

# Quantitative influence of macromolecular crowding on gene regulation kinetics

Marcin Tabaka, Tomasz Kalwarczyk and Robert Holyst\*

Institute of Physical Chemistry, Polish Academy of Sciences, 44/52 Kasprzaka, 01-224 Warsaw, Poland

Received June 14, 2013; Revised August 28, 2013; Accepted September 17, 2013

## ABSTRACT

**We introduce macromolecular crowding quantitatively into the model for kinetics of gene regulation in *Escherichia coli*. We analyse and compute the specific-site searching time for 180 known transcription factors (TFs) regulating 1300 operons. The time is between 160 s (e.g. for SoxS  $M_w = 12.91$  kDa) and 1550 s (e.g. for PepA<sub>6</sub> of  $M_w = 329.28$  kDa). Diffusion coefficients for one-dimensional sliding are between  $0.003 \mu\text{m}^2\text{s}^{-1}$  for large proteins up to  $0.4 \mu\text{m}^2\text{s}^{-1}$  for small monomers or dimers. Three-dimensional diffusion coefficients in the cytoplasm are 2 orders of magnitude larger than 1D sliding coefficients, nevertheless the sliding enhances the binding rates of TF to specific sites by 1–2 orders of magnitude. The latter effect is due to ubiquitous non-specific binding. We compare the model to experimental data for LacI repressor and find that non-specific binding of the protein to DNA is activation- and not diffusion-limited. We show that the target location rate by LacI repressor is optimized with respect to microscopic rate constant for association to non-specific sites on DNA. We analyse the effect of oligomerization of TFs and DNA looping effects on searching kinetics. We show that optimal searching strategy depends on TF abundance.**

## INTRODUCTION

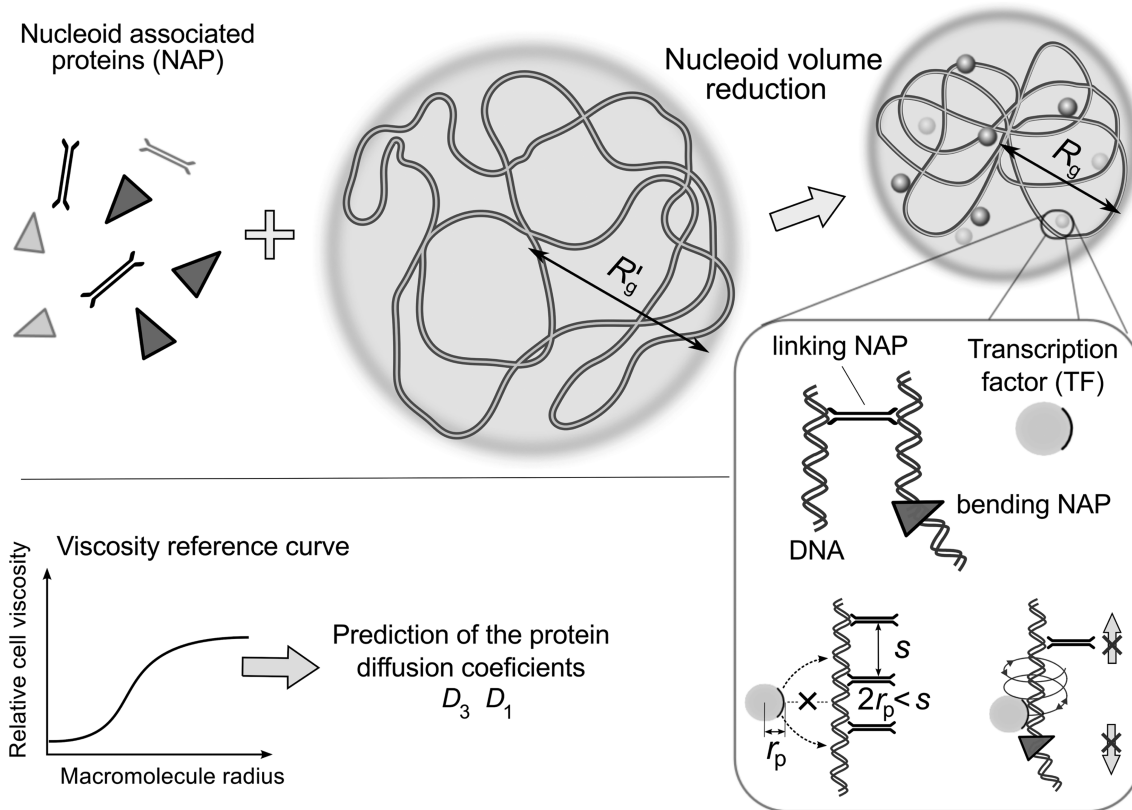
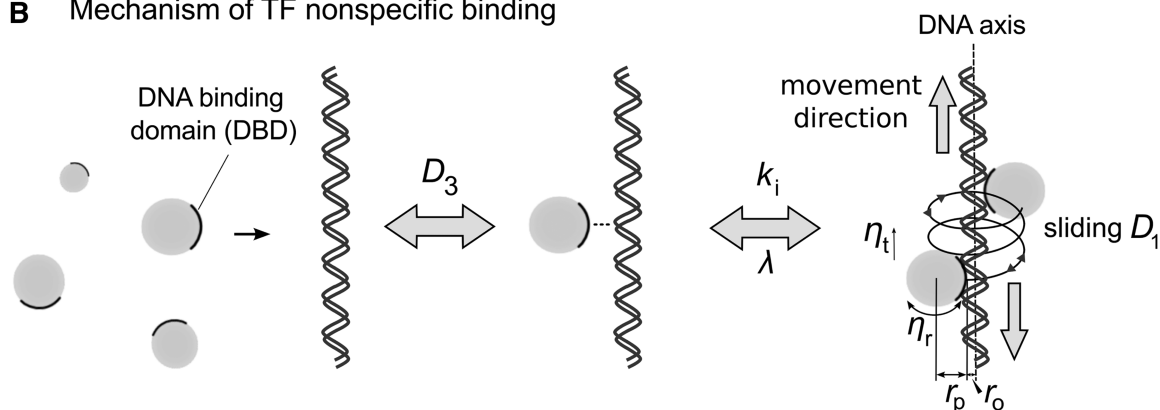
For many decades, macromolecular crowding has been well advertised as a factor strongly affecting all biochemical reactions and other processes in living cells (1–4). Yet, rarely this effect has been quantitatively analysed. We follow classical works on facilitated target location (5–11), extended by Li *et al.* (12) to analyse gene regulations in *Escherichia coli* by transcription factors (TFs). We explicitly include new factors to the model: reduction of mobility of TFs in the cytoplasm and on DNA due to macromolecular crowding (13,14) and reduction of

nucleoid volume caused by nucleoid-associated proteins (NAPs) covering DNA. We use the model of TF diffusion along DNA (sliding) of Blainey *et al.* (15,16). We calibrate the model using non-specific binding constant of LacI to DNA calculated from single-molecule experimental data (17,18). Next we apply the model to all known 180 TFs that bind specifically to *cis*-regulatory elements and regulate 1300 operons. In particular we determine 3D and 1D diffusion constants for TFs, enhancement rates due to the sliding and times for location of operators by TFs. We discuss various strategies used by TFs for rapid target location *in vivo*. We provide a large number of data, which can be used for analysis of genetic networks in systems biology. These data can also guide future experiments, since quantitative experimental data *in vivo* are available for only one repressor (LacI) regulating only one operon.

TFs use diffusion driven by thermal fluctuations to search for specific sites on DNA. The search process combines 3D diffusion of TF through a cytoplasm, subsequent non-specific binding of TF to DNA, next 1D diffusion along a DNA and finally unbinding into cytoplasm (and subsequent repeat of the previous steps) or target recognition when TF encounters an operator during sliding (Figure 1). This process is known as the facilitated target location (5–11). Typical bacterial genome consists of  $\sim 10^6$  base pairs, and location of the specific site on DNA is through enormous amount of bindings to non-specific sites. One-dimensional diffusion along DNA increases effective target size on DNA and speeds the searching up to 2 orders of magnitude (8,12,18). Recent *in vivo* single-molecule experiment has shown that LacI repressor dimer needs  $\sim 350$  s to locate the operator (17) and on average slides along 36 bp when non-specifically bound (18) to DNA. The diffusion coefficient for sliding has been measured in the buffers only (16,17).

