

Review

RNA Polymerase and Transcription Mechanism: Forefront of Physicochemical Study as chemical reactions

Nobuo Shimamoto^{1, 2,*}, and Masahiko Imashimizu³

¹ National Institute of Genetics Mishima, Shizuoka-ken Japan ; nshima@nig.ac.jp

² Present Address ; Veritas Kitayama, 30-1-104 Shimogamo-Minamishiba-cho, Sakyo-ku, Kyoto, Japan 606-0841

³ Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology, Tokyo, 135-0064, Japan

* Correspondence: nshima@nig.ac.jp; Tel.: +81-90-2139-1019

Abstract: Transcriptional regulations have been widely studied as one of the main bridges between biology and other basic sciences as well as medicine. The traffic across it has been mostly unidirectional: chemistry and physics provided a lot of tools for biology, although the supply is now saturating. The traffic in opposite direction, the supply of subjects to develop chemistry and physics, has been only a little. However, if there are any, the supply will be at least from transcription, because the notion of chemical reaction is the strongest. This topic is aimed to increase the opposite traffic by introducing the forefront of physicochemical studies of transcription.

(105 words)

Keywords: transcriptional regulation; reaction theory; prediction of promoters; one-dimensional diffusion; rate equation; detailed balance; antenna effect; physicochemical techniques

1. Introduction

The intellectual horizon of biological studies has been extended over the early distinction as replication, transcription, and translation for several decades. In a historical view, transcription has been developed in a unique feature that its development is synchronized with that of physicochemical and biophysical techniques. This is because a fully-active reconstitution system with purified components was established earliest among the three. In consequence, it tended to show a depth of chemistry and physics. The recent advance and degeneracy of the techniques are summarized by [Buckle et al.](#) in this issue.

Studies of transcription were blessed with a historical fortune [1]. It started from the construction of cell-free system as other biological research areas, but the early finding of the initiation factor in *E. coli*, sigma-70 subunit, simplified the biochemical analysis of transcription afterwards. The single protein can switch on the synthesis of most native transcripts, contributing to the reconstitution with purified factors and the identification of promoters.

The full-active reconstitution system was expected for interpreting physiological phenomena deductively in terms of the principles of chemistry and physics, unifying basic principles of biology, chemistry and physics. Contrary to this expectation, the mechanistic study has been inductive rather than deductive. The DNA-dependent RNA polymerase (RNAP) is a distinct enzyme with its cofactor, the template DNA. It encodes genetic information, and thus the mechanistic biochemical studies were combined with genetics to involve the inductive study from genetic information.

This tendency is typically shown in the prediction of transcriptional promoters, which is one of [the article by Imashimizu et al.](#) in this special issues. The consensus sequence for promoters for *E. coli* major sigma factor is known as consensus promoter sequence, which is familiar in most textbook of molecular biology [2, 3]. These sequences were empirically induced as the most frequently conserved

ones among the identified promoters. The actual agreements with the motifs is not enough, and there is only one promoter preserving these motifs among several thousands in *E. coli* K-12 strain. Moreover, the conservation shows no correlation with promoter strength. One of the strongest promoters, A1 promoter of bacteriophage T7, has a sequence far from the motifs. This lack of correlation may rather reflect a likely hypothesis that *E. coli* has not evolved so as to strengthen all of its promoters. The deductive prediction of promoters and their strengths is still a challenge.

Genetics has been powerful tool in mechanistic studies, but reach saturation in the field of transcriptional regulation. There are movement beyond the conventional limit of genetics, as Chatterji and his colleagues are challenging in this issue: the polymorphic feature of a single gene product.

What DNA has introduced in the study of transcription is not limited to its sequence information. As a polymer with repetitive structure, some transcription factors, RNAP and transcription complexes can translocate along DNA. It happens to RNAP during promoter search, to elongation complex in Brownian ratchet mechanism, to initiation as well as elongation complexes in backtracking. Not only the effect on limited steps, this diffusion may be a distinct mechanism of transcriptional regulation in combination with a new notions, timescale and chemical ratchet which are introduced later. However, theoretical analysis for one-dimensional diffusion remains primitive and is not based on the real polymeric structure of DNA as discussed later.

The phenomena in transcription are composed of chemical reactions as other biological processes, that are not necessarily described with rate equations that have a definite limit of application [4]. The ignorance of the limit caused confusion in the treatment of diffusion and Brownian motions. We would like to emphasize here that consideration on the timescales of reactions is critical for further sound development of this field.

2. Implicit assumption of rate equations and a danger of applying them to proteins and DNA

In chemistry, the rate equation is the most representative analytical method which have been developed for several centuries. A reaction rate or reaction velocity is thus described as

$$k[\text{reactant}] \quad (1)$$

$$\text{or } k[\text{reactant}_1][\text{reactant}_2] \quad (2),$$

for a unimolecular reaction or a bimolecular reaction, respectively, by using the time-independent constant k . However, these equations are not universally true and they implicitly based on several assumptions. The most essential one is the homogeneity of the reactant in terms of the reaction. In Eqs. 1 and 2, we are supposed to assume that all the reactant molecules share the same probability of being converted into the product. If the reactant is inhomogeneous, unreacted reactant would be enriched by less active fractions in time, contradicting the time-independence of the rate constant.

