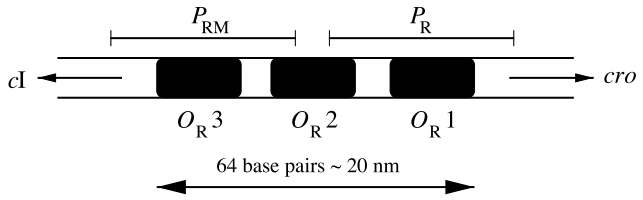

Fig. 1. Schematic illustration of the operator O_R of the λ phage genome. O_R has three binding sites, O_R1, O_R2, and O_R3, where repressor dimers of CI and Cro are able to bind. The promoter regions P_{RM} and P_R indicate where RNA polymerase binds to initiate transcription of the *c*I and *cro* genes. The arrows associated with *c*I and *cro* indicate the transcription direction of these genes, respectively.

the P_{RM} and P_R activities are supposed to be proportional to the probability for binding RNAP to the respective promoters (for details, see [12,13]).

One should note that Eq. 1 will be modified in the case when the left operator (O_L) is present in the λ genome, because this leads to a typical situation in a lysogen where O_R and O_L are linked via a stabilizing octamer [14]. However, in the experiments we compare our model to this DNA loop is not present, because P_R data are obtained at zero CI concentration and P_{RM} data are obtained for an allele without O_L.

2.3. Non-specific binding

The non-specific binding is assumed to be homogeneous, i.e. we calculate an average free energy upon NSB (ΔG_{NSB}) for CI and Cro that obeys the equilibrium [15,16]

$$\exp(-\Delta G_{NSB}/RT) = \frac{[X_2D]}{[X_2][D]} \quad (2)$$

where $[X_2]$ is the free concentration of dimeric species X (CI or Cro), $[X_2D]$ is the concentration of non-specifically bound dimeric proteins, and $[D]$ is the concentration of unoccupied non-specific binding sites. In the regime we consider in this work $[X_2D] \ll [D]$, such that D approximately corresponds to the total number of bases that is 4.6×10^6 for *E. coli*. The conservation equation of the total concentration of species X yields

$$[X_{tot}] = [X_1] + 2[X_2] + 2N_{DNA}([X_{O_R}] + [X_2][D]\exp(-\Delta G_{NSB}/RT)) \quad (3)$$

where the first and second terms on the right hand side are the free monomeric and dimeric concentrations of species X, respectively. The last term in Eq. 3 takes into account the concentrations of specifically (first part) and non-specifically (second part) bound dimers. N_{DNA} is the number of genome equivalents per *E. coli* cell (on average there are 2.3 DNA copies per cell in the CI culture and 2.7 DNA copies per cell in the Cro culture we compare our model to [17]) and $[X_{O_R}]$ is the concentration of bound dimers to O_R per DNA (specific binding) calculated from our thermodynamic model [12,13]. ΔG_{NSB} for CI and Cro are then obtained from an optimized fit (least squares error method) to experimental data of P_{RM} activity [18] and of P_R activity [19,20], respectively.

3. Results and discussion

Comparison between experiment and our model represented by Eqs. 1 and 3 yields the ΔG_{NSB} best fit values -4.1 ± 0.9 kcal/mol for CI and -4.2 ± 0.8 kcal/mol for Cro[1]. In Fig. 2 we show the quality of the fit. Without non-specific binding ($\Delta G_{NSB}\to\infty$) the model overestimates the free concentration of the regulatory proteins, leading to a shift of the activity curves to the left (thin lines in Fig. 2), compared to the experimental data and the best fit with NSB. Thus, the model without NSB cannot reasonably reproduce the in vivo data. Including NSB in the model improves the description of the in vivo experimental data significantly, and NSB clearly emerges as an indispensable ingredient for a proper description of the λ switch regulation. Both data sets for CI and Cro show excellent agreement with the NSB-based model. The ΔG_{NSB} for CI obtained above is close to the in vitro result of Senear and Batey (−3.7 kcal/mol) [21] and Koblan and Ackers (−3.5 kcal/mol) [6]. However, one should note that since non-specific binding is mainly due to electrostatic interactions [22] and therefore likely depends upon ionic strength (in general unknown in vivo) the in vitro results only indicate the in vivo situation in cells. Aurell et al. [23] applied a similar model compared to the model in this work, although with some of the parameter values different, and concluded upon comparison to data from Johnson et al. [24] that $\Delta G_{NSB} < -2$ kcal/mol for CI.