The prokaryotic cytoplasm is a highly crowded environment (14,20–22). Two aspects of macromolecular crowding affect regulation of gene expression (Figure 1A): (i) cytoplasmic crowding reduces 3D and 1D diffusion constant of TF and (ii) crowding on DNA disturbs TF non-specific binding and 1D diffusion along DNA due to presence of NAPs (12,23,24). However,

\*To whom correspondence should be addressed. Tel: +48 22 343 3123; Fax: +48 22 343 3333; Email: rholyst@ichf.edu.pl

**A** Effect of macromolecular crowding**B** Mechanism of TF nonspecific binding

**Figure 1.** Influence of macromolecular crowding on gene regulation kinetics. (A) NAPs cause shrinking of available searching volume by linking and bending of DNA chains. NAPs when bound to DNA disturb non-specific association of TFs and 1D diffusion along DNA (12). Macromolecular crowding increases viscosity of the cytoplasm and diminishes TF 1D and 3D-mobility constants [Equations (1–2)]. The viscosity experienced by a translocating macromolecule is a stretched exponential function of size of the macromolecule (13,14). (B) TF searches for specific site on DNA through numerous DNA–non-specific associations. Three-dimensional diffusion brings TF close to DNA. Next, TF binds to DNA with microscopic association rate constant  $k_i$  (8). Non-specifically bound TF slides performing curvilinear motion (16,19) along DNA and dissociates with rate constant  $\lambda$ .

NAPs additionally shrink the nucleoid, thereby decrease the size of space where the process of facilitated target location occurs. Elowitz *et al.* (25) have shown that protein diffusion in the *E. coli* cytoplasm is the normal diffusion, but significantly reduced as compared with its mobility in water. This reduction is a non-linear function of protein molecular weight (21,22). We have introduced

the phenomenological model of cytoplasm length-scale dependent viscosity (13,14) to describe all available experimental data for diffusion of small ligands, proteins and DNA plasmids in *E. coli*. We have previously tested the model (13) for motion of proteins in buffers crowded by polymers (13,26) or micelles (27). In this article, we apply the model to describe the influence of macromolecular

crowding on 1D diffusion along DNA. There is a marked difference between diffusion in the cytoplasm and sliding. During sliding on DNA, the protein follows the path along major grooves and thus rotates around its own axis of rotation and also around the DNA. The friction for rotational motion is proportional to the volume of the protein (15,16), whereas for the translational diffusion in the cytoplasm, the friction is proportional to the radius of the protein. One-dimensional diffusion coefficients are substantially (2 orders of magnitude) smaller than 3D diffusion constants for the typical sizes of TFs.

## MATERIALS AND METHODS

### Macromolecular crowding

The prokaryotic cell is a densely packed structure with concentration of macromolecules between 200 and 400 g/l (20–30% of cellular volume) (20,28). This macromolecular crowding in the cytoplasm reduces in the first place diffusional mobilities of TFs. The reduction of mobility is proportional to the scale-dependent viscosity (13,26). This viscosity describes the mobility in complex liquids (13,26), in cytoplasm of the eukaryotic cells (HeLa and Swiss 3T3 fibroblasts) (13) and prokaryotic cells (*E. coli*) (14). This model based on the viscosity describes protein–protein association kinetics in *HeLa* cytoplasm (29). The model is based on the Stokes-Sutherland-Einstein relation  $D = k_B T / f(r_p)$ , where  $D$  is a diffusion constant of macromolecule,  $k_B T$  is the Boltzmann constant times the absolute temperature and  $f$  is the length-scale-dependent hydrodynamic drag friction. For 3D diffusion coefficient  $D_3$ , this drag has the following form:

$$f(r_p) = 6\pi\eta_t(r_p)r_p \quad (1)$$

Thus  $f(r_p)$  is proportional to the length-scale-dependent viscosity  $\eta_t(r_p)$ . At fixed temperature, the viscosity  $\eta_t$  experienced by an object of hydrodynamic radius  $r_p$  is described by the formula (13,26)  $\eta_t(r_p) = \eta_0 \exp[(R_{eff}/\xi)^\alpha]$ , where  $\eta_0$  is the solvent viscosity,  $\xi$  is the correlation length,  $\alpha \sim 1$  is a constant and effective radius  $R_{eff}$  is defined as  $R_{eff}^{-2} = R_h^{-2} + r_p^{-2}$ , where  $R_h$  is the effective hydrodynamic radius of macromolecular crowding agents. We find for *E. coli* these characteristic parameters by fit to experimental data:  $\xi = 0.51$  nm,  $R_h = 42$  nm and  $\alpha = 0.53$  (14). The length-scale-dependent viscosity of *E. coli* cytoplasm change by a factor of  $10^4$  from 0.001 Pas for small solutes to 13 Pas for large plasmids (250 nm). The calculation of the 1D coefficient  $D_1$  is based on an observation that TFs spin while non-specifically bound to DNA (16). The friction experienced by TF of size  $r_p$  during its curvilinear motion along DNA is given by the following expression (19):

$$f(r_p) = 6\pi\eta_t(r_p)r_p + B^2(8\pi\eta_r(r_p)r_p^3 + 6\pi\eta_t(r_p)(r_o + r_p)^2r_p) \quad (2)$$

where the first term corresponds to frictional force experienced by TF bound to DNA during translational motion parallel to the DNA axis, the second term describes rotational friction that arises from TF rotation

along its axis, and the last term is the rotational friction experienced by TFs that spin around DNA axis (see Figure 1B). TF makes full  $360^\circ$  rotation ( $2\pi$  rotation) around DNA after translating over the distance of 3.54 nm, i.e. 10.4 base pairs along DNA ( $B = 2\pi/10.4l$ ,  $l = 0.34$  nm is a length between consecutive base pairs).  $r_o$  accounts for the protein offset from the DNA axis (Figure 1B). We use the value of  $r_o = 0.6$  nm (19,30). Brownian dynamics simulations (21) have shown that rotational viscosity experienced by rotating proteins of molecular weights in the range  $10 - 10^3$  kDa in *E. coli* cytoplasm is  $\sim 10$  times larger than water viscosity. Therefore, we assume that  $\eta_r(r_p) = 10\eta_0$ .

### Nucleoid volume

NAPs reduce volume of the nucleoid (Figure 1A), by bending and linking DNA strands. There are two types of proteins belonging to NAP family that affect nucleoid volume: bending NAPs that bends DNA and change its Kuhn length and linking NAPs that connect neighbouring DNA chains. We describe the volume reduction using a model of a randomly cross-linked polymer chain of the Kuhn length  $\kappa$ , consisting of  $N$  base pairs subjected to  $M_l$  uncorrelated cross-linking constraints (31). The radius of gyration  $R_g$  of the polymer is (31):

$$R_g = \begin{cases} \kappa(0.26N'/M_l)^{1/2} & \text{if } \epsilon \ll \epsilon_1 \\ \kappa(0.34\epsilon/\kappa)^{1/2}(N'/M_l)^{1/4} & \text{if } \epsilon_1 \ll \epsilon \ll \epsilon_2 \\ \kappa(N'/6)^{1/2} & \text{if } \epsilon \gg \epsilon_2 \end{cases} \quad (3)$$

where  $\epsilon$  is a mean distance between monomers that form a cross-link,  $N'\kappa$  is a contour length of the polymer and  $\epsilon_1 = \kappa(N'/M_l)^{1/2}$ ,  $\epsilon_2 = \kappa(N'M_l)^{1/2}$ . The first case in Equation (3) concerns rigid cross-linking polymer regime where  $\epsilon \sim \kappa$  or smaller, second soft cross-linking regime and the last one describes free polymer chain regime. The volume occupied by nucleoid is calculated from  $V_n = \frac{4}{3}\pi R_g^3$ . The mean distance between bending proteins is  $N/M_b$ , where  $M_b$  is a number of bending NAPs. The Kuhn length is a non-linear function of the mean distance between bending proteins (32,33).

### Model of facilitated target location

We follow the classical works of Berg *et al.* (6–8) with the extension by Li *et al.* (12). The TF association rate constant to a specific site on DNA is as follows:

$$k = \frac{k_{assoc}}{1 + K_{ns}cp(r)}(L) \quad (4)$$

The association rate constant is equal to the rate constant per binding site  $k_{assoc}$  diminished by the presence of non-specific sites on DNA.  $K_{ns}$  is the equilibrium constant for non-specific binding of TF to DNA,  $c$  is the concentration of non-specific DNA,  $p(r)$  is the probability distribution that the binding site is free of NAPs and  $r$  is the diameter of TF.