There is a reason for the reactant to be homogeneous under a condition which is usually satisfied. Let me consider the case of the bimolecular binding reaction $A+B \rightarrow C$, for example. Among the A reactant molecules, different molecules are differently distant from their nearest B molecules with different velocities, namely different initial conditions or different molecular histories. The possibility of the reaction of an A molecule is in principle depends on its history. Obviously, the shorter the distance, the larger the possibility. Their locations and velocities are randomized due to the numerous collisions with solvent molecules as well as those with other A molecules and even with B molecules in non-productive collisions. Thus, the ensemble average of the possibility is expected to become less dependent on the histories, and the average approaches a value common to all A molecules in time. Even when a fraction of the reactants have large velocities, namely a higher local temperature, all the fraction of the reactants will have a common temperature in time due to the collisions before the reaction in consideration. This time-consuming convergence is conventionally termed "thermally equilibrated". Instead of this word, we hereafter use the term "molecular shuffling" because the term "thermal equilibrium" is also used between the two states in a chemical reaction to avoid the ambiguity of "temperature".

For enough molecular shuffling to occur, molecules have to be confined to the reactant state, of course. In the condition where molecular shuffling is completed much earlier than the binding

reaction, the reaction possibility or the rate constant would be thus uniquely determined. In other words, the rate constant can be defined only when the “shuffling” is enough to homogenize reactant molecules (see more strict statistical discussion in [4])

3. Chemical reactions beyond rate equations

The shuffling, however, costs time and thus is not always enough. The most famous extrapolation beyond the essential assumption of homogeneity is the rate constant for a diffusion-controlled bimolecular reaction, where every collision between reactant molecules significantly results in the complex formation. The rate constant is given by Debye–Smoluchowski equation.

$$k_{diff} = \frac{4\pi r}{1000} (D_{protein} + D_{DNA}) N_A \quad (3).$$

In this form, r is the encounter distance between protein and DNA, and $D_{protein}$ and D_{DNA} are their diffusion coefficients. N_A denotes Avogadro’s number, and the number 1000 must be added when the concentration and length are expressed in molar and in cm, respectively.

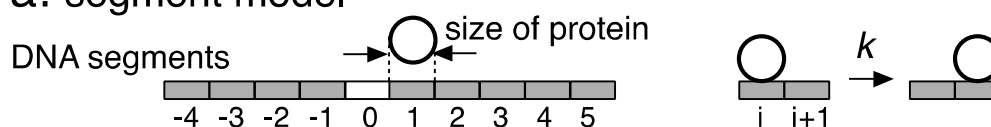
Since the elementary process of the encounter is diffusion of the reactant molecules. In the case of diffusion-controlled binding, molecular shuffling is largely limited, and the closer the distance between A and B, the higher the probability all the time. Namely, there are no homogeneity of reactants in the diffusion-controlled binding. Therefore, k_{diff} is calculated on theoretically inconsistent framework [5].

In contrast, if the bimolecular reaction is not diffusion-controlled, namely most collision events between A and B result in their dissociation and only a small fraction of the collision is productive, the shuffling is mediated by the non-productive collision to rationalize Eq. 2. In fact, the inconsistent rate constant of diffusion-controlled binding is a kind of harmless over-extrapolation, because it approximately shows the upper limit of association rate constant.

4. Difficulties in segment models of one-dimensional diffusion

For the analysis of one-dimensional diffusion of protein along DNA, the segment model of DNA is frequently introduced as an approximation. In this model, a linear DNA is divided into the segments and one-dimensional diffusion is described as the transfer between the segments with a rate constant (Figure 1a). The size of the segment is usually set to the protein size without any overlaps. There would be no time for shuffling before the transfer, because the transfer itself is an elementary process of the molecular shuffling, and because there are no confinement to the end of the segment.

a: segment model



b: real distribution of protein binding site

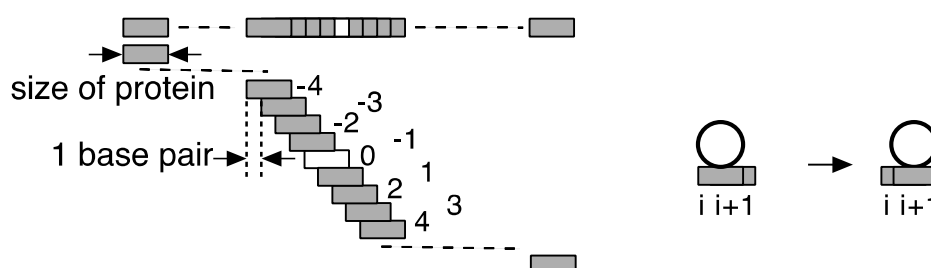


Figure 1 Segment model for one-dimensional diffusion and its difficulty. DNA is considered to be a sequence of binding sites, and the specific site is illustrated as open box and a nonspecific one as gray box. (a) In segment model, the binding sites are not overlapped with each other and the site size is assumed to be the size of the protein (open circle), typically 20 bp for example. The diffusion is

replaced by a transfer reaction into the next box with a rate constant. (b) More real definition and distribution of protein binding sites on DNA. A protein binding site exists at every base pair to identified by the DNA base sequence as long as the protein size. This real distribution reflects the repetitive structure of nucleotide and different from the site size. Therefore, DNA is a sequence of overlapping boxes and one dimensional diffusion is a movement of the gravity center of the protein along DNA, that cannot be described with rate constant because of the absence of shuffling.

