Dependence of DNA length on binding affinity between TrpR and trpO of DNA

Nobuo Shimamoto1,7,*, Mikito Toda2,*, Shigetoshi Nara3, Tamiki Komatsu1,5,9, Kiyoto Kamagata6,9, Takashi Kinebuchi1,8,9 & Jun-ichi Tomizawa1,9,10,10

We scrutinize the length dependency of the binding affinity of bacterial repressor TrpR protein to trpO (specific site) on DNA. A footprinting experiment shows that the longer the DNA length, the larger the affinity of TrpR to the specific site on DNA. This effect termed “antenna effect” might be interpreted as follows: longer DNA provides higher probability for TrpR to access to the specific site aided by one-dimensional diffusion along the nonspecific sites of DNA. We show that, however, the antenna effect cannot be explained while detailed balance holds among three kinetic states, that is, free protein/DNA, nonspecific complexes, and specific complex. We propose a working hypothesis that slow degree(s) of freedom in the system switch(es) different potentials of mean force causing transitions among the three states. This results in a deviation from detailed balance on the switching timescale. We then derive a simple reaction diffusion/binding model that describes the antenna effect on TrpR binding to its target operator. Possible scenarios for such slow degree(s) of freedom in TrpR–DNA complex are addressed.

How proteins encounter and bind DNA is one of the most intriguing subjects in biology. An exceptional feature of DNA–protein interactions is that a DNA molecule has multiple sites for a protein binding, yielding various isomeric protein–DNA complexes. They are generally classified into two types: those formed at most DNA sites with similar affinities and those at limited DNA sites with far larger affinities1. The former are called nonspecific complexes and are weakly maintained by sequence-independent interactions such as electrostatic forces between DNA backbone and a protein. The latter are called specific complexes and are stabilized by additional non-covalent interactions between residues of DNA and the protein. The genetic commands are mostly performed by the specific complex. Two binding pathways are possible. In a direct pathway, proteins form the specific complex via direct binding from bulk. On the other hand, proteins bind to non-specific DNA, diffuse along DNA as nonspecific complexes, and the diffused proteins eventually form specific complexes with their specific DNA sites2,3. Nonspecific sites can mediate the binding between a protein in the bulk and a specific site, if they are located near the specific site.

Surby and Reich reported 20-fold enhanced affinity of a monomeric enzyme, EcoRI methyltransferase, when the DNA fragment harboring a single specific site is extended from 14 to 775 bp DNA4. By von Hippel's and Lu's groups, 10–1,000-fold enhancement of affinity was reported for LacI5,6. This huge enhancement was interpreted by the fact that a single tetrameric LacI molecule simultaneously binds to two additional minor operators with a relatively high affinity as well as the major lacO1 operator. However, a 20-fold enhancement remained for DNA fragments that have only lacO1, a significant antenna effect5. Namely, as DNA length increases, the protein affinity

1National Institute of Genetics, Mishima 411-8540, Japan. 2Research Group of Physics, Faculty Division of Natural Sciences, Nara Women’s University, Kita-Uoya-Nishimachi, Nara 630-8506, Japan. 3Department of Electrical and Electronic Engineering, Okayama University, Okayama 700-8530, Japan. 4Research Institute for Electronic Science, Hokkaido University, Kita 20 Nishi 10, Kita-Ku, Sapporo 001-0020, Japan. 5Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21 Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan. 6Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan. 7Present address: Veritas Kitayama, 30-1-104 Minamishiba-Cho, Sakyo-ku, Kyoto 606-0841, Japan. 8Present address: Olympus Corporation, Quality Assurance and Regulatory Affairs, 2951 Ishikawa-machi, Hachioji-shi, Tokyo 192-8507, Japan. 9These authors contributed equally: Tamiki Komatsu, Kiyoto Kamagata, Takashi Kinebuchi and Jun-ichi Tomizawa. 10Jun-ichi Tomizawa is deceased. 11email: nshima@nig.ac.jp; toda@ki-rin.phys.nara-wu.ac.jp
to the target site gets larger, which was named ‘antenna effect’ representing that DNA fragments (antenna) harbor the target site (detector) to receive proteins (signal) and bring them to that site or that the long antenna of insects collecting information to carry it to the central nerve system. Similarly, the binding of *E. coli* RNA polymerase to its strongest promoter, T7A1 promoter, was shown to depend on DNA length. The dependence was once interpreted to accelerate the association rate for specific complex formation by harboring the DNA sequence in the nonspecific complexes. Later, the effect was considered not to be kinetic enhancement, but rather enhancement of the affinity at “equilibrium”, because the protein–DNA system was supposed to be “equilibrated” in the experimental condition. The antenna effect is expected to be an important factor in enabling proteins to form the specific complex at their specific sites more efficiently in the presence of tremendous non-specific sites in a cell.