$$k_{assoc} = \frac{2\pi D_3 l}{\ln(\zeta/2b) + 2\pi D_3 l / k_l} \quad (5)$$

where  $k_i$  is a microscopic (intrinsic) association rate constant (Figure 1B),  $l$  is a distance between consecutive base pairs and  $\zeta$  is a mean distance between neighbouring DNA chains and  $b$  is the DNA radius.  $\zeta$  is calculated as in (8,12) from  $\pi(\zeta/2)^2 Nl = V_n$ . The probability that a binding site of size  $r$  on DNA is free from NAPs is as follows (12,24):

$$p(r) = \left(f + \frac{1-f}{d}\right) \exp\left(-\frac{r(1-f)}{df}\right) \quad (6)$$

where  $d$  is a mean weighted size of NAPs, i.e.  $d = \sum_i M_i d_i / M$ , with  $M$  being the total amount of NAPs per nucleoid and  $M_i, d_i$  are amount and size of  $i$ -th NAP (see Supplementary Table S1), respectively. The fraction of DNA that is free from NAPs is  $f = 1 - \sum_i M_i d_i / N$ . The sliding enhances the rate constant by effectively increasing the mean target size  $\langle L \rangle$ . The size (in bp), under the assumption of semi-stationary positions of NAPs (12), is as follows:

$$\begin{aligned} \langle L \rangle &= \int_r^\infty P(g)(1+L(g-r))dg \\ &= p(r) + P(r) \left(1/2\beta \left(\frac{l_s(1-f)}{4df}\right) l_s^2 - \frac{l_s df}{1-f}\right) \end{aligned} \quad (7)$$

where  $P(g)$  is the probability density function calculated from Equation (6),  $P(g) = \left(f + \frac{1-f}{d}\right) \left(\frac{1-f}{df}\right) \exp\left(-\frac{g(1-f)}{df}\right)$ , and  $L(g) = l_s \tanh(2g/l_s)$  is an effective target size for specific site placed symmetrically between two NAPs (8,12) with  $l_s = 2\sqrt{D_1/\Lambda}$ .  $\beta$  is the  $\beta$ -function (34) given by  $\beta(x) = \frac{1}{2}[\psi(\frac{x+1}{2}) - \psi(\frac{x}{2})]$  with  $\psi$  being an Euler psi function ( $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$ ). The factor 1 denotes that at least one binding site is scanned on TF binding.  $\Lambda$  is the macroscopic dissociation constant (dissociation to the distance  $\zeta/2$  where TF can bind to uncorrelated DNA chain) equal to microscopic dissociation constant  $\lambda$  times the probability of reaching  $\zeta/2$  (8), that is

$$\Lambda = \frac{\lambda(2\pi D_3 l / k_i)}{\ln(\zeta/2b) + 2\pi D_3 l / k_i} \quad (8)$$

## RESULTS

### The presence of NAPs highly reduces nucleoid volume

The length of *E. coli* nucleoid is  $N = 4.64 \times 10^6$  bp. The Kuhn length of undisturbed DNA is 300 bp,  $\kappa = 102$  nm. For the free polymer regime (without NAPs bound to DNA), the nucleoid radius of gyration and volume would be  $R'_g = 5.2 \mu\text{m}$  and  $V'_n = 582 \mu\text{m}^3$ , respectively, that is much larger than *E. coli* cell size. Abundances of DNA binding proteins [NAPs (35,36), TF (37) and RNA polymerase (RNAP) (38)] for early exponential phase of *E. coli* growth in rich media are collected in Supplementary Table S1. The calculated fraction of DNA that is free from DNA binding proteins is  $f = 0.85$  (total level of the proteins is  $M = 36\,600$ ). Number of bending NAPs  $M_b = 26\,000$  shows that mean distance between these NAPs (178 bp) is below Kuhn length of

naked DNA. The random placement of bending protein every 200 bp reduce the persistence length by a factor 2.2 (39). We use the value  $\kappa = 46$  nm in calculations. Thus, we predict that nucleoid shrinking is due to presence of DNA linking and bending proteins. The observed volume of nucleoid is within range  $1-0.1 \mu\text{m}^3$  (40-43). In Supplementary Figure S1, we show that the number of cross-links estimated from (44,45) is enough to shrink nucleoid to the observed values. The length of relaxed MukB homodimer is  $\sim 100$  nm, while on binding to DNA, the protein is compacted to the size below the Kuhn length of DNA (46). Similar case is for SeqA and MatP proteins (47). This is significantly below calculated value of  $\epsilon_1 = 1 \mu\text{m}$ . Therefore, the nucleoid is in the first regime of Equation (3). The mean volume of the nucleoid is between  $V_n = 0.56 - 0.83 \mu\text{m}^3$  at the early exponential phase of growth (Supplementary text and Figure S1). Two other NAPs are commonly attributed to the group of NAPs forming connections between DNA chains: H-NS/StpA dimers (48,49) and Dps dodecamers through the oligomerization (50). However, presence of divalent cations ( $\text{Mg}^{2+}$ ) precludes formation of bridges between DNA chains by these proteins (50,51).

### Non-specific binding of LacI to DNA is activation-limited

We use the results of single-molecule experimental studies (17,18) of specific-site binding kinetics of LacI dimer (fused with Venus proteins). We obtain parameters for the model in the following fitting procedure: (i) In Ref. (18), the fit giving sliding length  $s_l = \sqrt{2D_1/\Lambda}$  does not include presence of NAPs. We develop a simulation procedure (Supplementary text and Supplementary Figure S2) for this experiment, including NAPs on DNA as crowders. The best fit to the model gives  $s^2 = D_1/\Lambda = 1000 \text{ bp}^2$  that corresponds to sliding length equal to  $s_l = 45$  bp (for the vacancy  $f = 0.85$ , and mean weighted size of NAPs  $d = 19$  bp). The sliding length is larger in our model since NAPs weaken the spatial correlations between operators. (ii) The 3D diffusion constant of LacI dimer without DNA binding domains equals (17)  $D_3 = 3 \mu\text{m}^2\text{s}^{-1}$ . We use this value for calculation of hydrodynamic radius of LacI dimer fused with Venus proteins from Equation (1). We get  $r_p = 4.55$  nm. The calculated 1D diffusion constant for *in vivo* conditions is  $D_1 = 0.025 \mu\text{m}^2\text{s}^{-1}$ . (iii) From point (i) and (ii), we calculate  $\Lambda = 215 \text{ s}^{-1}$ . (iv) The non-specific equilibrium constant  $K_{ns}$  is calculated from the equation  $D_3^{eff} = (D_1 t_u + D_3 t_b)/(t_b + t_u) = (D_3 + K_{ns} c p(r) D_1)/(1 + K_{ns} c p(r))$ , where  $t_b$  and  $t_u$  are the average times spent by the protein diffusing in the cytoplasm (with diffusion constant  $D_3$ ) and spent diffusing on DNA (with diffusion constant  $D_1$ ), respectively. The concentration  $c$  of non-specific sites is  $NG/V_c$ , where  $G$  is the amount of DNA in genome equivalents and  $V_c$  is the cell volume (Supplementary text). For  $D_3^{eff} = 0.4 \mu\text{m}^2\text{s}^{-1}$  (17), we get the value  $K_{ns}^{LacI} = 1410 \text{ M}^{-1}$ . (v) From Equation (8), we calculate microscopic parameters:  $\lambda = 265 \text{ s}^{-1}$  and  $k_i = 3.7 \times 10^5 (\text{Ms})^{-1}$ . The  $1/\lambda$  gives the average time spent by the protein on DNA smaller than 5 ms as

observed *in vivo* (17). We get  $k_{assoc} = 3.0 \times 10^5 \text{ (Ms)}^{-1}$ . Such small value of  $k_{assoc}$  suggests that the process of binding of the protein to DNA is activation-limited since in the case of the diffusion-limited association this parameter would have the value  $16.2 \times 10^5 \text{ (Ms)}^{-1}$ . (vi) Using parameters from points (i-v) we find from Equations (4–8) that searching of the single specific site takes  $V_c/(k < O >) = 475 \text{ s}$  for LacI dimer fused with Venus proteins, where  $< O >$  is the average number of operators in the cell volume (Supplementary text). The lower value of  $s^2$  in the fit from pt. (i) is  $s^2 = 700 \text{ bp}^2$  and this gives the time  $385 \text{ s}$ . The experimentally determined rate is  $0.74 \text{ min}^{-1}$  and there are 3–5 LacI dimers (18). Thus, the *in vivo* binding time is in the range (240–405 s).