Moreover, there is a more serious problem. In this model, the site size is confused with the distance between contiguous binding site. The distance should be one base pair irrespective of the size of the protein because a protein binding site is defined at every structural repeat of the polymer, namely one base pair (Figure 1b). Therefore, the segment model is under-estimates the number of protein binding sites. Since the site size is usually equal to the size of the protein on DNA, typically 15~40 base pair, unless the protein forms a ring or helical polymer with DNA at its axis like LecA [6]. Therefore, the segment model under-estimates the density of protein binding site by ten-fold or more.

Moreover, the model ignores competitive nature of the binding within the protein size. If the site size is 20 bp, a binding at a site should block the additional binding at the 19 neighboring sites. This characteristic mode of competition and the high density of the nonspecific sites are nature of double-stranded DNA, but cannot be approximated by the segment model that mistakenly suppose the molecular shuffling within the segment. If there are potential barriers along non-specific DNA, they should exist at every base pair not at the ends of the virtual segments. Therefore, one-dimensional diffusion cannot be described by a rate equation between segments of finite size, but can be analyzed so far by diffusion equation on a continuum.

In single-molecule analysis, protein concentrations are sometimes increased to the level where several or more protein molecules are complexed on a single DNA chain to get enough sample numbers for statistical averaging. Because the segment model ignores the characteristic mode of competition in protein binding to nonspecific sites, and because it under-estimate the number of possible nonspecific binding sites order of magnitude, it spoils the reliability of the obtained quantitative results. At present, reliable analysis can only made at low protein concentration making one or less protein molecule on the DNA by using composite equation of rate equation and diffusion equation. The solution according to the overlapping binding site shown in Figure 1b has been a big challenge for theorists in biology.

5. Timescale: Index showing how fast a reaction becomes stationary

The timescale is defined as a time required for a converging reaction to be stationary. In the case of exponential decay of $e^{-t/\tau}$, it is the time constant τ . Otherwise, it is the time required for the deviation from the stationary value to reduce by half or e^{-1} . In kinetic analysis, the reactions with much faster (or short) timescales are considered to be equilibrated, and those with much slower (or longer) timescales are assumed to be stopped. The reactions with similar timescales are called coupled and must be treated as a dynamic phenomenon. This coarse graining of time is the core heart of temporal analysis.

Notably, the timescale defined as above is specified in neither of forward or backward reaction but in both. If the reaction is $A \xrightleftharpoons[k^-]{k^+} B$ or $A+B \xrightleftharpoons[k^-]{k^+} C$, the timescale is $\tau = (k^+ + k^-)^{-1}$ or $\tau = (k^+[\overline{A}] + k^-)^{-1}$, respectively, where \overline{A} is the temporal average of A that exists in excess over the other reactant B.

Now we can rephrase the necessary condition for the use of rate equation as that the timescale of the molecular shuffling is much shorter than that of the reaction in consideration. It is usually satisfied for a chemical reaction of small molecules. However, for macromolecules, a change in their conformational changes could be slower than the reaction in consideration: the timescales of which are minutes, hours, and even years [7-9], making the timescale coupled or slower than the chemical reaction in consideration.

5. Transcriptional regulation by timescale matching and mismatching

When there are two or more pathways in sequence, namely in a stepwise process, the timescale of the whole process is their sum. A transcription process contains many stepwise reactions and its regulation is critically determined by their timescale. Let us consider one of the most typical transcriptional regulations, the competitive binding between a repressor and RNAP at a promoter overlapping with an operator. It may be widely believed that the transcriptional inhibition is determined only by two, the concentration of the repressor and its affinity for the operator. This is true only when the timescale of the repressor matches to the timescale of production of mature transcript as follows.

Suppose that there are a repressor (I), and RNAP (R), and the promoter DNA, and both I and R exist in excess over the promoter in a cell. The binding reactions of I and R to the DNA are assumed to be more rapid than the following formation of the first phosphodiester bond formation, which may be a likely assumption in a cell. The promoter is then fractionated into free promoter, inactive repressor complex, and active open complex by the ratio of $1 : [I]/K_I : [R]/K_R$, respectively, where K_I is the dissociation constant of the repressor-operator complex, and K_R is that of RNAP-promoter complex. During initiation, the total amount of the promoter DNA available in this equilibrium is kept reduced by the formation of ternary complex retaining nascent transcripts. Each of the first four phosphodiester bonds is formed in 30-100 milliseconds on the strongest T7A1 promoter in initiation [10]. The inhibitor binding causes deceleration of the formation of the ternary complex by reducing fraction of $[R]/K_R$, by typically less than a second. When the concentration of I is close to or more than K_I , the deceleration of the formation is significant in the production of ternary complex (Figure 2a; unless $R \gg K_R$). However, when the timescale of the ternary complex delayed by the repressor binding is still faster than the timescale of the production of the mature transcript, typically in the order of minute, the repressor hardly inhibits the transcription due to the mismatching of the timescales (Figure 2b).