Despite increased examples of the antenna effect, the theoretical description faces difficulty. Protein affinity to DNA is defined and determined at thermal equilibrium. At this condition, detailed balance should hold, which requires equality between forward and reverse reaction flows in any pathway. If the association flow of a protein to a specific site is enhanced in longer DNA by one-dimensional diffusion along the DNA, the dissociation flow of proteins from the specific DNA site should be also enhanced. Furthermore, the binding affinity is uniquely determined by the free energy difference between the dissociated components and the complex. When detailed balance holds, the protein affinity for the specific site should be independent of DNA length, as shown later. This suggests deviation from detailed balance for those protein–DNA systems showing the antenna effect.

Here, we propose a working hypothesis that slow degrees of freedom in protein/DNA system exist to switch different reaction mechanisms or different potentials of mean force which makes the system deviate from detailed balance in the switching timescale even while each reaction system may attain equilibrium. We postulate multiple potentials of mean force, and show what kinds of topographical features of the multiple potentials give rise to a larger deviation from detailed balance in the timescale of switching. We term this hypothesis ‘chemical ratchet’ hereinafter. A ratchet model was originally presented by Smolchowski and was also discussed by Feynman. The idea of ratchets has been applied to various biological processes where some macroscopic quantities are in non-equilibrium. van Kampen pointed out that the internal degrees of freedom may not reach equilibrium within the timescale of reactions. However, this possibility has not been explored in the study of ratchets. As an example, we consider a chemical ratchet alternating between two different potentials of mean force for a protein–DNA system composed of three kinetic states, free proteins and DNA, nonspecific complexes, and the specific complex. When the alternations occur slower than equilibration timescales on the potentials, the overall system deviates from detailed balance among the three kinetic states.

In the following, we first present the experimental evidence of the antenna effect on *E. coli* repressor TrpR binding to its specific DNA *trpO*. Second, we examine a one-dimensional diffusion model along the nonspecific sites of DNA preceded by protein–DNA binding at the specific site at equilibrium, and show that the antenna effect cannot be predicted while detailed balance holds. Third, we propose the concept of chemical ratchet as a working hypothesis to make the system deviate from detailed balance at some timescale, and revise the reaction diffusion/binding model to represent the antenna effect on TrpR binding to *trpO* by indirectly reflecting the idea of chemical ratchet. Finally, we address possible scenarios for such slow degree(s) of freedom in TrpR–DNA complex. Supplementary Information (SI) contains a detailed discussion of the model and explains the concept of the chemical ratchet. The definitions and the dimensions of the parameters are listed in Table 1 of the SI.

**Results**

**Experiments revealed that TrpR binding to trpO shows antenna effect of 10,000 fold.** We first examined the DNA length dependency (from 18 bp to 5.2 kbp) of the TrpR binding affinity to the specific site *trpO*. All the DNA harbor 18 bp *trpO* at their centers (Fig. 1a). The *trpO* of *trpR* gene was selected among the five operators in *E. coli* because it forms a one-to-one complex with a TrpR. We conducted a hydroxyl radical footprinting for the binding assay to identify to which sequences along DNA TrpR binds at 1 bp spatial resolution. Briefly, a fixed amount of 32P-end labeled DNA fragment was pre-incubated with various excess amounts of TrpR to form TrpR–*trpO* complex for 2 h. Then, hydroxyl radical was added to cleave DNA at various points along the chain. If protein is bound to its specific site, the cleavage at that site is known to be enhanced. The DNA fragments with different lengths were separated by electrophoresis in a sequencing polyacrylamide gel, and the radioactivity of individual fragments was detected by autoradiography. In Fig. 1b, uncut DNA contained trace amounts of DNA fragments as shown in the top half of the autoradiogram and the majority remained at the gel top as an over-exposed dense band (lane U). After the cleavage reaction, DNA was randomly cleaved by hydroxyl radical to generate uniform length distribution as shown in lane 0 where there exist no dense bands at single-cutting condition. This condition enables us to quantify the statistics of cleavage of DNA of the same length. A similar uniform distribution was obtained in the absence of the physiological corepressor, L-tryptophan (lane-trp). In contrast, when L-tryptophan exists, a preferential cut was observed according to increasing concentration of TrpR as indicated by the dense and broad band. The location of the dense and broad band by referencing the G-ladder bands (lane G) corresponds to the specific binding to the *trpO* site. At several positions in the both strands within *trpO*, the cleavage was dependent on the concentration of TrpR but only with L-tryptophan. This is consistent with the fact that TrpR can form the specific complex only in the presence of L-tryptophan. Thus, the enhancement of the band intensity reflects the specific binding of TrpR to *trpO*. The intensities of the specific complex at each [TrpR] were well fitted to the equation expected for one-to-one binding (Fig. 1c):

\[
\text{intensity with no TrpR} - \frac{\alpha \cdot [\text{TrpR}]}{\beta + [\text{TrpR}]}.
\] (1)
is the concentration of total DNA, and its saturation level is the total concentration of [TrpR], respectively. As long as the left-hand side of Eq. (1) quantifies the concentration of TrpR-complex and free trpO in the interval of (triangle). The uncleaved DNA (lane U), cleaved but without TrpR (lane 0), and cleaved in the absence of corepressor L-tryptophan (−trp) were controls. Uncut DNA bands were over-exposed as shown in red in this autoradiogram, but total intensity of a lane were measured independently in a short exposure time. The position of trpO is indicated in the left, and the band most sensitive to the binding, −1T in the case of the upper strand, also shown by the arrow in the right. (c) The band intensities at various concentrations of TrpR. A measured band intensity was first normalized with the total intensity of the same lane for the correction of loading amount, and then the background intensity of no protein was subtracted from it. The intensities thus obtained are plotted against [TrpR] (open circle), and the least square fit to Eq. (1) (solid curve) gives the value of β. (d) The obtained values of β plotted against DNA length showing 10,000-fold variation (see also Fig. 3).