### LacI repressor and RNA polymerase are optimized for fast target location

The average time for target location is a function of microscopic association and dissociation constants and can be reduced by changing these two constants. In the previous paragraph, we have discussed the experiment involving LacI dimer, but LacI protein in the native form attaches to the operator as the tetramer. Molecular weight of LacI tetramer is 154.36 kDa. Its hydrodynamic radius (in nm) is determined from the following equation (52):

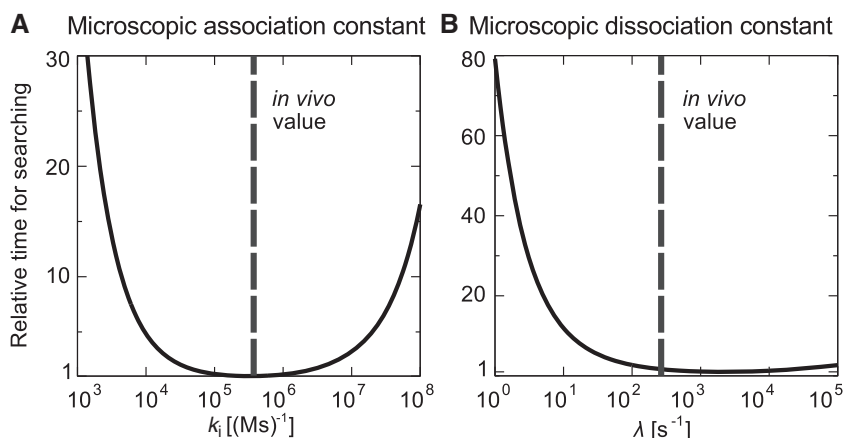
$$r_p = 0.0515M_w^{0.392} \quad (9)$$

The calculated mobility constants (14)  $D_3 = 1.7$ ,  $D_1 = 0.01 \mu\text{m}^2\text{s}^{-1}$  give the overall search time equal to 730 s. Figure 2 depicts the relative average time of location of the specific site as a function of one of the microscopic rate constants (the second is kept constant). We find that the searching time has a minimum as a function of microscopic association rate constant  $k_i$ . The minimization is an effect of two opposite processes. Increase of microscopic constant enhances the rate of TF binding to non-specific site and increase the overall association rate [Equation (4)]. The increase is limited

as  $k_{assoc}$  becomes diffusion-limited, i.e. for  $k_i \rightarrow \infty$ . Additionally it increases non-specific equilibrium constant and decrease the overall rate by tight binding to non-specific sites. For large values of  $k_i$ , TF is recaptured to the same site after dissociation, and thus the search process is limited by 1D sliding. The association constant for *in vivo* conditions differ from that in the minimum by a factor 1.002.

The reduction of association rate due to transient binding to non-specific sites [by a factor  $(1+K_{ns}cp)$ , denominator of Equation (4)] is larger than decrease of the effective target length due to the sliding ( $L$ ). Therefore, the time for target location reaches a constant value before sliding length is reduced to 1 bp. The minimum is for  $K_{ns}cp \approx 1$  (8). It corresponds to  $\lambda = 2200 \text{ s}^{-1}$  (Figure 2B), i.e. it is 1 order of magnitude larger than calculated for *in vivo* conditions. However, the minimum is shallow and the calculated difference in search time for *in vivo* conditions and at the minimum is only 66%. We suspect that high dissociation rate for optimal searching time is not possible for a large macromolecule.

We have done similar analysis for *E. coli* single promoter localization time during transcription initiation. Calculation of the mobility constants for RNA polymerase holoenzyme complex (with  $\sigma^{70}$  subunit bound,  $M_w^{\text{RNAP}} = 458 \text{ kDa}$ ) gives  $D_3 = 0.47$  and  $D_1 = 0.0016 \mu\text{m}^2\text{s}^{-1}$ . The microscopic dissociation constant,  $\lambda = 34.2 \text{ s}^{-1}$ , for RNAP has been recently determined (53) from single molecule measurements. The equilibrium constant for non-specific binding of RNAP to DNA has not been measured so far *in vivo*, but its estimation for cellular conditions gives a value  $K_{ns}^{\text{RNAP}} = 300 \text{ M}^{-1}$  (54) and *in vitro* measurements for high ionic strength with divalent cations that mimics milieu of *E. coli* cytoplasm gives  $3000 \text{ M}^{-1}$  (55). In Supplementary Figure S3, we show that minimum of searching time as a function of microscopic association time is between  $k_i$  values calculated from the above equilibrium constants. It has been observed recently in



**Figure 2.** Relative changes of the average time for location of the operator by LacI tetramer. (A) Microscopic association constant is optimized for fast target location. The times are given with respect to the minimal value. (B) The minimum of the time as a function of the microscopic dissociation rate constant is shifted by 1 order of magnitude from determined value. However, the decrease of the time is only by a factor 1.66. The dashed lines point to determined values of microscopic rate constants.

independent *in vitro* experiments that RNA polymerase with  $\sigma^{70}$  (53) or  $\sigma^{54}$  (56) factors does not perform long-range sliding when searching for its promoter. Our calculations also give short-range sliding of the length 28 bp for RNAP- $\sigma^{70}$  complex in the cell. The measured lifetime of RNAP- $\sigma^{54}$  with non-specific DNA is equal to 4.7 s but such long non-specific binding was not observed for  $\sigma^{70}$  complex (53) pointing that this subunit may interact differently with non-specific DNA.

#### Diffusion constants and average time for the location of a single site on DNA for all recognized TFs

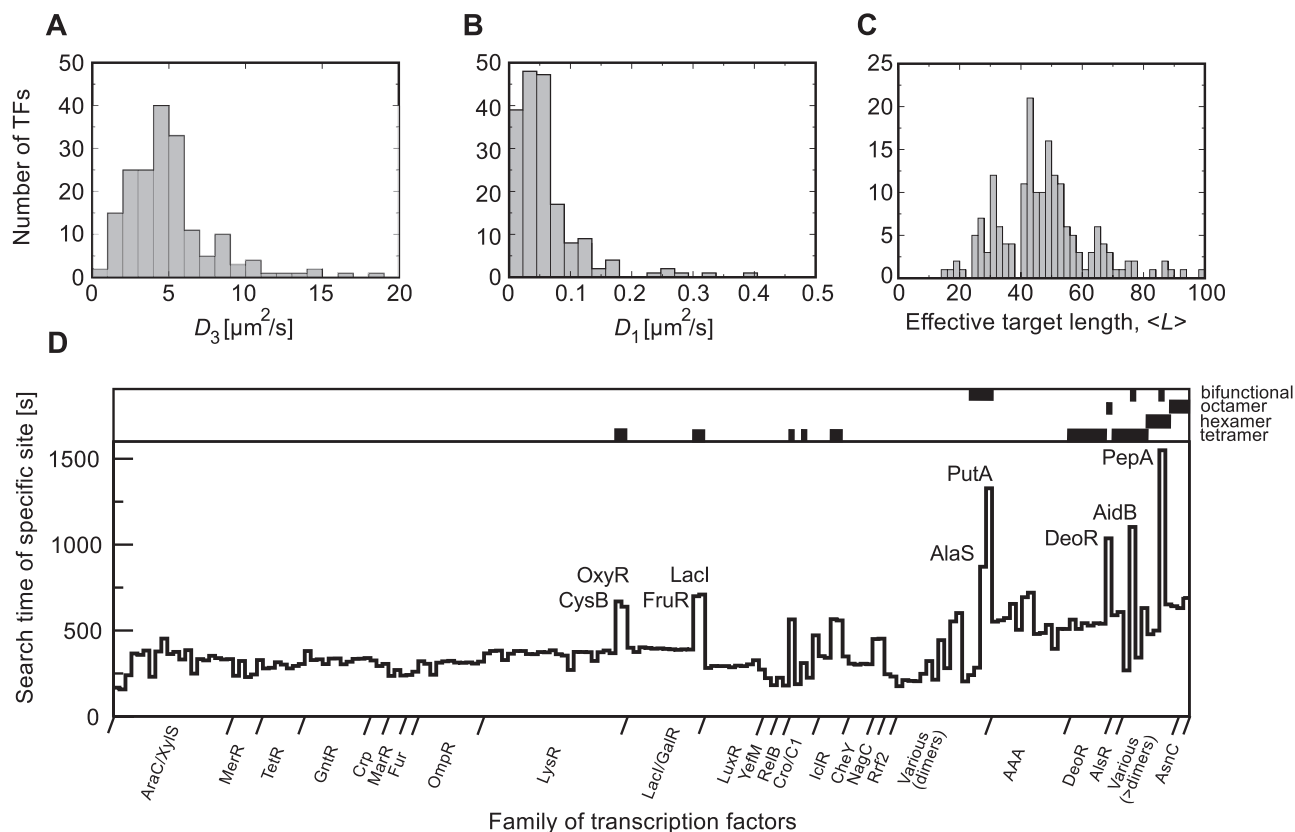
The *E. coli* genome contains  $\sim 300$  genes that encode proteins that regulate promoter activity (57). So far, more than half of them have been verified experimentally. Here we analyse kinetic and dynamic parameters characterizing the process of target location for all 180 verified TFs (regulating  $\sim 1300$  operons) for *in vivo* conditions. We get a list of all TFs and information about their monomer molecular weights (Supplementary Table S2) from *E. coli* EcoCyc (58) and RegulonDB (59) databases. We extend the EcoCyc database for TF oligomerization levels by mining of this parameter from the literature (References in Supplementary Table S2). We assume oligomerization level that is most common in a given TF family, if this information is unavailable for a given TF from this family. In Figure 3A–B, we show histograms for calculated 3D and 1D diffusion constants for all TFs. The TF mobility on DNA is  $\sim 2$  orders of magnitude slower than its cytoplasm mobility. The calculation of single-site localization times on DNA by TFs requires additional assumption about non-specific microscopic rate constants. The unknown rate parameters for all TFs ( $k_i$  and  $\lambda$ ) are assumed to be equal to that for *lac* repressor. The assumption is justified by the fact that nature of non-specific interactions of TF with DNA is primarily electrostatic (30,60,61) between the positively charged side chains of proteins and the negatively charged backbone phosphate groups of the DNA. Dahirel *et al.* (62) have shown that DNA–protein non-specific interaction free energy has a minimum at finite separation between interacting interfaces equal to 0.6 nm that is in agreement with the distance observed for the non-specific LacI–DNA complex (30). The minimum is an effect of the concave shape of DNA–proteins and an enhancement of the osmotic pressure at short separation distances. They performed also statistical analysis of 77 DNA binding proteins finding for sequence-specific DNA proteins similar interface area and mean charge density. Thus, it is expected that non-specific interaction should not vary between TFs. Moreover, *in vivo* experiments give similar values of the fraction of various non-specifically bound TFs on DNA:  $\lambda$  phage regulatory protein CI in  $\lambda$ -infected *E. coli* cells was measured to be 86% bound to non-specific sites (63) [non-specific equilibrium constants for CI and Cro repressors were identical (63)], 87% of LacI dimers fused with Venus proteins from single-molecule experiments (17) and 90% of LacI tetramers from measurements of the distribution of repressors in minicell-experiment (64).

