Eur. Phys. J. Plus (2023) 138:604
https://doi.org/10.1140/epjp/s13360-023-04191-y

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

# RNA polymerase is a unique Maxwell's demon that converts its transcribing genetic information to free energy for its movement

**Tatsuaki Tsuruyama**[1,2,3,4,a]

[1] Department of Physics, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan
[2] Department of Drug Discovery Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan
[3] Tazuke Kofukai Medical Research Institute, Osaka 530-8480, Japan
[4] Department of Molecular Biosciences, Radiation Effects Research Foundation, Hiroshima 732-0815, Japan

Received: 21 April 2023 / Accepted: 13 June 2023


**Abstract** RNA polymerase (RNAP) catalyzes RNA synthesis from template DNA via translocation on the DNA. The experimental value of the free energy change required for RNAP translocation exhibits an unexplainable discrepancy. To address this, we propose a transcription system model based on information thermodynamics. The state function of RNAP was defined from its position on the template DNA ($m$), its migration direction ($d$), and the deoxyribonucleotide (dNTP) that it transcribes ($N$). Based on the state function, the free energy change in the RNAP translocation was defined as the sum of $\Delta G_d$ in the movement fluctuation or the mutual entropy term $k_BT\log P(N)$ from the appearance probability $P(N)$ of $N$ to be transcribed by RNAP. In conclusion, a discrepancy in free energy change values is due to either $\Delta G_d$ or $k_BT\log P(N)$ involvement. The discrepancy highlights that RNAP is a unique information-work converter or Maxwell's demon that can feed back the obtained dNTP-type information into its self-movement along the DNA.

## 1 Introduction

The biophysical analysis of cell movement and molecular machines has advanced recently, with membrane transport [1], signal transduction [2], and molecular motors [3] being extensively studied. RNA polymerase (RNAP) is one of the most extensively studied molecular machines [4], consisting of a complex of protein subunits. RNA synthesis is one of the essential biological processes required for cell activity, and it precedes protein translation. DNA is transcribed into messenger RNA (mRNA), which is translated into amino acids by ribosomes to form a protein [5]. In terms of information transduction, transcription is the transformation of genetic information encoded in the deoxyribonucleic acid (DNA) double-strand sequence of deoxyribonucleotides (dNTPs) into a complementary sequence of ribonucleotides (rNTPs) in the single-strand RNA molecule. RNAP is the enzyme that catalyzes this transcription of DNA to mRNA. At the beginning of the transcription reaction, RNAP binds to a DNA sequence called the promoter, which is necessary for determining the starting position. Next, using one strand of double-stranded DNA as a template, RNAP sequentially adds rNTPs to the RNA precursor during its translocation along the DNA template, with fluctuations between the $3' \rightarrow 5'$ and $5' \rightarrow 3'$ orientations [6]. This RNAP movement has been postulated to be "powered" by the chemical energy released during rNTP hydrolysis while adding rNTPs [7].

Several experimental studies have estimated the mean free energy change ($\Delta G$) of RNAP during translocation. Thomen et al. [8], Wang et al. [9], and Guo et al. [10] estimated $\Delta G$ to be $-$ 1.3 $k_BT$, $-$ 0.2 to 3.4 $k_BT$, and 0.0 $k_BT$, respectively, where $k_B$ is the Boltzmann's constant, and $T$ is the temperature of the transcription system(Fig. 1). The significant difference (almost 5.0 $k_BT$) between these values has puzzled scientists [9, 11]. To address this inconsistency, we assumed that there is an unknown thermodynamic parameter; therefore, an information thermodynamics model of the transcription system is proposed. Among RNAPs, T7 RNAPI (RNA polymerase I) has been used to measure the force required for RNAP movement, and thus, this enzyme was utilized for the modeling [12, 13].