where α and β are parameters independent of [TrpR], indicating the saturation level and the half-saturating [TrpR], respectively. As long as the left-hand side of Eq. (1) quantifies the concentration of TrpR-trpO complex, its saturation level is the total concentration of trpO. Thus, the left-hand side and α can be replaced by [TrpR-trpO complex] and [free trpO] + [TrpR-trpO complex], respectively, converting Eq. (1) into Eq. (2).

\[
\beta = \frac{[\text{free trpO}][\text{TrpR}]}{[\text{TrpR-trpO complex}]}.
\]

Then, the parameter β could be called the dissociation (equilibrium) constant, Kβ, according to the mass-action law. In fact, this term has been used for the parameter corresponding to β\text{NS-trpO} in Ref. 9. However, this term is sometimes associated with implicit theoretical frameworks which must be examined in terms of detailed balance. Thus we reserve this expression until the examination. The obtained values of β are plotted against DNA length in Fig. 1d.

A theoretical model holding detailed balance does not show antenna effect. How can one rationalize the antenna effect of the protein–DNA system? Here we built a simple theoretical model specialized for the protein–DNA system and examined whether the model exhibits the antenna effect of TrpR. The model involves reaction processes composed of specific and nonspecific complexes and free proteins in the bulk as well as diffusion of protein along DNA. The concentration of the specific complex is denoted as nF(t), and its dimension is [M]. On the other hand, we here assume “one-dimensional linear chain of DNA” specified by coordinate x with trpO and “continuum approximation” in treating diffusion of protein along the DNA chain. Then, the concentration of nonspecific complex per a unit length at x and time t is denoted as nNS(x,t), and its dimension is [M length^{-1}]. Thus, nNS(x,t)Δx denotes the concentration of nonspecific complex with protein located in the interval [x, x + Δx]. The concentrations of DNA with empty site at x is n^0_{DNA}(x,t) = n^0_{tot} = n_{NS}(x,t)Δd for x ≠ x₀ and n^0_{DNA}(x_0,t) = n^0_{tot} = n_{NS}(x_0,t)Δd, where n^0_{tot} is the concentration of total DNA, and Δd is the distance between neighboring nonspecific sites, namely 1 bp. The definitions and dimensions of the quantities and parameters are listed in Table 3 in SI. The net flows of three components are generally described using the following equations:

\[
\frac{\partial n_{NS}(x,t)}{\partial t} = D \frac{\partial^2 n_{NS}(x,t)}{\partial x^2} - k^{F-NS}_- n_{NS}(x,t) + k^{F-NS}_+ n^F_{prot} n^F_{DNA}(x,t) - J(t) \delta(x-x_0),
\]

\[
\frac{dn_{NS}(t)}{dt} = -k^{F-S}_- n_S(t) + k^{F-S}_+ n^F_{prot} n^F_{DNA}(x_0,t) + J(t),
\]
Here, \( D \) is the diffusion coefficient of the protein along DNA, and the rate constants have a superscript showing the reaction pathway (F–S, F–NS, or S–NS) and subscript expressing forward reaction (+), or backward reaction (−). Note that the dimensions (as physical units) of rate constants are defined by taking into account the difference of the dimensions between \( n_{NS}(x,t) \) and the other quantities \( n_{S}(t), n_{F}^{DNA}(x,t) \) and so on. Assuming that the DNA length is finite and detailed balance holds between the isomerization of specific and nonspecific complexes at the specific site \( J(t) = k_{+}^{S-NS}n_{NS}(x_{S},t) - k_{-}^{S-NS}n_{S}(t) = 0 \) in Eqs. (3) and (4), the dissociation constant of protein was calculated at some timescale slower than that of protein–DNA binding as follows:

\[
K_{d} = \frac{k_{+}^{F-S}}{k_{-}^{F-S}}. \tag{6}
\]

Since both \( k_{+}^{F-S} \) and \( k_{-}^{F-S} \) are independent of DNA length, \( K_{d} \), and thus the protein affinity for the specific site, is also independent. The analysis done in the discussion of Eq. (S11) in SI shows that detailed balance still holds at equilibrium irrespective of the addition of a diffusion term for finite DNA length. One may interpret that as DNA length increases, the number of forward pathways of TrpR from bulk to the specific site via nonspecific sites increases, enhancing the association to the specific site. Also, that of backward pathways increases, promoting the dissociation from the specific site. Such cancellation results in no dependence of DNA length on the affinity of TrpR. This conclusion is considered to be free from our choice of the one-dimensional diffusion/bindng model.