The site-specific location times for higher-order oligomers and bifunctional TFs are substantially decreased. The association rate effective enhancement factor  $\langle L \rangle$  (Figure 3C) due to the sliding changes by 1 order of magnitude (between 10 for TF with large molecular weights/high level of oligomerization to 100 for TFs that are monomers or small dimers). In Figure 3D, we present the calculated TF searching times for single specific site. The TFs are grouped in the families they belong to (Supplementary Table S2). We mark in the upper belt the TFs that form stable higher-order oligomers in solution (oligomerization level  $> 2$ ) or are bifunctional, i.e. contain both DNA binding domain (DBD) and active centres responsible for enzymatic activity (e.g. AlaS - alanyl-tRNA synthetase, PutA - flavo-protein that acts as enzyme catalysing reactions of the proline degradation, AidB - isovaleryl-CoA dehydrogenase activity, PepA - aminopeptidase). Within the same family, the searching times are usually similar (except they belong to one of above groups).

#### TF's searching strategies (higher oligomers versus dimers)

A TF binding time to the regulatory region of the operon can be reduced by increasing TF concentration. Downside of this is an additional cost of protein production (65,66). The best strategy for fast regulation seems to be reached by small monomers with bipartite helix-turn-helix (HTH) motifs (as for MarA, SoxS) and single binding site in the regulatory region. However, this is not the case observed for *E. coli* regulatory motifs. Frequently, regulatory regions of the operons contain many binding sites and transcription regulation is a complex function of the number of these sites (67,68). TF dimers are the most frequent oligomeric forms. Stable higher-order oligomers occur for TFs that regulate transcription of the operons involved in transport and catabolism of sugars [e.g. lactose (LacI), glycerol-P (GlpR), glucitol (GutR), L-ascorbate (UlaR), deoxyribonucleoside (DeoR), N-acetylgalactosamine (AgaR), cell wall sugars (MurR), D-allose (AlsR)]. Higher oligomers are capable of forming DNA loops (69) that efficiently stabilize the TF–DNA complex with dissociation rates, e.g. for LacI tetramer as low as  $\sim 10^{-4} \text{ s}^{-1}$  (9,70). DNA looping reduces additionally searching time for the main operator (12) within regulatory region. Many motifs are used to take advantage of DNA looping: presence of auxiliary operators within structural genes (e.g. LacI, GlpR), within regulatory regions (DeoR, AgaR, PepA, IclR), or presence of single regulatory region with multiple operators within divergent operons (AgaR, MurR, AlsR, YqjI). If auxiliary operators are separated by distances higher than sliding length of TF, they reduce searching time by a factor equal to the number of all operators (12,18). On the other side, other repressors, e.g. GalR and GalS, use DNA looping but binds to operators as dimers. Therefore, we pose a question: what is the optimal strategy for fast regulation in case when TFs use DNA looping.

We analyse the simplest searching strategies in which regulatory region is composed of  $m$  independent operators and TF is a complex with oligomerization level of dimers equal to  $o$ . We consider two scenarios of formation of an



**Figure 3.** Analysis of TF mobility and kinetic parameters for the set of recognized TFs (Supplementary Table S2) from *E. coli* EcoCyc (58) and RegulonDB (59) databases. Histograms of TF (A) three-dimensional,  $D_3$  [Equation (1)] and (B) one-dimensional diffusion constants,  $D_1$  [Equation (2)]. (C) The histogram of the effective target enhancement factors  $\langle L \rangle$  [Equation (7)]. (D) The average time for the location of single specific binding site by TFs [ $V_c/(k \langle O \rangle)$ ]. For simplicity, we assume that  $\langle O \rangle = G$ . TFs are grouped into families (the order is the same as in Supplementary Table S2). In the upper belt of the panel (D), TFs that form stable higher oligomers (oligomerization level higher than 2) and/or are bifunctional (TF poses DNA binding domain and various enzymatic activity) are marked with black stripes.

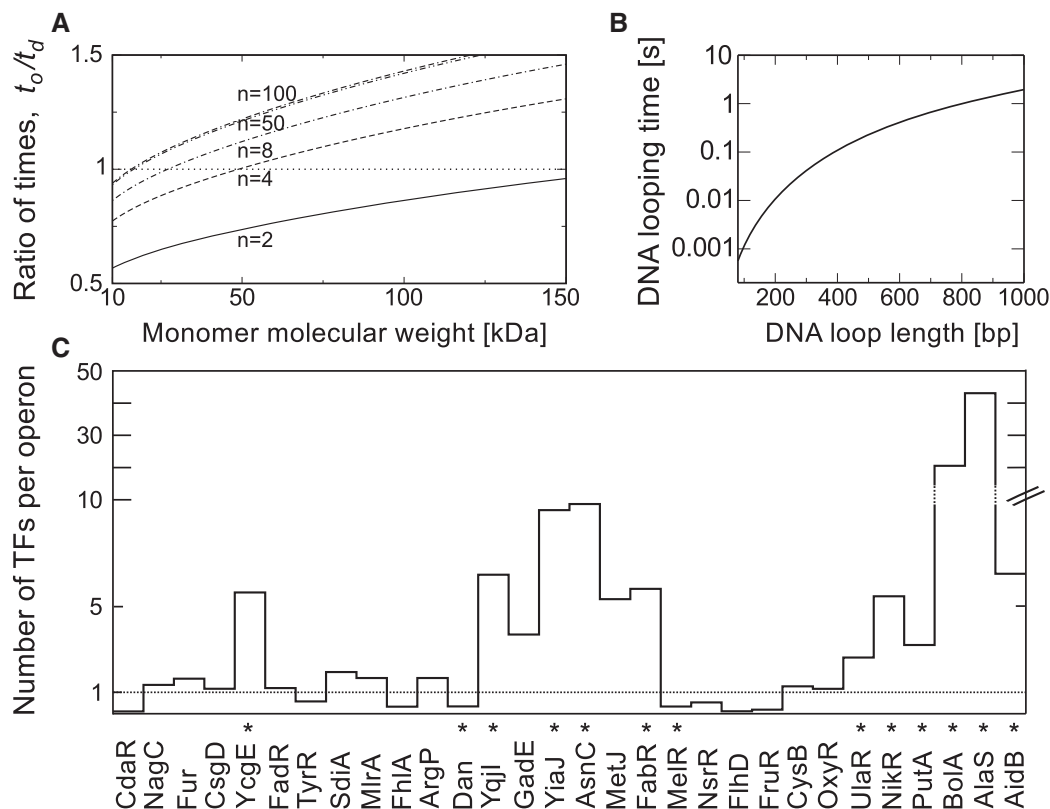
oligomeric TF-operators complex: (i) via independent searching by  $n$  dimers and oligomerization of dimers when specifically bound via DNA looping (formation

time  $t_d = \sum_{i=0}^{o-1} 1/(n-i)(m-i)k$  or (ii) via searching by

$n/o$  oligomers and formation of DNA loops by means of auxiliary operators with time  $t_o = o/mnk$ . The optimal strategy depends on the number of searchers ( $n$  dimers versus  $n/2$  tetramers, case  $m = o$ ) (Figure 4A) and molecular weight of the TF monomer. Assumption of equal number of TF monomers excludes effect of costs of TF production. Case with one stable tetramer is favoured over independent searching by two dimers. The second scenario is the best strategy for high concentration of TFs. We exclude from calculations the characteristic times of DNA loop formation, since this process for *in vivo* cellular crowding conditions is expected to take place orders of magnitude faster than specific site searching (Figure 4B).