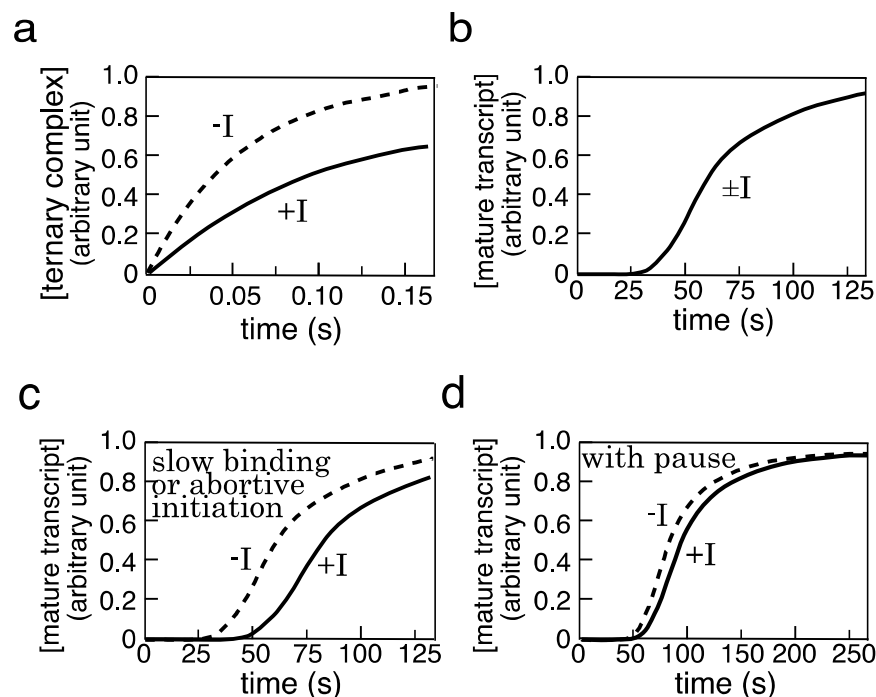


Figure 2 Competitive repressor and transcription inhibition The time courses of the formation of ternary complex (a) and the production of the mature transcript (b-d) are shown, when a repressor (I) and RNAP (R) competitively bind to a promoter-operator overlapping domain of DNA. These time courses are generally dependent on the many experimental parameters as well as RNA degradation, and thus presented as example images. The scale of coordinate is also symbolic to indicate the differences between the timescales. (a) The addition of a repressor can significantly delay the formation of ternary complex. The timescales is decreased in half in this example. (b) When the timescale of the formation mismatch that of the production of the mature product, the repressor

cannot inhibit transcription at all. Two curves with or without the repressor almost overlap with each other. (c) When slow binding or abortive initiation makes the timescale of the formation slower (longer) then the inhibition is recovered. (d) If a pause is introduced, the mismatch is again recovered and the inhibition is reduced. The RNA processing can exert a similar effect.

To realize the inhibition by the repressor, its binding reaction can be made much slower. One realistic way is to introduce a time-consuming covalent bond formation/cleavage in the binding as some regulators like AraC. The phosphorylation of the two-component system may work similarly in addition to its main function of signal transduction. Alternatively, the timescales can match by decelerating the formation of active ternary complex. Abortive transcription can have this function by forming inactive ternary complex, and this may be one of its biological significances. With the timescale matching, repressor introduces a longer delay of the production of mature transcript as the longer time-lag shown in Figure 2c.

Furthermore, if there is an time-consuming elongation pause or RNA processing, the timescale matching can be again hampered by the further deceleration of the production of mature transcript (Figure 2d). Namely, by introducing a longer time-lag in the production, the sensitivity to the repressor is reduced. Since these changes by timescale matching and mismatching can work at least as fine tuning of gene expression, it will be interesting to classify the regulatory mechanisms and to find a new cross-talk in transcription from this viewpoint.

In a cell, transcripts are not necessarily stable and some are degraded rapidly. The coexistence of the synthesis and degradation is named futile cycle or more positively push-pull mechanism. Although there is 2000-fold difference in the timescale between the examples shown in Panels a and d, the RNA degradation alters the timescale of RNA accumulation faster (shorter), making the timescale matching easier than symbolically illustrated in Figure 2. This can be another function of RNA degradation.

6. Chemical ratchet

For a long time, the first step in a kinetic analysis was supposed to define the homogeneous reactants. However, at first, one must examine whether or not the reaction contain chemical ratchet, which has been proposed recently for protein-DNA binding [11]. The new notion is the extension of the ratchet mechanism in physics, which is driven an external energy source [12] or internal ones [13, 14] similarly in this case.

Figure 3a shows the kinetic scheme of the simplest unimolecular reaction, inter-conversion between A and B (Figure 3a, "Mechanism a" hereafter). A and B are two states of the reaction. It is actually not limited to unimolecular one. If the reaction is a binding, A is composed of two dissociated reactants and B is their complex. The thermodynamic rule of detailed balance holds at equilibrium.

When B exists in two conformations B_1 and B_2 with different reactivities, the kinetic scheme becomes Mechanism b (Figure 3b) with the conformational change of k_{\pm}^c step. The conformation can change in the direct pathway of k_c^{\pm} and/or in the stepwise $k_1^{\pm} + k_2^{\pm}$ pathway via A. The forward reaction of k_1^{\pm} stochastically coexists with that of k_2^{\pm} all the time. At equilibrium, detailed balance holds in this case, too.