## 2 Results

### 2.1 Transcription system model

We introduced an RNAP state function into a transcription system model (Fig. 2), assuming an isobaric and isothermal system. This model consists of RNAP, a DNA template strand comprising a dNTP ($N$: dGTP, dATP, dCTP, or dTTP), an RNA precursor

[a] e-mail: tsuruyam@kuhp.kyoto-u.ac.jp (corresponding author)

Published online: 10 July 2023

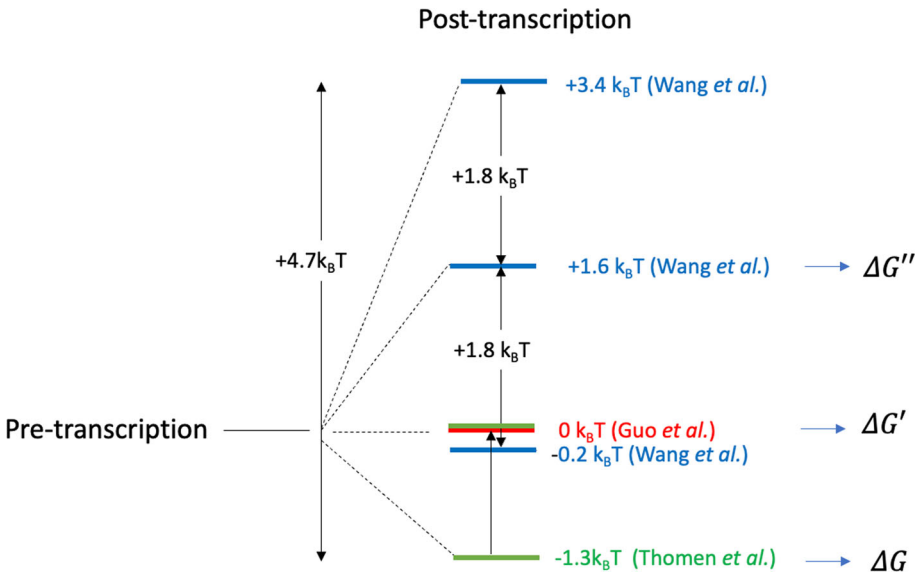


**Fig. 1** Difference in free energy levels between pre- and post-DNA transcription via RNAP. The coloured horizontal lines represent the RNAP energy levels estimated in different studies; the blue lines depict the range of − 0.2 to + 3.4 $k_BT$ reported by Wang et al. [9]; the green line depicts the value reported by Thomen et al.[8]; the red line depicts the value reported by Guo et al.[10]. $\Delta G$, $\Delta G'$, and $\Delta G''$ indicate the free energy change variations in the text

**Table 1** Experimental RNAP translocation data

| Force | Work | Free energy | | | |
|---|---|---|---|---|---|
| $f$ ($10^{-12}$ (N)) | $W$ ($10^{-21}$ (J))/$k_BT$ | $(\Delta G/k_BT)_{\min}$ | $\Delta G'/k_BT$ | | $(\Delta G''/k_BT)_{\max}$ |
| | | | [8] $\log P_+(\Delta s_{tot}(m)/P_-(-\Delta s_{tot}(m)) = 1.39$ | [10] $\log P_+(\Delta s_{\text{tot}}(m)/P_-(-\Delta s_{\text{tot}}(m)) = 1.02$ | |
| 10 [11] | 0.82 | − 0.82 | − 0.56 | − 0.2 | 1.95 |
| 11 [14] | 0.9 | − 0.9 | − 0.48 | − 0.12 | 1.87 |
| 15 [14] | 1.23 | − 1.23 | − 0.15 | 0.21 | 1.54 |
| 18 [22] | 1.48 | − 1.48 | 0.1 | 0.46 | 1.29 |
| 20 [23] | 1.64 | − 1.64 | 0.26 | 0.62 | 1.13 |
| 25 [11] | 2.05 | − 2.05 | 0.67 | 1.03 | 0.72 |
| Mean | | | | | |
| 16.5 | 1.35 | − 1.35 | − 0.03 | 0.33 | 1.42 |