Reinitz and Vaisnys [20] performed a similar analysis of Cro compared to this work, although with a different protein–DNA affinity set, and concluded that introduction of non-specific binding in their model did not improve their fit. However, one should note that Reinitz and Vaisnys [20] did not convert the Cro concentrations taken from [19] to absolute amounts, as we have done here. Since the non-specific binding itself does not alter the shape of the activity curve much, it is not surprising that no detectable difference with or without non-specific binding was obtained by Reinitz and Vaisnys [20]. Takeda et al. [25] also performed in vitro measurements of non-specific Cro binding; however, these are obtained for very short DNA strains (21 bp) yielding a considerably different statistics than treated here and, thus, comparison to our results is not obvious. With a similar analysis as we performed in this work Aurell et al. [23] estimated ΔG_{NSB} for Cro to be −3 kcal/mol by a comparison between

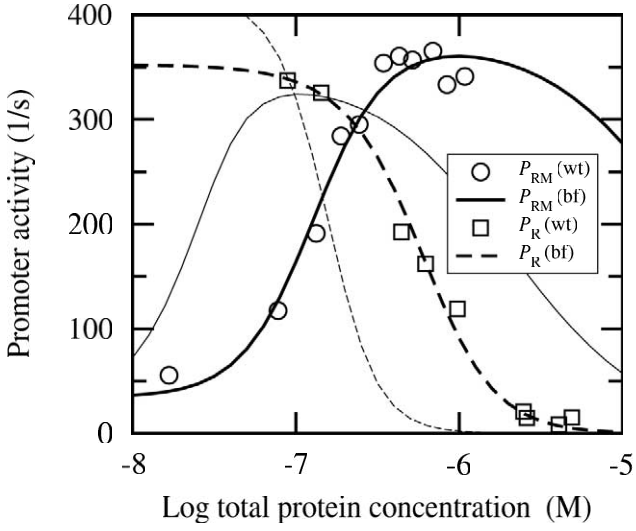

Fig. 2. P_{RM} (LacZ) activity versus total CI concentration (log scale) and P_R (β-gal) activity (rescaled by a factor 0.15) versus total Cro concentration (log scale). P_{RM} (wt), experimental P_{RM} activity data [18]; P_R (wt), experimental P_R activity data [19,20]; bf, best fit of experimental data optimized with respect to ΔG_{NSB}. The two thin curves correspond to P_{RM} activity (solid line) and P_R activity (dashed line) calculated from our thermodynamic model without NSB ($\Delta G_{NSB}\to\infty$).

[1] The error estimates of both ΔG_{NSB} are calculated by including the individual error estimates of the ΔGs [4,6,7] and of the N_{DNA} [17].

the model and half repression of the experimental P_R activity data of Reinitz and Vaisnys [20]. However, we note that some of the parameter values Aurell et al. employed were different compared to the ones we apply in this work. Furthermore, the fit of our model to the experimental data is performed over a large concentration range, leading to an improved determination of ΔG_{NSB} for Cro.

Since there are many parameters involved in the model (16 parameters when both CI and Cro are accounted) it might be interesting to investigate the sensitivity of ΔG_{NSB} upon systematic variations of the other parameters. As an example we choose $\Delta G_{NSB} = -3.6$ kcal/mol for CI, which is a value around the in vitro values of Senear and Batey [21] and Koblan and Ackers [6]. By applying this NSB strength we are not able to fit experimental data satisfactorily by varying, one by one, the other parameters within the expected experimental error (≤ 0.5 kcal/mol). However, with an individual change of typically 1 kcal/mol, compared to wild-type value (absolute value), for five out of the seven relevant protein–DNA binding affinities (when [Cro] = 0) we obtain a satisfactory fit of our model to experimental P_{RM} data [19,20]. Furthermore, with $\Delta G_{NSB} = -3.6$ kcal/mol for CI we have to increase the wild-type value of N_{DNA} by a factor 2 or alternatively the free RNAP concentration must be changed to 150 nM, in order to obtain acceptable fits to the experimental data when the other parameters are fixed to the wild-type values. 150 nM is significantly larger than the originally assumed (and usual) RNAP concentration (30 nM). We note that without NSB in the model, it is not possible to obtain a satisfactory fit of the above mentioned experimental P_{RM} and P_R data for any parameter value, except for one parameter. The latter refers to increasing the number of *E. coli* genomes by the (unrealistic) factor 20 (corresponds to 50–60 genomes per cell).

We estimate that the fraction of non-specifically bound CI versus the total CI concentration is $>80\%$ for total CI concentrations >100 nM. In particular, in a lysogen (~ 250 CI monomeric equivalents per cell) this fraction is $\sim 86\%$. We note that Kao-Huang et al. concluded that a comparable amount (over 90%) of *E. coli lac* repressors (in vivo) is non-specifically bound [11], but we should stress that the *lac* repressor system differs from the λ switch we study here (e.g. *lac* repressors bind to DNA as tetramers). For Cro the non-specific binding fraction is somewhat less than for CI at a given total concentration, but still significant. For total Cro concentrations over 100 nM more than 50% is non-specifically bound.