A chemical ratchet is shown in Panel c, and it is composed of alternative reaction pathways temporally switching. This kinetic scheme for chemical ratchet looks similar to Mechanism b in appearance. The critical difference exists in molecular or microscopic level. The two pathways, k_1^{\pm} and k_2^{\pm} are alternative in chemical ratchet, while they coexist in Mechanism b. In chemical ratchet, the pathway is k_1^{\pm} for t_1 and then switches to k_2^{\pm} for t_2 and again switches back to k_1^{\pm} for t_3 , and so on. In the pathway k_1^{\pm} or k_2^{\pm} , B exists in the form of either B_1 or B_2 , respectively, and they never coexist. In other words, in Mechanism b, the conversion between B_1 and B_2 is possible all the time, in the stepwise pathway via A and/or direct pathway, while these conversion pathways do not exist in chemical ratchet. In this mechanism, the interconversion is only possible at the time of the switching, and thus the potential mean force switches alternatively as shown in Figure 3c.

The timescale of the switching, the average of $(t_{odd} + t_{even})$ is coupled with or slower (larger) than the timescales of the reaction, $(k_1^+ + k_1^-)^{-1}$ and $(k_2^+ + k_2^-)^{-1}$, otherwise Mechanisms b and c

become identical. This means that the internal degree(s) of freedom corresponding to the switch is the slowest or close to the slowest among all degrees of freedom. Because detailed balance holds at the equilibrium of the reaction with the slowest timescale [4], and because the timescales of k_1^\pm and k_2^\pm reaction are not the slowest, detailed balance on the reaction k_1^\pm or k_2^\pm is not guaranteed. The deviation from detailed balance by switching the potential mean has already been reported [13, 14].

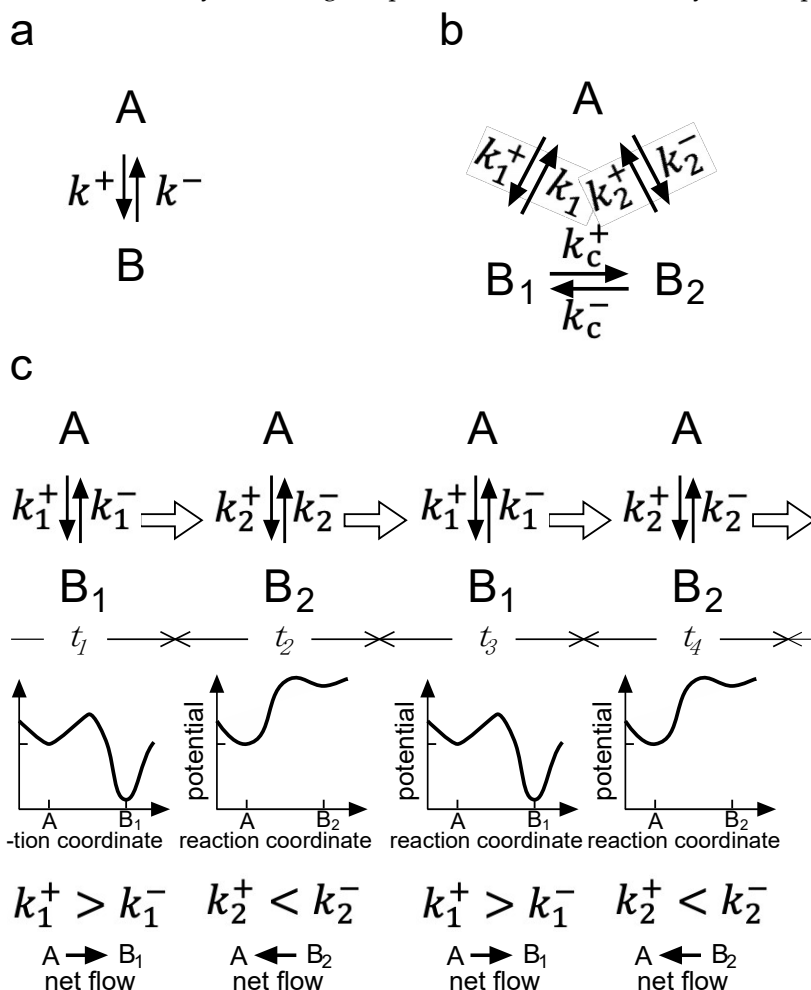


Figure 3 Chemical ratchet (a) The simplest chemical scheme composed of two states A and B, describable with rate equations with the rate constants (k^\pm). (b) The chemical scheme when B state is divided because of inhomogeneity in the reactant with two different activities. (c) Chemical ratchet with alternating two potential mean forces which are shown below the time duration t_i of the available potential. The switching is indicated with open arrows, and its timescale must not be faster (shorter) than that of the conversion between A and B_i . Since the potential AB_1 is inclined to B, while the potential AB_2 to A, there is net flows after a potential switching, generating an oscillating flow.

Moreover, in chemical ratchet, the reaction k_1^\pm or k_2^\pm is not equilibrated after the switching but oscillates between the two imaginary equilibria of the reactions. Since $k_1^+ > k_1^-$ in the example shown in Panel c, there is a net flow from A to B_1 to attain the equilibrium between A and B_1 . Since $k_2^+ < k_2^-$ as shown in the potential k_2^\pm there is an opposite net flow from B_2 to A to attain the equilibrium between B_2 and A. Therefore, in microscopic level, periods at non-equilibrium exist in chemical ratchet.