$f$ denotes the force of RNAP movement. The work done on RNAP by the DNA template is denoted by $W$. A Shapiro–Wilk test of the given physical work ($W$) data revealed that $P > 0.05$, indicating that these data may follow a normal distribution. Accordingly, the mean values of $\Delta G/k_BT$, $\Delta G'/k_BT$, and $\Delta G''/k_BT$, which are − 1.35, − 0.03 or 0.33, and 1.42, respectively, can be used to validate the numerical values in Eqs. (14), (17), and (23). $\Delta G'/k_BT$ and $\Delta G''/k_BT$ were calculated according to Eq. (16) [8] and Eq. (21), respectively. The superscripts "min" and "max" represent the minimal and maximum value, respectively

consisting of an rNTP ($R$: rCTP, rUTP, rGTP, or rATP), and an rNTP pool (Table 1). Transcription proceeds in four steps: reading the dNTP of the template (dGTP in Fig. 2A) to be transcribed, recruitment of rNTP (rCTP in Fig. 2A) to be incorporated into the transcriptional system (rNTP recruitment), hybridization between the transcribed dNTP and the incorporated rNTP (rCTP in Fig. 2B), addition of rNTP (rCTP in Fig. 2C) to the RNA precursor (elongation) [13, 14], and RNAP translocation (Fig. 2D) [4]. During transcription, RNAP moves at a one-nucleotide interval step-by-step, fluctuating between forward ($3' \to 5'$:$m \to m$+1) and backward ($5' \to 3'$:$m$+1$\to m$) orientations along the template DNA from the first to the $m$th dNTP ($N_m$). In the example in Fig. 2, $N_m$ = dCTP and $R_m$ = rGTP. The abbreviations are summarized in Table 2.

## 2.2 Stochastic thermodynamics of translocation

The state of RNAP is described using the following parameters: (i) the RNAP binding position on the DNA template, denoted by the coordinate $m$, (ii) the stepwise movement direction, represented by $d$, which can be $3' \to 5'$ (+) ($m \to m + 1$) or $5' \to 3'$ (−)($m + 1 \to m$), while fluctuating between two dNTPs, and (iii) the type of dNTP on the template DNA, denoted by $N$. To characterize these parameters, we define the probability density functions as $P(m)$ for (i), $P(d)$ for (ii), and $P(N)$ for (iii). Here, we assumed that the