**Chemical ratchet results in a deviation from detailed balance.** Here we propose a working hypothesis to make the system deviate from detailed balance for some timescale. The key assumption of detailed balance is that the system attains a stationary distribution having time reversal symmetry. However, this assumption may not necessarily hold in all timescales in systems composed of modes with diverse timescales (unless canonical distribution would be prepared a priori as an ensemble). Even after fast degrees of freedom for protein/DNA system attain their equilibria, slower degrees of freedom may not, leading to deviation from detailed balance in the timescale of slow degrees of freedom. Because deviation from detailed balance accompanies chemical changes in the system, we term this hypothesis as a ‘chemical ratchet.’

Here we consider a simple chemical ratchet model switching between two potentials of mean force at every time interval \( t_{int} \). As shown in Fig. 2a, the system has three states 1, 2, and 3 corresponding to dissociated protein/DNA, nonspecific complexes, and specific complex, respectively. The ratchet switches between two different potentials A and B as in Fig. 2b,c. Let us suppose that on each potential the system satisfies the detailed balance
between the three states. The timescales to reach equilibria on potentials A and B are denoted by $t_A$ and $t_B$, respectively. On the potential A shown in blue, state 3 is the most stable, and the flow is directed from state 1 to state 3 through state 2, i.e., from dissociated protein/DNA to specific complex via nonspecific complexes. On the potential B shown in red, the state 1, which is dissociated protein/DNA, becomes most stable. Therefore, the flow is directly from the state 3 to 1, i.e., dissociation from specific complex. This means that when the system switches alternatively between the two potentials for a timescale much slower than the timescales of equilibration on the potentials A and B, $t_A$, $t_B$, nonzero net flow occurs among the three kinetic states 1-3. In turn, when the system switches on a timescale much faster than $t_A$ and/or $t_B$, the system experiences only the averaged potential of the two, resulting in no net flow (Fig. 2d,e at the limit of $t_{int} / \max (t_A, t_B) \to 0$).

The above are two approximate explanations relevant only for extreme cases in terms of the ratio. Then, we would like to ask what will happen when the ratio is in an intermediate value.

Let us formulate this situation in the framework of master equations (see SI Eq. (S23)). The transition matrix $A$ is, for example, defined as follows: $A_{ij} = \exp \left( - \frac{E_{ij}^{(A)} - E_{ij}^{(A)}}{T} \right)$ for $i \neq j$, where $E_{ij}^{(A)}$ and $E_{ij}^{(A)}$ denote the potential of the $i$-th state, and the potential of the saddle separating the $i$-th and the $j$-th states on the potential A, respectively, with the Boltzmann constant set to 1 (the matrix B is also defined likewise for potential B). Suppose that the reactions described by transition matrices A and B alternate every time interval of timescale $t_{int}$. Then, the net flow from the $j$-th to the $i$-th states ($j \neq i$) of the ratchet is given by

$$J_{j \to i} = \int_{0}^{t_{int}} \left( B_{ij}P_{j}(t) - B_{ji}P_{i}(t) \right) dt + \int_{0}^{t_{int}} \left( A_{ij}P_{j}(t) - A_{ji}P_{i}(t) \right) dt,$$

where $P_{i}(t)$ denotes the distribution of the $i$-th state at time $t$. To test our ratchet model, we calculate the time course of the net flow for two different potential landscapes (Fig. 2d,e). The net flow increases almost linearly for the ratio $t_{int} / \max (t_A, t_B)$ being smaller than 1. On the other hand, the increase of the net flow $f$ starts to deviate from linearity and saturates as the ratio exceeds 1.

The idea that the chemical ratchet induces deviation from detailed balance is applicable to any set of potentials. However, the amount of nonzero net flow that is generated depends on the timescale of switching and the topology of those potentials. For example, comparing the potentials B between Fig. 2b,c, we notice that the potential barrier of the saddle from state 2 to state 1 is larger in (b) than that in (c). Thus, more flow is generated from state 2 to state 3 for the ratchet in (b) than in (c), resulting in a larger nonzero net flow as shown in Fig. 2d,e. Moreover, the steeper the initial increase of the net flow is, the larger the saturated net flow is. This initial linear increase can be well reproduced by the time average of the matrices A and B, as shown by green lines in Fig. 2d,e (see SI Eq. (S36)). This implies that the average matrix provides us with a good approximation of the ratchet for $t_{int} \lesssim \max (t_A, t_B)$.

Thus, the master equation is useful for a general description of the chemical ratchet. In the following, we present our model to explain the experimental results utilizing the rate equation rather than the master equation. The reason is that the rate equation is more suitable to take into account the specific mechanism involved in the reaction. The correspondence between the master equation and the rate equation is discussed in the reference\cite{17}. Moreover, for a relevant model of the DNA/protein system, we only need to pay attention to a model corresponding to the matrix which is constant in time, without specifying the alternating matrices A and B (see SI Eq. (S36)).