The microscopic non-equilibrium is an oscillation, and the time average of many cycles of oscillation converges to a constant as the averaging period becomes longer. Similarly, because the phases of many microscopic oscillations are random, ensemble average also converges to a constant as the size of ensemble becomes larger. Therefore, in chemical ratchet, microscopic non-equilibrium states exist in a macroscopic stationary state. Although this stationary state is the time-independent

state to which the system converge, the term “equilibrium” is here reserved to describe it, because it contains non-equilibrium.

7. A possible molecular example and detection of chemical ratchet

A possible example of molecular model for B_1 and B_2 may help understand the reality of chemical ratchet. DNA B-helix has the rigidity with the persistent length of ca. 150 bp. The B-helix in the protein complex is made even more rigid because of the specific interactions maintaining the specific complex. This enhanced rigidity of DNA can consist a chemical ratchet (Figure 4a).

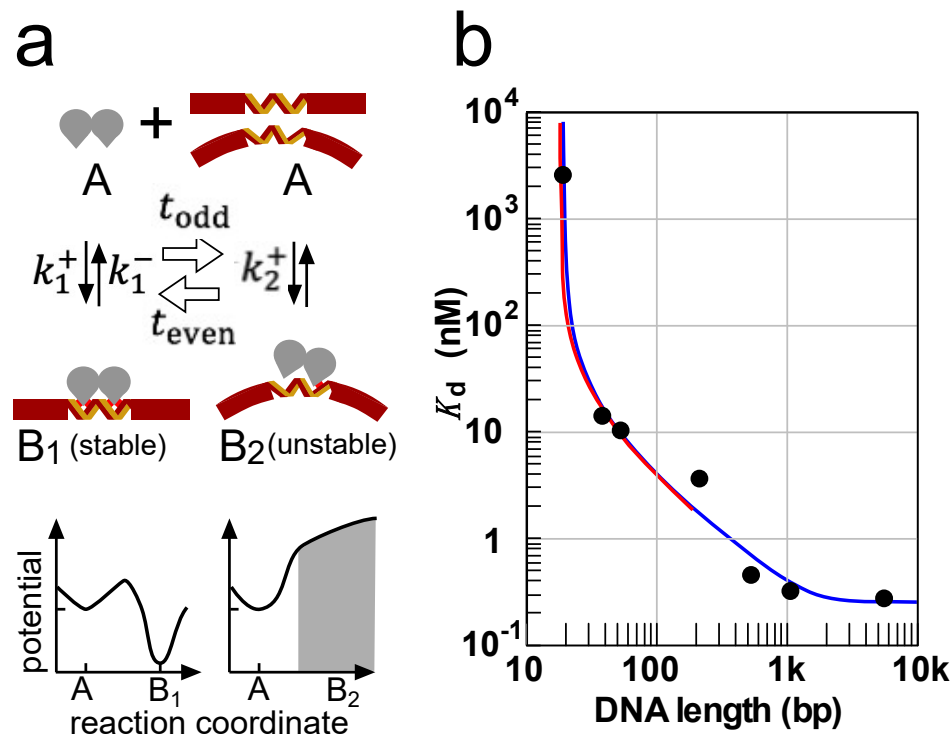


Figure 4 One of possible molecular models for chemical ratchet in protein-DNA binding (a). Reaction scheme of the chemical ratchet and its alternative potentials 1 and 2. A DNA binding protein (gray droplets) binds to its specific site (yellow and brown strands) on DNA (brown bar). The protein stably binds to its specific site on straight DNA with many cooperative interactions (four small red boxes in B₁) in reaction k_1^+ , and its potential mean force has two local minima. In the presence of one-dimensional diffusion, the both steps k_1^+ and k_1^- are accelerated for longer DNA. In the alternative reaction, $A \rightleftharpoons B_2$, B₂ has only a little interaction (a small red box) because of a bent at the specific site, and distributes on the high-energy slope (gray) in the potential mean force. Because, B₂ is unstable and because the DNA bent inhibits sliding out of the specific site, it is destined to be dissociated without molecular shuffling (impossible to define k_2^-). Thus, it cannot be converted into B₁, or B₁ and B₂ do not coexist. The forward switch rarely occurs, or t_{odd} is larger than the timescale of the binding, although t_{even} may be small. (b) The antenna effect of TrpR binding to *trpO*. The values of dissociation constant (closed circles) were obtained with the site-specific hydroxyl radical footprinting of TrpR on *trpO* DNA of various lengths. The differential equations composed of rate equations and diffusion equations corresponding to the one-dimensional diffusion of TrpR along the DNAs was solved in stationary state to give the best-fit curve (blue line) with the site size of 18 bp and the diffusion distance of 625 bp (blue) [11]. Another theoretical calculation of chemical ratchet under the assumption of rapid diffusion for short DNAs gave the best-fit curve with the site size of 18 bp (red curve; see the text). This panel is taken from [16].