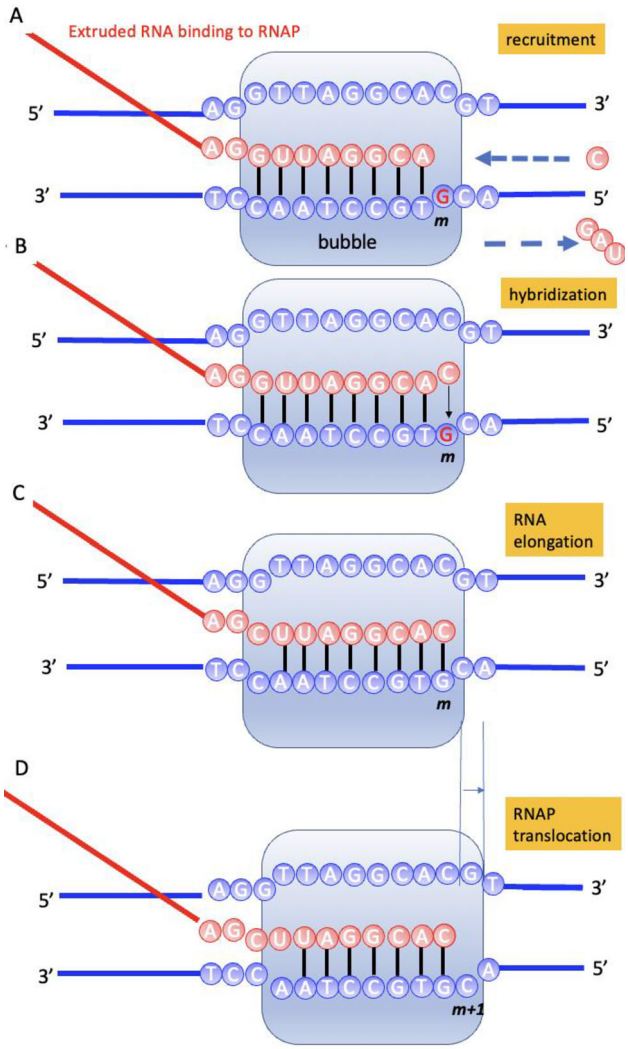


**Fig. 2** Model of transcriptional reaction: The four steps of RNA transcription. The blue lines and circles represent template DNA, while the red lines and circles represent the RNA precursor. (**A**) dGTP ($N_m$) is being transcribed, rCTP ($R_m$) is being recruited, and rGTP, rATP, and rUTP are not being recruited (red circles). (**B**) Recruited rCTP hybridizes with dGTP (hybridization). (**C**) Hybridizing rCTP is incorporated at the 3′-end of the RNA (RNA elongation). (**D**) Translocation of RNAP by one dNTP

position, movement direction, and dNTP type are independent of each other for simplicity. Therefore the total probability function $P(m,\ d,\ N)$ is approximately described by Eq. (1):

$$P(m,d,N) \cong P(m)P(d)P(N) \tag{1}$$

In the above assumption, it was considered that upon completing the movement between two dNTPs the state of the RNA transcription system is reset to the initial state each time. In addition, the variation in RNAP transition rate due to the type of dNTP is negligible. Herein, the probability for RNAP state is simplified and described in Eq. (1). Furthermore, the entropy of the transcriptional system is defined by Eq. (2) [15]:

$$S(m,d,N) = -k_B P(m,d,N) \log P(m,d,N) \tag{2}$$

Equation (2) calculates the expected value for $-\log P(m,\ d,\ N)$, which we can further define as:

$$s(m,d,N) = -k_B \log P(m,d,N) \tag{3}$$

where $s(m,\ d,\ N)$ is the stochastic entropy of RNAP, which is described by the sum of entropy, $s_{th}\ (m) = -k_B \log P(m)$, $s_d = -k_B \log P(d)$, $s_N = -k_B \log P(N)$.

**Table 2** Abbreviations used in the text

| | |
|---|---|
| RNAP | RNA polymerase |
| $d$ | Movement direction |
| $d(+)$ | Operation for RNAP translocation directions, $(m \rightarrow m+1)$ |
| $d(-)$ | Operation for RNAP translocation directions, $(m+1 \rightarrow m)$ |
| dNTP | Deoxyribonucleotide |
| dATP | Deoxyadenosine triphosphate |
| dCTP | Deoxycytidine triphosphate |
| dGTP | Deoxyguanosine triphosphate |
| dTTP | Deoxythyamine triphosphate |
| rNTP | Ribonucleotide |
| rATP | Adenosine triphosphate |
| rCTP | Cytidine triphosphate |
| rGTP | Guanosine triphosphate |
| rUTP | Uridine triphosphate |
| $m$ | Position on the DNA |
| $N_{\mathrm{m}}$ | Deoxyribonucleotide type of the $m$ position |
| $N/R$ | Deoxyribonucleotide type / Ribonucleotide type |
| $P(m,d,n)$ | RNAP state function |
| $Q_{m+}$ | Heat flow when RNAP moves from $m \rightarrow m+1$ for $3' \rightarrow 5'$ and $m+1 \rightarrow m$ for $5' \rightarrow 3'$ |
| $Q_{m-}(t)$ | Heat flow when RNAP moves from $m+1 \rightarrow m$ for $5' \rightarrow 3'$ |
| $\Delta s$ | Entropy change |
| $\Delta S_{tot}$, $\Delta s_{tot}$ | Total entropy changes |
| $\Delta S_{th}$, $\Delta s_{th}$ | Thermodynamic entropy changes |
| $\Delta S_d$, $\Delta s_d$ | Entropy changes in the movement fluctuation |
| $S(N_m)$,$s(N_m)$ | Shannon entropy from DNA sequence consisting of $N_m$ |
| $W$ | Work done by the template DNA to RNAP |
| $\Delta G$ | Gibbs free energy change |
| $k_B$ | Boltzmann constant |
| $t_{\mathrm{m}+}$ | Time necessary for RNAP movement $m \rightarrow m+1$ |
| $t_{\mathrm{m}-}$ | Time necessary for RNAP movement $m+1 \rightarrow m$ |
| $h(R_m, N_m)$ | Mutual entropy in gaining $N_m$ dNTP type |
| $s(R_m \mid N_m)$ | Conditional entropy of $R_m$ given $N_m$ |