Operons are regulated by low-copy number of TFs. We use data from single-molecule studies of protein copy number in *E. coli* cell (37) to get abundances of oligomeric TFs. We compare the values with the numbers of regulated operons [taken from RegulonDB (59)]. TFs for which both information are available are

shown in Figure 4C. Divergent operons with the same operator are counted as a single unit. The results shed light on the problem of trade-off between TF's speed of transcription regulation and costs of TF synthesis. TFs that regulate large regulons are present approximately in copies that correspond to single TF per operon. Contrary, regulons that consist of few operons (marked by stars in Figure 4) appeared to be regulated by TF levels higher than one copy per operon. In this group, the most abundant TFs are bifunctional proteins that reduce significantly searching times for specific sites. Possible explanation of these results arises from the fact that the average binding time of TF to the operator of the specific operon in a regulon decreases with the regulon size (Supplementary Figure S4). We consider the simplified problem in which the number of TFs equals the number of regulated operons and every operon contains single operator. Even regulon that consists of four operons and four TFs reduces, on average, the binding time to the particular operator by 50%. When the number of operons increases to 40, the reduction is 90%. The same effect can be achieved when TF regulates single operon by increasing the number of TFs 2- and 10-fold, respectively.



**Figure 4.** Searching strategy for the formation of TF complex at the operator depends on the number of searchers. (A) Ratio,  $t_o/t_d$ , of the complex formation time via searching by oligomers (abundance  $n/o$ ) to the time of independent searching by dimers (abundance  $n$ ). Case with  $o = 2$  (TF complex is a tetramer), and two independent specific sites on DNA within operator region ( $m = 2$ ). (B) DNA looping times as a function of base pairs that form a loop for *in vivo* macromolecular crowding. The time calculation is based on DNA relaxation times (71,72) of size  $L$  from  $k_f = \frac{\pi k_b T \ln(L/2b)}{4\eta L^2}$ . The hydrodynamic radius of DNA is  $r_{DNA} = C * M_w^\alpha$  with  $C = 0.024$  and  $\alpha = 0.57$  (73) and molecular weight of the 1 bp is 660Da. The viscosity  $\eta$  experienced by moving DNA is calculated as previously (14). (C) Mean number of oligomeric TFs per regulated operon. Data are taken from single-molecule studies of protein copy number in *E. coli* cell (37). The dotted line corresponds to the ratio value equal to 1. By means of stars we marked TFs that regulate only one or two operons.

## DISCUSSION

We present a first comprehensive study of the impact of macromolecular crowding on the gene regulation kinetics in *E. coli*. Our approach combines the theory of scale-dependent diffusion (13,26), nucleoid shrinking by NAPs (31) and model of facilitated target location (8,12). This methodology is crucial for predictions of 3D and 1D diffusion constants and specific-site location times for all currently known TFs. From the parameterization of the model, we find that binding of LacI repressor to non-specific DNA is activation-limited reaction. Moreover, the microscopic association rate is optimized for fast target location for *in vivo* conditions.

From the seminal work of Riggs *et al.* (74) on association kinetics of LacI repressor to the operator site, there is broad interest in processes involved in a target location by TF (75). Recent measurements of TF searching (17,18) point that TF make use of the processes of facilitated diffusion to locate its specific site *in vivo*. Here, our analysis shows that presence of 1D diffusion speeds up the specific-site location for all TFs by 1–2 orders of magnitude as compared with the case where the TF binds with the same affinity to non-specific sites but does not slide.

The previous models concerning *in vivo* facilitated diffusion of LacI repressor (76,77) pointed that the repressor search can be close to the optimal conditions for target location. Koslover *et al.* (76) used value of LacI dissociation constant equal to  $200 \text{ s}^{-1}$  that comes from the observation that the upper value of LacI unbinding time is 5 ms (17). Both models pointed that the DNA configuration has insignificant impact on the search time. In our work, we calculated all the parameters for LacI repressor on the basis of *in vivo* data. Specifically, taking the results of experiments by Hammar *et al.* (18) and the macromolecular crowding present in the cell, we calculated the value of microscopic parameters proving that LacI optimizes the target location with respect to the binding rate constant. Thus the parameters include implicitly (78) macromolecular crowding present in the cell.

Measured 1D diffusion constant *in vitro* is  $0.046 \mu\text{m}^2\text{s}^{-1}$  (17) in a reaction buffer characterized by water viscosity. The value is less than predicted from Equation (2). The difference is explained by presence of small energy barrier in helical sliding of TF along DNA during transition between consecutive base pairs (16). This effect was incorporated by additional factor  $F(\epsilon) = \exp(-(\epsilon/k_B T)^2)$  (79) multiplying right-hand side of Equation (2), where  $\epsilon$  is



root-mean-squared roughness of the interaction potential. The value of  $\epsilon$  for LacI was determined to be  $1.35k_{BT}$  (16) for experiment carried out without  $Mg^{2+}$  in the reaction buffer. One-dimensional diffusion on DNA is much faster in the presence of  $Mg^{2+}$  (70,80). It has been shown (70) that magnesium cations change the diffusion constant of LacI repressor from  $0.15 \mu m^2 s^{-1}$  in the presence of 10 mM  $Mg^{2+}$  to  $0.045 \mu m^2 s^{-1}$  for buffers without  $Mg^{2+}$  [the divalent cations stabilize the double helix of DNA (81) and concentration of  $Mg^{2+}$  *in vivo* is evaluated to be 5–10 mM (61) or even 100 mM (82)]. More studies are required to quantify this effect as a function of  $Mg^{2+}$  concentration. We simply assume that protein 1D diffusion constant is decreased only by cytoplasm viscosity. Including energy barrier  $\epsilon = 1.35k_{BT}$  into the model gives the following values: 1D diffusion constant  $D_1 = 4 \times 10^{-3} \mu m^2 s^{-1}$ , microscopic dissociation constant  $\lambda = 36 s^{-1}$  and target location time 3000 s. The values disagree with the experimental observation for which  $1/\lambda < 5$  ms and searching times are 1 order of magnitude smaller.

TF must scan DNA base pairs (18) to recognize specific site (23,83,84). We neglected scanning in the model at non-specific sites because the scanning rates are unknown. It has been shown by Sheinman *et al.* (84) that model in which the scanning energy barrier varies between non-specific sites gives appreciable increase of searching times. The free-energy calculations between DNA and protein give argument for the existence of such barriers (62). The model allows also for optimal search that is equivalent to the classical model of facilitated target location. Since the calculated value of searching time for LacI dimer is similar to the experimentally measured one, we suspect that the TF trapping at non-specific sites does not affect significantly binding times.

We assume that motion of all TFs and RNAP is diffusive. This is confirmed by many experimental results for proteins of large range of sizes [(14,22,25) and references therein]. Genomic loci (85) and large RNA–protein complexes (86) subdiffuse in the *E. coli* cytoplasm. The reason of subdiffusive motion is confined diffusion in a polymer chain and activation processes (87).

We assume in our model that TFs are initially uniformly distributed within a cell. Therefore, our calculations concern a case in which initial position of TF is not correlated with the region of its synthesis (88–90). The assumption is justified by the observation that for fast-growth conditions, considered here, DNA is less densely packed and distribution of LacI does not depend on the spatial location of its encoding gene (89). Also repression strength as a function of intergenic distance is well reproduced by model assuming spatial homogeneity of the repressor (89). Intersegmental jumps that speed up the target location in coiled DNA (91–93) are implicitly embedded in classical model of facilitated diffusion. Since the values of the microscopic and macroscopic dissociation constants are comparable, most of the unbinding events lead to the association to uncorrelated DNA sites.

Our findings also shed light on trade-off between facilitated target location by TFs due to DNA looping,

increase of TF concentration and costs of TF production. The increase of TF concentration weakens also the fluctuations in the gene products (94–97). The precise regulation of transcription initiation is easier for dimers with regulation at the dissociation step and direct interactions between dimers. Here, more parameters can be adjusted to reach a gene transfer function that effectively couples the expression of genes to environmental signals. Searching by dimers (98) with further oligomerization at the operator is also preferred in AAA family of TFs (ATPase associated with diverse cellular activities) (Supplementary Table S3). These TFs regulate gene expression in  $\sigma^{54}$ -RNAP-dependent promoters. AAA TFs when bound to upstream activating sequences (UAS) function by coupling the energy yielded from ATP hydrolysis to the isomerization of RNAP from closed to open complex. TF oligomerization is required for the formation of functional ATPase and it is promoted by the binding to UAS (99). Therefore, searching by high-oligomeric TF is disadvantageous owing to consumption of ATP when not bound with UAS.

System biology studies of genetic networks require enormous amount of data. Most of them are still unavailable from experiments. Database built here on TF mobilities and searching kinetics can now be used in studies of functioning of these networks with proper temporal resolution to give accurate predictions of gene products fluctuations.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

M.T. and R.H. designed research; M.T. and R.H. performed research; M.T., R.H. analysed data with input from T.K.; and M.T. and R.H. wrote the article. We thank N. Bobrowska for help with preparation of figures.