The reaction k_1^+ with Potential 1 is supposed to be a protein binding to straight specific site mediated by one-dimensional diffusion with tracking the DNA glove, sliding. The longer the DNA, the larger the rate constants k_1^+ and k_1^- . Thus, sliding out from the specific site is the major

dissociation pathway. The potential mean force switches to Potential 2, when the rigid specific site in stable complex B₁ is bent by high-energy thermal fluctuations to form an unstable complex B₂. The energy-requiring bending happens infrequently so that, the average of t_{even} as well as the time scale of the switching, the average of $(t_{odd} + t_{even})$, can be longer than that of the binding reaction. the complex B₂ distributes according to the bent angle on the high-energy slope of the potential mean force as grizzled in Figure 4a. Because the groove structure in B₂ is distorted by the bent, its dissociation by sliding out is inhibited. Therefore, B₂ is destined to dissociate directly into A independently of DNA length. In this way, the potential mean forces 1 and 2 are alternative. Notably, because B₂ is not at a local minimum and confined by no potential barriers, there is little molecular shuffling before the dissociation, indicating that rate constant k_2^- cannot be defined. Irrespective of this difference from what is shown in Figure 3c, this example satisfies the requirement for chemical ratchet.

In Potential 1, both of k_1^+ and k_1^- reactions are increased for longer DNA, while their ratio, the affinity, is independent of DNA length if there is no switching, which is also deduced from detailed balance. If there are switching, the major association is through potential 1 which is length dependent, while the dissociation is significantly proceed through potential 2 which is length independent. In consequence, the affinity of the protein for the specific site is increased by DNA length.

Theoretical framework of chemical ratchet was proposed by M. Toda on the experimental results and interpretation by T. Kinebuchi and N. Shimamoto with the mathematical support by S. Nara [11]. In the experiment, *E. coli* TrpR binding to *trpO*, the apparent dissociation equilibrium changed by 10,000-fold according to the length of DNA harboring *trpO*, which has been named antenna effect [15, 5]. The scenario shown above is one of the possible candidates of molecular models for the chemical ratchet. As shown in Figure 4b, the observed values of the dissociation equilibrium constant, the inverse of the affinity, are well fitted by the solution of differential equation composed of diffusion equation and rate equation (blue curve of Figure 4b)[11]. For short DNAs, where sliding is supposed to be equilibrated before dissociation from DNA, the concentration of the complex can be dynamically determined from kinetic equations and the apparent dissociation constant was given its temporal average over the period longer than the timescale of the switching (red curve of Figure 4b). The fitting are satisfactory with the consensus site size of 18 bp for *trpO*[16].

At present, there is another mechanism exerting antenna effect, looping mechanism. If a protein molecule has two binding sites for DNA, the complex formed in the first binding can be further stabilized by the second binding. The longer the DNA, the more possibilities for the second binding with the same DNA molecule by forming a DNA loop. This has been proved for several proteins with two DNA binding sites [17-21]. However, there are no proteins with a single DNA binding site like TrpR. Moreover, we denied this possibility by using a DNA connection that retains DNA looping but blocks one-dimensional diffusion [16]. Moreover, this study also provided evidence for the existence of one-dimensional diffusion *in vivo*. Therefore, all these lines of evidence are so far consistent to the chemical ratchet of TrpR binding to *trpO*.

8. Biological significance and indications of chemical ratchet

There was a serious contradiction between the values of the affinities of TrpR for *trpO* measured in vitro [22] and estimated from the TrpR protein level and the number of *trpO* sites in vivo [23]. The difference is about 100-fold. However, this can be solved by the antenna effect of TrpR due to the possible chemical ratchet in the presence of one-dimensional diffusion [16].

A new type of cross-talk can be predicted from the antenna effect caused by the chemical ratchet in the presence of one-dimensional diffusion. When a regulatory protein exerts the type of antenna effect, a binding of a foreign protein near its binding site will decrease the affinity of the regulatory protein by hindering one-dimensional diffusion. This cross-talk at a distance can be universal. In fact, the binding of a repressor near a promoter is expected to block the diffusion into the promoter by RHAP polymerase even when there is no direct steric hindrance between their bindings.

How can we decide whether the mechanism is a chemical ratchet or not? The most direct is to detect the imbalance of microscopic flow in macroscopically stationary state in a single-molecule

experiment. But it is usually too tedious experiment just for the selection of a reaction scheme. Furthermore, a quantitative analysis of single-molecule experiments is always exposed to the danger of artifact caused by the surface effects and fixing methods, and thus good control experiment is essential.

The biological significance of the microscopic flow other than antenna effect is not necessarily clear at present. The contribution of chemical ratchet could be pointed out a long time ago, because of the universal property of chemical ratchet. For example, in forty years ago, the experimental model of slow dimerization of yeast enolase with slow conformational change of monomer [24] was attacked as a violation against thermodynamics [25,26], but it can be a chemical ratchet with switching conformations.

It is a chance now to reconsider transcriptional regulations on the correct framework of reaction theory of DNA-binding proteins, to find a new mechanism and a new cross-talks. The maturation of transcription study allows a challenging feedback to chemistry with its depth of both physics and biology.