In order to generate nonzero net flow permanently using the chemical ratchet, some external sources must exist to keep the slow degree(s) of freedom in non-equilibrium. In our experiments, however, we do not have to keep non-equilibrium conditions forever. It only suffices that non-equilibrium conditions are kept long enough so that nonzero net flow is observed in the experiment. As we will point out in Discussion, the timescales of slow degrees of freedom can be much longer than that of the experiment. Thus, we extend the concept of the ratchet to non-equilibrium steady conditions maintained by slow degrees of freedom. Note that the mechanism of our model is the same as the ratchets proposed by e.g.,14,15 in the sense that fluctuation of potentials gives rise to directional flow. Therefore, we call our model a “chemical ratchet” to highlight the point that it is an extension of those ratchets which operate under external energy sources. The number of cycles our ratchet performs during the experiment depends on the ratio of the two timescales: one for the experiment and the other for slow degrees of freedom to cause switching of the potentials.

**One-dimensional diffusion-binding model for protein–DNA system explains antenna effect observed in TrpR-trpO system.** How can one incorporate the concept of chemical ratchet in modeling TrpR–DNA system in which a 104-fold antenna effect was observed in the affinity to trpO operator? Direct incorporation of chemical ratchet may require us to model at least two potentials of mean force among these three kinetic states, free TrpR and DNA, nonspecific complexes, and specific complex, which should be dependent on value(s) of the presumptive slow degree(s) of freedom buried in TrpR–DNA system. Instead, in this paper, we indirectly incorporate the concept of chemical ratchet into the framework of the one-dimensional diffusion/ binding model. As discussed above [see SI Eq. (S11)], when detailed balance holds between the specific complex and the nonspecific complexes near the specific site, it was found that the dissociation constant of protein–DNA system cannot show any antenna effect at the timescale of the experiments. This observation suggests that some deviation from detailed balance is required primarily at the specific site in order to understand the antenna effect. Under stationary condition, the flux $f(x)$ (Eq. 5) becomes constant, denoted by $f_0$, which is represented as $f_0 = (D_{NS}(x)/dx|_{x=0}) - (D_{NS}(x)/dx|_{x=0})$, obtained by integrating the left and right hand sides of Eq. (3) from $x = 0$ to $x = 0+$ with respect to $x$. While we assume the value of $f_0$ is nonzero, it is difficult to postulate the value of $f_0$, and hence we will determine the corresponding value (including its sign) so as to be consistent with the antenna effect observed in TrpR–DNA system. Here, the protein concentration in the bulk is set in
enough excess over the DNA concentration for $n_{\text{NS}}^{\text{prot}} \cong n_{\text{NS}}^{\text{prot tot}}$ to hold. In addition, we introduce new approximations $n_{\text{NS}}^{\text{DNA}}(x,t) \cong n_{\text{NS}}^{\text{DNA tot}} (x \neq S_0)$ and $n_{\text{NS}}^{\text{DNA}}(S_0,t) \cong n_{\text{NS}}^{\text{DNA tot}} - n_{S}(t)$, which are rationalized in the SI after Eqs. (S1–S3), to eliminate some variables as

$$\frac{\partial n_{\text{NS}}(x,t)}{\partial t} = D \frac{\partial^2 n_{\text{NS}}(x,t)}{\partial x^2} - k_{-}^{\text{NS}} n_{\text{NS}}(x,t) + k_{+}^{\text{NS}} n_{\text{NS}}^{\text{prot tot}} - J_0 \delta(x) = 0,$$

(8)

$$\frac{dn_{S}(t)}{dt} = -(k_{-}^{\text{S}} + k_{+}^{\text{S}} n_{\text{NS}}^{\text{prot tot}}) n_{S}(t) + k_{+}^{\text{S}} n_{\text{NS}}^{\text{prot tot}} + J_0 = 0,$$

(9)

where the specific site locates at $x = 0$, and $\delta(x)$ denotes Dirac’s delta function. $J_0$ corresponds to Eq. (5) at $x = 0$, and positive (negative) $J_0$ expresses incoming (outgoing) diffusion flow into (from) the specific site from (to) nonspecific parts. The flow $J_0$ is found to be represented by $J_0 = \frac{2Dn_0}{d_0} \tanh \left( \frac{l}{d_0} \right)$ [see SI Eqs. (S16) and (S18)]. In brief, $J_0$ arises from $D \frac{\partial \delta n_{\text{NS}}}{\partial x}(x) = D \frac{n_0}{d_0}$ where $n_0$ is some constant having the same dimension as $n_{\text{NS}}(x)$. $l$ is the length of nonspecific DNA segment from trpO to an end. $d_0$ denotes the average length of DNA covered with protein diffusion without dissociation from DNA, and equals to Eq. $\sqrt{D/k_{-}^{\text{NS}}}$. The term $\tanh(l/d_0)$ in $J_0$ represents the following: for instance, if TrpR associates with any of nonspecific parts on the DNA and diffuses into the specific site, TrpR may leave from the nonspecific parts before it finds the specific site. That is, for the total length of nonspecific parts $l$, the diffusion flow along the DNA should be “saturated” for $l > d_0$, because any region more distant than $d_0$ gives no contribution to flow. Note that, while non-vanishing $J_0$ makes the system deviate from detailed balance, the value of $J_0$ will be determined so as to be consistent with the observed antenna effect of TrpR–DNA system.