The large fraction of non-specifically bound proteins found here implies e.g. in a lysogen that only 10–20 free CI dimers exist at any time. This small number of regulatory proteins, within the framework of the statistical model, may lead to problems in explaining the high stability of the λ switch [26,23]. In this respect one might argue that the fluctuations of the free proteins are buffered (damped) against the non-specifically bound proteins, such that effectively the total number of cellular proteins of each species counts and, thus, the impact from noise is smaller than one would expect from the small number of free regulatory proteins.

It has been speculated that the specific binding of proteins may profit from the non-specifically bound portion of the proteins: one-dimensional diffusion of NSB proteins along the DNA combined with three-dimensional diffusion may enhance the efficiency of the search for the specific binding sites [27–29,22]. It will be interesting to pursue this idea further in future studies, in particular in view of our results reported herein, according to which the number of non-specifically bound regulatory proteins is estimated to be large. Finally we note that our ΔG_{NSB} results are significant, as they are based on in vivo data.

Acknowledgements: We thank K. Sneppen for suggesting this work. Discussions with J.S. Høye and K. Sneppen are greatly acknowledged.

References

[1] Ehrenberg, M., Elf, J., Aurell, E., Sandberg, R. and Tegner, J. (2003) Genome Res. 13, 2377–2380.
[2] Ptashne, M. (1992) A Genetic Switch: Phage λ and Higher Organisms, 2nd edn., Cell Press and Blackwell, Cambridge, MA.
[3] Brooks, K. and Clark, A.J. (1967) J. Virol. 1, 283–293.
[4] Shea, M.A. and Ackers, G.K. (1985) J. Mol. Biol. 181, 211–230.
[5] Sarai, A. and Takeda, Y. (1989) Proc. Natl. Acad. Sci. USA 86, 6513–6517.
[6] Koblan, K.S. and Ackers, G.K. (1992) Biochemistry 31, 57–65.
[7] Darling, P.J., Holt, J.M. and Ackers, G.K. (2000) J. Mol. Biol. 302, 625–638.
[8] Ackers, G.K., Johnson, A.D. and Shea, M.A. (1982) Proc. Natl. Acad. Sci. USA 79, 1129–1133.
[9] Koblan, K.S. and Ackers, G.K. (1991) Biochemistry 30, 7817–7821.
[10] Darling, P.J., Holt, J.M. and Ackers, G.K. (2000) Biochemistry 39, 11500–11507.
[11] Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D.W. and von Hippel, P.H. (1977) Proc. Natl. Acad. Sci. USA 74, 4228–4232.
[12] Bakk, A., Metzler, R. and Sneppen, K. (2004) Biophys. J. 86, 58–66.
[13] Bakk, A., Metzler, R. and Sneppen, K. (2004) Israel J. Chem. (in press).
[14] Révet, B., von Wilcken-Bergmann, B., Bessert, H., Barker, A. and Müller-Hill, B. (1999) Curr. Biol. 9, 151–154.
[15] von Hippel, P.H., Revzin, A., Gross, C.A. and Wang, A.C. (1974) Proc. Natl. Acad. Sci. USA 71, 4808–4812.
[16] Bakk, A. and Metzler, R. (2004) (submitted).
[17] Bremer, H. and Dennis, P.P. (1996) in: *Escherichia coli* and *Salmonella*, Vol. II, 2nd edn. (Neidhardt, F.C., Ed.), pp. 1553–1569, ASM Press, Washington, DC.
[18] Dodd, I.B., Perkins, A.J., Tsemitsidis, D. and Egan, J.B. (2001) Genes Dev. 15, 3013–3022.
[19] Pakula, A.A., Young, V.B. and Sauer, R.T. (1986) Proc. Natl. Acad. Sci. USA 83, 8829–8833.
[20] Reinitz, J. and Vaisnys, J.R. (1990) J. Theor. Biol. 145, 295–318.
[21] Senear, D.F. and Batey, R. (1991) Biochemistry 30, 6677–6688.
[22] Gerland, U., Moroz, J.D. and Hwa, T. (2002) Proc. Natl. Acad. Sci. USA 99, 12015–12020.
[23] Aurell, E., Brown, S., Johanson, J. and Sneppen, K. (2002) Phys. Rev. E 65, 051914.1–051914.9.
[24] Johnson, A.D., Poteete, A.R., Lauer, G., Sauer, R.T., Ackers, G.K. and Ptashne, M. (1981) Nature 294, 217–223.
[25] Takeda, Y., Ross, P.D. and Mudd, C.P. (1992) Proc. Natl. Acad. Sci. USA 89, 8180–8184.
[26] Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Science 297, 1183–1186.
[27] Berg, O.G., Winter, R.B. and von Hippel, P.H. (1981) Biochemistry 20, 6929–6948.
[28] von Hippel, P.H. and Berg, O.G. (1989) J. Biol. Chem. 264, 675–678.
[29] Shimamoto, N. (1999) J. Biol. Chem. 274, 15293–15296.