## FUNDING

Ministry of Science of Poland for the Iuventus-Plus program [IP2010 052570 (2011) to M.T.]; National Science Center [DEC1-2011/01/N/ST3/00865 to T.K.]; Foundation for Polish Science for START scholarship [to T.K.]; National Science Center for MAESTRO grant [DEC-2011/02/A/ST3/00143 to R.H.]. Funding for open access charge: MAESTRO grant from National Science Center of Poland [DEC-2011/02/A/ST3/00143].

*Conflict of interest statement.* None declared.

## REFERENCES

- Zimmerman, S.B. and Minton, A.P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.*, **22**, 27–65.
- Ellis, R. (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.*, **11**, 114–119.

3. Ellis, R. (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.*, **26**, 597–604.
4. Zhou, H., Rivas, G. and Minton, A. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Ann. Rev. Biophys.*, **37**, 375–397.
5. Richter, P.H. and Eigen, M. (1974) Diffusion controlled reaction rates in spheroidal geometry. Application to repressor–operator association and membrane bound enzymes. *Biophys. Chem.*, **2**, 255–263.
6. Berg, O.G. and Blomberg, C. (1976) Association kinetics with coupled diffusional flows. Special application to the lac repressor–operator system. *Biophys. Chem.*, **4**, 367–381.
7. Berg, O.G. and Blomberg, C. (1977) Association kinetics with coupled diffusion. An extension to coiled-chain macromolecules applied to the lac repressor–operator system. *Biophys. Chem.*, **7**, 33–39.
8. Berg, O.G., Winter, R.B. and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
9. Winter, R.B., Berg, O.G. and von Hippel, P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry*, **20**, 6961–6977.
10. von Hippel, P.H. and Berg, O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.*, **264**, 675–678.
11. Halford, S.E. and Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
12. Li, G.W., Berg, O.G. and Elf, J. (2009) Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat. Phys.*, **5**, 294–297.
13. Kalwarczyk, T., Ziebac, N., Bielejewska, A., Zaboklicka, E., Koynov, K., Szymański, J., Wilk, A., Patkowski, A., Gapiński, J., Butt, H.J. *et al.* (2011) Comparative analysis of viscosity of complex liquids and cytoplasm of mammalian cells at the nanoscale. *Nano Lett.*, **11**, 2157–2163.
14. Kalwarczyk, T., Tabaka, M. and Holyst, R. (2012) Biologistics–diffusion coefficients for complete proteome of *Escherichia coli*. *Bioinformatics*, **28**, 2971–2978.
15. Blainey, P.C., van Oijen, A.M., Banerjee, A., Verdine, G.L. and Xie, X.S. (2006) A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA*, **103**, 5752–5757.
16. Blainey, P.C., Luo, G., Kou, S.C., Mangel, W.F., Verdine, G.L., Bagchi, B. and Xie, X.S. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.*, **16**, 1224–1229.
17. Elf, J., Li, G.W. and Xie, X.S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**, 1191–1194.
18. Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G. and Elf, J. (2012) The lac repressor displays facilitated diffusion in living cells. *Science*, **336**, 1595–1598.
19. Bagchi, B., Blainey, P.C. and Xie, X.S. (2008) Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B*, **112**, 6282–6284.
20. Fulton, A. (1982) How crowded is the cytoplasm? *Cell*, **30**, 345–347.
21. McGuffee, S. and Elcock, A. (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comp. Biol.*, **6**, e1000694.
22. Mika, J.T. and Poolman, B. (2011) Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotechnol.*, **22**, 117–126.
23. Slutsky, M. and Mirny, L.A. (2004) Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.*, **87**, 4021–4035.
24. Flyvbjerg, H., Keatch, S. and Dryden, D. (2006) Strong physical constraints on sequence-specific target location by proteins on DNA molecules. *Nucleic Acids Res.*, **34**, 2550–2557.
25. Elowitz, M.B., Surette, M.G., Wolf, P.-E., Stock, J.B. and Leibler, S. (1999) Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.*, **181**, 197–203.
26. Hołyst, R., Bielejewska, A., Szymański, J., Wilk, A., Patkowski, A., Gapiński, J., Zywoćński, A., Kalwarczyk, T., Kalwarczyk, E., Tabaka, M. *et al.* (2009) Scaling form of viscosity at all length-scales in poly(ethylene glycol) solutions studied by fluorescence correlation spectroscopy and capillary electrophoresis. *Phys. Chem. Chem. Phys.*, **11**, 9025–9032.
27. Szymanski, J., Patkowski, A., Wilk, A., Garstecki, P. and Holyst, R. (2006) Diffusion and viscosity in a crowded environment: from nano-to macroscale. *J. Phys. Chem. B*, **110**, 25593–25597.
28. Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.*, **222**, 599–620.
29. Tabaka, M., Sun, L., Kalwarczyk, T. and Holyst, R. (2013) Implications of macromolecular crowding for protein–protein association kinetics in the cytoplasm of living cells. *Soft Matter*, **9**, 4386–4389.
30. Kalodimos, C.G., Biris, N., Bonvin, A.M., Levandoski, M.M., Guennegues, M., Boelens, R. and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science*, **305**, 386–389.
31. Vilgis, T. and Solf, M. (1997) Cross-linked polymer chains: scaling and exact results. In: Rubi, M. and Perez-Vicente, C. (eds), *Complex Behaviour of Glassy Systems*. Springer, Berlin Heidelberg, pp. 403–413.
32. van Noort, J., Verbrugge, S., Goosen, N., Dekker, C. and Dame, R.T. (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc. Natl Acad. Sci. USA*, **101**, 6969–6974.
33. Nir, G., Lindner, M., Dietrich, H.R., Girshevitz, O., Vorgias, C.E. and Garini, Y. (2011) HU protein induces incoherent DNA persistence length. *Biophys. J.*, **100**, 784–790.
34. Gradshteuin, I.S. and Ryzhik, I.M. (2000) In: Jeffrey, A. and Zwillinger, D. (eds), *Table of Integrals, Series, and Products*. Academic Press, Burlington.
35. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J. Bacteriol.*, **181**, 6361–6370.
36. Petrushenko, Z.M., Lai, C.H. and Rybenkov, V.V. (2006) Antagonistic interactions of kleisins and DNA with bacterial Condensin MukB. *J. Biol. Chem.*, **281**, 34208–34217.
37. Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X.S. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, **329**, 533–538.
38. Bremer, H., Dennis, P. and Ehrenberg, M. (2003) Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie*, **85**, 597–609.
39. Czapla, L., Swigon, D. and Olson, W.K. (2008) Effects of the nucleoid protein HU on the structure, flexibility, and ring-closure properties of DNA deduced from Monte Carlo simulations. *J. Mol. Biol.*, **382**, 353–370.
40. Hecht, R., Taggart, R. and Pettijohn, D. (1975) Size and DNA content of purified *E. coli* nucleoids observed by fluorescence microscopy. *Nature*, **253**, 60–62.
41. Van Ness, J. and Pettijohn, D.E. (1979) A simple autoradiographic method for investigating long range chromosome substructure: size and number of DNA molecules in isolated nucleoids of *Escherichia coli*. *J. Mol. Biol.*, **129**, 501–508.
42. Valkenburg, J.A. and Woldringh, C.L. (1984) Phase separation between nucleoid and cytoplasm in *Escherichia coli* as defined by immersive refractometry. *J. Bacteriol.*, **160**, 1151–1157.
43. Wang, W., Li, G.W., Chen, C., Xie, X.S. and Zhuang, X. (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, **333**, 1445–1449.
44. Badrinarayanan, A., Reyes-Lamothe, R., Uphoff, S., Leake, M.C. and Sherratt, D.J. (2012) *In vivo* architecture and action of bacterial structural maintenance of chromosome proteins. *Science*, **338**, 528–531.
45. Cagliero, C., Grand, R.S., Jones, M.B., Jin, D.J. and O’Sullivan, J.M. (2013) Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription. *Nucleic Acids Res.*, **41**, 6058–6071.
46. Matoba, K., Yamazoe, M., Mayanagi, K., Morikawa, K. and Hiraga, S. (2005) Comparison of MukB homodimer versus