By solving for $n_S$ using Eq. (9) and inserting $n_S$, $n_{\text{NS}}^{\text{prot}}$, and $n_{\text{NS}}^{\text{DNA}}(0)$ corresponding to [TrpR–trpO complex], [TrpR], and [free trpO] into the definition of $\beta$ in Eq. (2) [see the discussion after Eq. (S21) in SI], we obtain Eq. (10),

$$\beta = \frac{n_{\text{NS}}^{\text{prot tot}} n_{\text{NS}}^{\text{DNA}}(0)}{n_S} \equiv \frac{k_{-}^{\text{S}} n_{\text{NS}}^{\text{prot tot}} n_{\text{NS}}^{\text{DNA}} + J_0}{k_{+}^{\text{S}} n_{\text{NS}}^{\text{prot tot}} n_{\text{NS}}^{\text{DNA}} + 2Dn_0 \tanh \left( \frac{l}{d_0} \right)}.$$

(10)

Equation (10) gives an expression of $\beta$ (usually referred at equilibrium) under the existence of nonzero diffusion flow. Recalling the experimental result that $\beta$ monotonically decreases about $10^{4}$-fold as the DNA length increases, this means that $n_0$ is positive (i.e., $J_0 > 0$), since all the other parameters are positive in Eq. (10). Positive $J_0$ implies that $n_S$ involves the incoming flow of TrpR to the specific site along DNA, $J_0(l)/k_{-}^{\text{S}}$, in addition to the term defined at equilibrium, $k_{+}^{\text{S}} n_{\text{NS}}^{\text{prot tot}} n_{\text{NS}}^{\text{DNA}} / k_{-}^{\text{S}}$. For fitting of the experimental result, we rewrite Eq. (10) as

$$\beta = \frac{m_1}{m_2 + \tanh \left( \frac{\text{DNA length} - m_0}{2d_0} \right)}.$$

(11)

Note that we take $(2l + m_0)$ as the DNA length where $m_0$ represents the finite size of specific site implicitly and the factor 2 is due to our experimental design where the length of nonspecific parts are equal for each side from the specific site. We determined $m_0, m_1, m_2$, and $d_0$ to be best fitted to Fig. 3 where log $\beta$ is plotted against log(DNA length). A satisfactory fitting is obtained for four order of magnitude of $\beta$. The constant $\beta$ could be considered as the dissociation constant $K_d$ extended in the non-equilibrium steady state. Note also that, when $J_0$ vanishes,
that thermally induced DNA bending merely triggers its DNA binding, and the DNA in a complex is largely bent.

Complex, which decreases the affinity of the protein at specific site. When the protein tracks the DNA groove dur-

small DNA bending can be thermally induced frequently with timescale only at several microseconds, which

ratchets may be inevitable for understanding reaction processes in biological systems.

Discussion
In this simplified model the deviation is expressed to exist permanently, but it should be noted that the devia-
tion from detailed balance should take place at least for the timescale of TrpR–DNA binding. In biomolecules,
there exist multiple timescales for various degrees of freedom across many different orders to reach equilibrium.

The timescale of DNA–protein binding is usually of the order of milliseconds to seconds, while that of protein-

freedom with longer timescales remain to deviate from detailed balance. Moreover, analyzing the mechanism of

Eq. (10) turns to the definition of dissociation constant $K_d = \frac{k_R}{k_F}$ at equilibrium. The errors were more sensitive
to the values of $m_0$ and $d_0$ (compared to those of $m_1$ and $m_2$) whose best fitted values 18 bp and 625 bp were
consistent with the previous studies. While the solid identification of the actual value of $d_0$ is limited only for
seven experimental data points, the positive $d_0$ reflects that the underlying TrpR and DNA binding interaction
energy at the specific site is greater than those at nonspecific sites, and enables protein of complexes at trpO to
dissociate from DNA, but not to transfer into the neighboring nonspecific parts by one-dimensional diffusion,
introducing asymmetry in association and dissociation.