References

1. Chamberlin, M. , RNA polymerase-An overview In *RNA polymerase book*; Losik, R. chamberlin M., Eds.; Cold Spring Harbor Laboratory; Cold Spring Harbor, NY U. S. A. 11724; 1976; pp17-67.
2. Rosenberg, M.; Court, D. Regulatory sequences involved in the promotion and termination of transcription. *Annu. Rev. Genet.* **1979**, *13*, 319-353.
3. Siebenlist, U.; Simpson, R. B.; Gilbert., W. *E.coli* RNA polymerase interacts homologously with two different promoters. *Cell* **1980**, *20*, 269-281.
4. van Kampen, N.G. *Stochastic Processes in Physics and Chemistry*, 3rd Ed.; Elsevier; The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB UK; 2007; pp. 171 and pp. 114-7.
5. Shimamoto, N. Nanobiology of RNA Polymerase: Biological Consequence of Inhomogeneity in Reactant. *Chem. Rev.* **2013**, *113*, 8400–8422.
6. Egelman, E. H. ; Stasiak, A. Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP-gamma-S or ATP. *J. Mol. Biol.* **1986**, *191*, 677–697.
7. Ashby, B.; Wootton, J. C.; Fincham, J. R. Slow conformational changes of a *Neurospora* glutamate dehydrogenase studied by protein fluorescence. *Biochem. J.* **1974**, *143*, 317–329.
8. Lin, S. X.; Neet, K. E. Demonstration of a slow conformational change in liver glucokinase by fluorescence spectroscopy. *J. Biol. Chem.* **1990**, *265*, 9670–9675.
9. Stivers, 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* **1994**, *33*, 327–339.
10. Shimamoto, N.; Wu, F. Y.; Wu, C. W. Mechanism of ribonucleic acid chain initiation. Molecular pulse-labeling study of ribonucleic acid syntheses on T7 deoxyribonucleic acid template. *Biochemistry* **1980**, *20*, 4745–4755.
11. Shimamoto, N.; Toda, M.; Nara, S.; Komatsuzaki, T.; Kamagata, K.; Kinebuchi, T.; Tomizawa, J-I. Dependence of DNA length on binding affinity between TrpR and *trpO* of DNA. *Sci. Rep.-UK* Stage of publication (accepted)
12. The Feynman Lectures on Physics. Vol. 1. Chapter 46, Ratchet and pawl. Available on line: https://www.feynmanlectures.caltech.edu/I_46.html (2020/08/12)
13. Astumian, R. D.; Bier, M. Fluctuation Driven Ratchets: Molecular Motors. *Phys. Rev. Lett.* **1994**, *72*, 1766-1769.
14. Prost, J.; Chauwin, J-F.; Peliti, L.; Ajdari, A. Asymmetric Pumping of Particles. *Phys. Rev. Lett.* **1994**, *72*, 2652-2655.

15. Shimamoto, N. One-dimensional diffusion of proteins along DNA. Its biological and chemical significance revealed by single-molecule measurements. *J. Biol. Chem.* **1999**, *274*, 15293-15296.
16. Kinebuchi, T.; Shimamoto, N. One-dimensional diffusion of TrpR along DNA enhances its affinity for the operator. *Sci. Rep.-UK* Stage of publication (under review).
17. Harmer, T.; Wu, M.; Schleif, R. The role of rigidity in DNA looping-unlooping by AraC. *Proc. Natl. Acad. Sci. USA.* **2001**, *98*, 427-431.
18. Hochschild, A.; Ptashne, M. Cooperative binding of λ repressors to sites separated by integral the DNA helix. *Cell* **1986**, *44*, 681-687.
19. Ushida, C.; Aiba, H. Helical phase dependent action of CRP: effect of the distance between the CRP site and the -35 region on promoter activity. *Nucleic Acids Res.* **1990**, *18*, 6325-6330.
20. Vanzi, F.; Broggio, C.; Sacconi, L.; Pavone, F. S. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nuc. Acids Res.* **2006**, *34*, 3409-3420.
21. Aki, T.; Adhya, S. Repressor induced site-specific binding of HU for transcriptional regulation *EMBO J.* **1997**, *16*, 3666-3674.
22. Carey, J. Gel retardation at low pH resolves *trp* repressor-DNA complexes for quantitative study. *Proc. Natl. Acad. Sci. USA.* **1988**, *85*, 975-979.
23. Gunsalus, R. P.; Miguel, A. G.; Gunsalus, G. L. Intracellular *trp* repressor levels in *Escherichia coli*. *J. Bacteriol.* **1986**, *167*, 272-278.
24. Xu, G.; Weber, G. Dynamics and time-averaged chemical potential of proteins: Importance in oligomer association. *Proc. Natl. Sci. USA.* **1982**, *79*, 5268-5271.
25. Berg, O. G. Time-averaged chemical potential of proteins and the detailed balance principle (An Alternative Viewpoint). *Proc. Natl. Sci. USA.* **1983**, *80*, 5302-5303.
26. Weber, G. Stability of oligomeric proteins and its bearing on their association equilibria (A reply) *Proc. Natl. Sci. USA.* **1983**, *80*, 5303-5304.

Acknowledgments: We thank Dr. Mikito Toda of Nara Woman's University for coining "chemical ratchet" and making the theoretical frameworks, when he and N.S. proposed the concept based on our experimental results. We thank Dr. Sigetoshi Nara of Okayama University, Dr. Tamiki Komatsuzaki of Hokkaido University, Dr. Kiyoto Kamagata of Tohoku University, and the late Dr. Jun-ichi Tomizawa of National Institute of Genetics Japan for discussion critical to the proposal.

Conflicts of Interest: The authors declare no conflict of interest.