- MukBEF complex molecular architectures by electron microscopy reveals a higher-order multimerization. *Biochem. Biophys. Res. Commun.*, **333**, 694–702.
47. Dupaigne, P., Tonthat, N.K., Espéli, O., Whitfill, T., Bocard, F. and Schumacher, M.A. (2012) Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chromosome. *Mol. Cell*, **48**, 560–571.
  48. Dame, R.T., Wyman, C. and Goosen, N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.*, **28**, 3504–3510.
  49. Dame, R.T., Luijsterburg, M.S., Krin, E., Bertin, P.N., Wagner, R. and Wuite, G.J. (2005) DNA bridging: a property shared among H-NS-like proteins. *J. Bacteriol.*, **187**, 1845–1848.
  50. Ceci, P., Cellai, S., Falvo, E., Rivetti, C., Rossi, G.L. and Chiancone, E. (2004) DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus. *Nucleic Acids Res.*, **32**, 5935–5944.
  51. Liu, Y., Chen, H., Kenney, L.J. and Yan, J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.*, **24**, 339–344.
  52. Dill, K.A., Ghosh, K. and Schmit, J.D. (2011) Physical limits of cells and proteomes. *Proc. Natl Acad. Sci. USA*, **108**, 17876–17882.
  53. Wang, F., Redding, S., Finkelstein, I.J., Gorman, J., Reichman, D.R. and Greene, E.C. (2013) The promoter-search mechanism of *Escherichia coli* RNA polymerase is dominated by three-dimensional diffusion. *Nat. Struct. Mol. Biol.*, **20**, 174–181.
  54. Klumpp, S. and Hwa, T. (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl Acad. Sci. USA*, **105**, 20245–20250.
  55. DeHaseth, P., Lohman, T., Burgess, R. and Record, M. Jr (1978) Nonspecific interactions of *Escherichia coli* RNA polymerase with native and denatured DNA: differences in the binding behavior of core and holoenzyme. *Biochemistry*, **17**, 1612–1622.
  56. Friedman, L.J., Mumm, J.P. and Gelles, J. (2013) RNA polymerase approaches its promoter without long-range sliding along DNA. *Proc. Natl Acad. Sci. USA*, **110**, 9740–9745.
  57. Pérez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.*, **28**, 1838–1847.
  58. Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñoz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T. et al. (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.*, **39**, 583–590.
  59. Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muñoz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., Garcia-Sotelo, J.S., López-Fuentes, A. et al. (2011) RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.*, **39**, 98–105.
  60. Record, M.T., deHaseth, P.L. and Lohman, T.M. (1977) Interpretation of monovalent and divalent cation effects on the lac repressor-operator interaction. *Biochemistry*, **16**, 4791–4796.
  61. von Hippel, P.H. (2007) From “simple” DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 79–105.
  62. Dahirel, V., Paillusson, F., Jardat, M., Barbi, M. and Victor, J.M. (2009) Nonspecific DNA-protein interaction: why proteins can diffuse along DNA. *Phys. Rev. Lett.*, **102**, 228101–228101.
  63. Bakk, A. and Metzler, R. (2004) *In vivo* non-specific binding of lambda CI and Cro repressors is significant. *FEBS Lett.*, **563**, 66–68.
  64. Kao-Huang, Y., Revzin, A., Butler, A.P., O’Conner, P., Noble, D.W. and Von Hippel, P.H. (1977) Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *Escherichia coli* lac repressor *in vivo*. *Proc. Natl Acad. Sci. USA*, **74**, 4228–4232.
  65. Dekel, E. and Alon, U. (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature*, **436**, 588–592.
  66. Tănase-Nicola, S. and ten Wolde, P.R. (2008) Regulatory control and the costs and benefits of biochemical noise. *PLoS Comput. Biol.*, **4**, e1000125.
  67. 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.
  68. Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T. and Phillips, R. (2005) Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.*, **15**, 125–135.
  69. Matthews, K. (1992) DNA looping. *Microbiol. Rev.*, **56**, 123–136.
  70. Barkley, M.D. (1981) Salt dependence of the kinetics of the lac repressor-operator interaction: role of nonoperator deoxyribonucleic acid (DNA) in the association reaction. *Biochemistry*, **20**, 3833–3842.
  71. Meiners, J.C. and Quake, S.R. (2000) Femtonewton force spectroscopy of single extended DNA molecules. *Phys. Rev. Lett.*, **84**, 5014–5017.
  72. Morelli, M.J., Ten Wolde, P.R. and Allen, R.J. (2009) DNA looping provides stability and robustness to the bacteriophage  $\lambda$  switch. *Proc. Natl Acad. Sci. USA*, **106**, 8101–8106.
  73. Robertson, R.M., Laib, S. and Smith, D.E. (2006) Diffusion of isolated DNA molecules: dependence on length and topology. *Proc. Natl Acad. Sci. USA*, **103**, 7310–7314.
  74. Riggs, A.D., Bourgeois, S. and Cohn, M. (1970) The lac repressor-operator interaction: III. Kinetic studies. *J. Mol. Biol.*, **53**, 401–417.
  75. Halford, S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem. Soc. Trans.*, **37**, 343–348.
  76. Koslover, E.F., Diaz de la Rosa, M.A. and Spakowitz, A.J. (2011) Theoretical and computational modeling of target-site search kinetics *in vitro* and *in vivo*. *Biophys. J.*, **101**, 856–865.
  77. Bauer, M. and Metzler, R. (2013) *In vivo* facilitated diffusion model. *PLoS One*, **8**, e53956.
  78. Morelli, M.J., Allen, R.J. and Wolde, P.R. (2011) Effects of macromolecular crowding on genetic networks. *Biophys. J.*, **101**, 2882–2891.
  79. Zwanig, R. (1988) Diffusion in a rough potential. *Proc. Natl Acad. Sci. USA*, **85**, 2029–2030.
  80. Berg, O.G. and Blomberg, C. (1978) Association kinetics with coupled diffusion: III. Ionic-strength dependence of the lac repressor-operator association. *Biophys. Chem.*, **8**, 271–280.
  81. Eichhorn, G.L. (1981) The effect of metal ions on the structure and function of nucleic acids. In: Eichhorn, G.L. and Marzilli, L.G. (eds), *Metal Ions in Genetic Information Transfer*. Elsevier, New York, pp. 1–46.
  82. Moncany, M. and Kellenberger, E. (1981) High magnesium content of *Escherichia coli* B. *Experientia*, **37**, 846–847.
  83. Hu, L., Grosberg, A.Y. and Bruinsma, R. (2008) Are DNA transcription factor proteins maxwellian demons? *Biophys. J.*, **95**, 1151–1156.
  84. Sheinman, M., Bénichou, O., Kafri, Y. and Voituriez, R. (2012) Classes of fast and specific search mechanisms for proteins on DNA. *Rep. Prog. Phys.*, **75**, 026601–026601.
  85. Weber, S.C., Spakowitz, A.J. and Theriot, J.A. (2010) Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys. Rev. Lett.*, **104**, 238102–238102.
  86. Golding, I. and Cox, E.C. (2006) Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.*, **96**, 098102–098102.
  87. Weber, S.C., Spakowitz, A.J. and Theriot, J.A. (2012) Nonthermal ATP-dependent fluctuations contribute to the *in vivo* motion of chromosomal loci. *Proc. Natl Acad. Sci. USA*, **109**, 7338–7343.
  88. Wunderlich, Z. and Mirny, L.A. (2008) Spatial effects on the speed and reliability of protein-DNA search. *Nucleic Acids Res.*, **36**, 3570–3578.
  89. Kuhlman, T.E. and Cox, E.C. (2012) Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol. Syst. Biol.*, **8**, 610.
  90. Pulkkinen, O. and Metzler, R. (2013) Distance matters: the impact of gene proximity in bacterial gene regulation. *Phys. Rev. Lett.*, **110**, 198101–198101.
  91. Gowers, D.M. and Halford, S.E. (2003) Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.*, **22**, 1410–1418.
  92. van den Broek, B., Lomholt, M.A., Kalisch, S.M., Metzler, R. and Wuite, G.J. (2008) How DNA coiling enhances target localization by proteins. *Proc. Natl Acad. Sci. USA*, **105**, 15738–15742.

93. Lomholt, M.A., van den Broek, B., Kalisch, S.M., Wuite, G.J. and Metzler, R. (2009) Facilitated diffusion with DNA coiling. *Proc. Natl Acad. Sci. USA*, **106**, 8204–8208.
94. Paulsson, J. (2004) Summing up the noise in gene networks. *Nature*, **427**, 415–418.
95. Thattai, M. and van Oudenaarden, A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA*, **98**, 8614–8619.
96. Shahrezaei, V. and Swain, P.S. (2008) Analytical distributions for stochastic gene expression. *Proc. Natl Acad. Sci. USA*, **105**, 17256–17261.
97. Ochab-Marcinek, A. and Tabaka, M. (2010) Bimodal gene expression in noncooperative regulatory systems. *Proc. Natl Acad. Sci. USA*, **107**, 22096–22101.
98. De Carlo, S., Chen, B., Hoover, T.R., Kondrashkina, E., Nogales, E. and Nixon, B.T. (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev.*, **20**, 1485–1495.
99. Tucker, N.P., Ghosh, T., Bush, M., Zhang, X. and Dixon, R. (2010) Essential roles of three enhancer sites in sigma54-dependent transcription by the nitric oxide sensing regulatory protein NorR. *Nucleic Acids Res.*, **38**, 1182–1194.