Materials and methods
Protein and DNA. *E. coli* TrpR protein was provided by Dr. Jannette Carey. All the DNA fragments con-
tained only the intact chromosomal DNA near the trpO site of *trpR* gene, except the 5.2 kbp and 18 bp ones.
The 5.2 kbp fragment included a 2.1 kbp vector sequence of pRPG16 in each end, and the 18 bp one was a hair-
pin duplex with a loop of five cytidine residues to prevent dissociation of the short stem duplex. The fragments
shorter than 500 bp were purified by electrophoresis in an 8% polyacrylamide gel, followed by simple diffusion
from the crushed gel slices. The longer fragments were purified by electrophoresis in a 0.7% agarose gel and
recovered using the Rapid Gel Extraction kit (Life Technologies).
Hydroxyl radical footprinting. All the binding experiments were performed in 10 mM sodium-phosphate buffer (pH 6.5) containing 25 mM NaCl, with or without 0.25 mM L-tryptophan as the coressor. To obtain footprints of TrpR protein on DNA, the 5’-end of the hairpin DNA was labeled with [γ-32P]ATP (6000 Ci/mol, NEN) using T4 polynucleotide kinase (Takara, Kyoto). The unreacted ATP was removed by passage through a SpinColumn G-50 (Pharmacia). For DNA fragments shorter than 500 bp, the 5’-end of the primer was similarly labeled before synthesis of the fragment by PCR. For longer fragments, a primer corresponding to the sequence 100 bp upstream of the trpO site was labeled, and then extended by Sequenase (Stratagene) using the footprinted DNA as a template. The cleavage reaction was performed for 3 min in the presence of 2.5 μM Fe2+ -EDTA mix25, 0.25 mM sodium ascorbate, and 0.015% hydrogen peroxide at room temperature. The concentrations of DNA were 20 nM, 2 nM, 1 nM, 1 nM, and 10 pM for the hairpin DNA, 36 bp, 50 bp, 200 bp, and the longer ones, respectively. To satisfy the single-cutting condition, we stopped the cleavage by an addition of glycerol to 20% when 20% of the full-length DNA fragment had disappeared. The radioactivity was measured by a phosphorimager BAS1500 (Fuji Film, Tokyo) and normalized within individual lanes for correction of loading errors. The band density in the absence of TrpR was subtracted as the background to calculate the enhanced Hydroxyl radical footprinting.

Received: 12 April 2020; Accepted: 17 August 2020
Published online: 24 September 2020

References

1. Ptashne, M. See for example: Designing an efficient DNA-binding protein. A Genetic Switch 2nd Ed. Cell Press and Blackwell Scientific Publications, Cambridge Massachusetts Appendix one (1991).
2. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. J. Biol. Chem. 264, 675–678 (1989).
3. Shimamoto, N. One-dimensional diffusion of proteins along DNA: Its biological and chemical significance revealed by single-molecule measurements. Biol. Chem. 274, 15293–15296 (1999).
4. Surby, M. A. & Reich, N. O. Facilitated Diffusion of the EcoRI DNA Methyltransferase Is Described by a Novel Mechanism. Biochemistry 35, 2209–2217 (1996).
5. Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor–operator interaction: equilibrium measurements. Biochemistry 20, 6948–6960 (1981).
6. Khoury, A. M., Lee, H. J., Lillis, M. & Lu, P. Lac repressor–operator interaction: DNA length dependence. Biochem. Biophys. Acta. 1087, 55–60 (1990).
7. Shimamoto, N. Nanobiology of RNA polymerase: biological consequence of inhomogeneity in reactant. Chem. Rev. 113, 8400–8422 (2013).
8. Ricchetti, M., Metzger, W. & Heumann, H. One-dimensional diffusion of Escherichia coli DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. Proc. Natl. Acad. Sci. USA. 85, 4610–4614 (1988).
9. Metzger, W., Schickor, P., Meier, T., Werel, W. & Heumann, H. Nucleation of RNA chain formation by Escherichia coli DNA-dependent RNA polymerase. J. Mol. Biol. 232, 35–49 (1993).
10. von Smoluchowski, M. Experimentell nachweisbare, der Ublichen Thermodynamik widersprechende Molekularphenomene. Phys. Zeitschrift 13, 1069 (1912).
11. Feynman, R. P. The Feynman Lectures on Physics, Vol. 1. Chapter 46 (Addison-Wesley, New York, 1963).
12. Peskin, C. S., Odell, G. M. & Oster, G. F. Cellular motions and thermal fluctuations: the Brownian ratchet. Biophys. J. 65, 316 (1993).
13. Magnasco, M. O. Forced thermal ratchets. Phys. Rev. Lett. 71, 1477 (1993).
14. Astumian, R. D. & Bier, M. Fluctuation driven ratchets: molecular motors. Phys. Rev. Lett. 72, 1766 (1994).
15. Prost, J., Chauwin, J.-F., Peliti, L. & Aidari, A. Asymmetric pumping of particles. Phys. Rev. Lett. 72, 2652 (1994).
16. van Oudenaarden, A. & Boxer, S. G. Brownian ratchets: molecular separations in lipid bilayers supported on patterned arrays. Science 285, 1046 (1999).
17. van Kampen, N. Stochastic Processes in Physics and Chemistry 3rd edn. (Elsevier, Berlin, 2007).
18. Gunsalus, R. P. & Yanofsky, C. Nucleotide sequence and expression of the Escherichia coli trp operon. Proc. Natl. Acad. Sci. USA 77, 7117–7121 (1980).
19. Kumamoto, A. A., Miller, W. G. & Gunsalus, R. P. Escherichia coli trp operon repressor binds multiple sites within the AROH and TRP operators. Genes Dev. 1, 556–564 (1987).
20. Klig, L. S., Carey, J. & Yanofsky, C. trp repressor interactions with the trp aroH and trp operators: comparison of repressor binding in vitro and repression in vivo. J. Mol. Biol. 202, 769–777 (1988).
21. Yang, J. et al. In vivo and in vitro studies of TrpR–DNA interactions. J. Mol. Biol. 258, 37–52 (1996).
22. Shiicker, P. & Heumann, H. DNA–protein interactions. ed. Kneale, G. G. Humana Press, Totowa, NJ 21–32 (1994).
23. Metzger, W., Hermann, T., Schatz, O., Le Grice, S. F. J. & Heumann, H. Hydroxyl radical footprint analysis of human immunodeciency virus reverse transcriptase–template-primer complexes Proc. Natl. Acad. Sci. USA 90, 5909–5913 (1993).
24. Kamagata, K., Murata, A., Itoh, Y. & Takahashi, S. Finding the target: characterization of the facilitated diffusion of transcriptional factor p53 along dna using ensemble kinetic and single-molecule fluorescence measurements. J. Photochem. Photobiol. C, 30, 36–50 (2017).
25. Yanofsky, C. & Crawford, I. P. The tryptophan operon. In Escherichia coli and Salmonella typhimurium (eds Ingraham, J. L. et al.) 1453–1472 (American Society of Microbiology, Washington, DC, 1987).
26. Ashby, B., Wootton, J. C. & Fincham, J. R. Slow conformational changes of a Neusporas glutamate dehydrogenase studied by the protein fluorescence. Biochem. J. 143, 317–329 (1974).
27. Lin, S. K. & Neet, K. E. Demonstration of a slow conformational change in liver glucokinase by fluorescence spectroscopy. J. Biol. Chem. 265, 9670–9675 (1990).
28. Sisters, J. T., Shuman, S. & Mildvan, A. S. Vaccinia DNA topoisomerase I single-turnover and steady-state kinetic analysis of the DNA strand cleavage and ligation reactions. Biochemistry 33, 327–339 (1994).
29. Prusiner, S. B. Prion diseases and the BSE crisis. Science 278, 245–251 (1997).
30. Stone, V. T., Shuman, S. & Mildvan, A. S. Vaccinia DNA topoisomerase I single-turnover and steady-state kinetic analysis of the DNA strand cleavage and ligation reactions. Biochemistry 33, 327–339 (1994).
31. Kuznetsov, S. V., Sugimura, S., Vivas, P., Brothers, D. M. & Ansari, A. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc. Natl. Acad. Sci. USA 103, 18515–18520 (2006).
32. See the TrpR structures of TrrR, 1TRo, and ICR. Protein Data Bank (http://www.rcsb.org).
33. Orth, P., Schnappinger, D., Hilfen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nat. Struct. Biol.* 7, 215–219 (2000).
34. Kamagata, K., Mano, E., Ouchi, K., Kanbayashi, S. & Johnson, R. C. High free-energy barrier of 1D diffusion along DNA by architectural DNA-binding proteins. *J. Mol. Biol.* 430, 655–667 (2018).
35. Naimushin, A. N., Fujimoto, B. S. & Schurr, J. M. Dynamic bending rigidity of a 200-bp DNA in 4 mM ionic strength: a transient polarization grating study. *Biophys. J.* 78, 1498–1518 (2000).
36. Cuculis, L., Abi, Z., Zhao, H. & Schroeder, C. M. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* 6, 7277 (2015).
37. Murata, A. *et al.* One-dimensional search dynamics of tumor suppressor p53 regulated by a disordered C-terminal domain. *Biophys. J.* 112, 2301–2314 (2017).
38. Nelson, S. R., Dunn, A. R., Kathe, S. D., Warshaw, D. M. & Wallace, S. S. Two glycosylase families diffusively scan DNA using a wedge residue to probe for and identify oxidatively damaged bases. *Proc. Natl. Acad. Sci. USA* 111, E2091-2099 (2014).

Acknowledgements
This work was supported by JSPS (Nos. 25287105, 25650044 to TK, 22654047, 19K03653 to MT), and the Research Program of ‘Five’ star Alliance in ‘NJRC Matter and Dev.’ (TK, KK, MT, SN). Institute for Chemical Reaction Design and Discovery (ICReDD) was established by WPI, MEXT, Japan. Also, one of the author (SN) appreciates MEXT-JSPS KAKENHI Grant Number JP24120707. The authors thank the anonymous referees for valuable comments. They have led us to correct mistakes in the original manuscript and to improve our explanation for the results. The authors also thank Dr. James Nicholas Taylor at Hokkaido University for checking and suggesting English expressions.

Author contributions
T. Kinebuchi performed all experiments with his advisor N.S. M.T. built a chemical ratchet model based on master equations and conducted the simulation. S.N. obtained analytical solution for diffusion-binding equation. N.S. conducted the fitting of the experimental data. N.S., M.T., S.N., T. Kinebuchi, T. Komatsuzaki and K.K. organized this study. J.T. pointed out the importance of deviation from detailed balance for understanding the antenna effect and critically advised N.S. and T. Kinebuchi.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-71598-3.

Correspondence and requests for materials should be addressed to N.S. or M.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020