### 2.3 Probability function and entropy of RNAP

First, we calculated the thermodynamic entropy change $\Delta S_{th}$ and the stochastic entropy change $\Delta s_{th}$ between the RNAP at the $m$ and the $m+1$ positions as [16, 17]:

$$\begin{aligned}\Delta S_{th} &:= \sum_{m=1}^{n} \Delta s_{th}(m) = \sum_{m=1}^{n-1} (-k_B \log P(m) + k_B \log P(m+1)) \\ &= -k_B \log P(m=1) + k_B \log P(m=n)\end{aligned} \tag{4}$$

with the stochastic entropy:

$$\Delta s_{th}(m) := -k_B \log P(m) + k_B \log P(m+1) \tag{5}$$

In the process of forward RNAP migration, in the first step, RNAP binds to the $3'$-end of the promoter element of the template DNA ($m=1$) with probability $P(m=1)$, and then, it moves to $m \rightarrow m+1$ $(1 \leq j \leq m)$ at the transition rate $k_{j+}$, finally reaching the $m$. In addition, consider that RNAP can move to $j+1 \rightarrow j$ at the transition rate $k_{j-}$, which is sufficiently small, finally reaching the $m$. Here, $Q_{m+}(t)$ and $Q_{m-}(t)$, which represent heat flow when RNAP moves from $m \rightarrow m+1$ for $3' \rightarrow 5'$ and $m+1 \rightarrow m$ for $5' \rightarrow 3'$, respectively, were introduced. The entropy change $\Delta S_d$ and the stochastic entropy change $\Delta s_d(m)$ in the RNAP movement at the $m$ and the $m+1$ positions are described as [17]:

$$\Delta S_d := \sum_{m=1}^{n} \Delta s_d(m) = \sum_{m=1}^{n} \frac{Q_{m^+}(t) - Q_{m^-}(t)}{T} = \sum_{m=1}^{n} k_B \log \frac{k_{m^+}}{k_{m^-}} \tag{6}$$

and

$$\Delta s_d(m) := \frac{Q_{m^+}(t) - Q_{m^-}(t)}{T} \tag{7}$$

In this case, the probability distribution function of RNAP moving forward at the $m$ position is given by [16, 17]:

$$P(m, d = +) = P(m = 1) \prod_{j=1}^{m} k_{j^+} \tag{8}$$

The probability distribution function of RNAP at the $m$ position moving backwards is given by:

$$P(m, d = -) = P(m = n) \prod_{j=n}^{m} k_{j^-} \tag{9}$$

The total stochastic entropy $\Delta s_{tot}\ (m)$ can be expressed from Eqs. (8–9) as [17]:

$$\begin{aligned} \Delta s_{tot}(m) &:= \Delta s_d(m) + \Delta s_{th}(m) = (\log P(m, d = +) - \log P(m, d = -)) \\ &= -k_B \log \frac{P(m = 1) \prod_{j=1}^{m} k_{j^+}}{P(m = n) \prod_{j=n}^{m} k_{j^-}} \\ &= -k_B \log \frac{P(m, d = +)}{P(m, d = -)} \end{aligned} \tag{10}$$

## 2.4 Path integral for RNAP movement

Here, we defined the path integral for each transcriptional orientation of the template DNA. $d\ (+)$ and $d\ (-)$ are operations for both RNAP translocation directions, $+\ (m \to m + 1)$ and $-\ (m + 1 \to m)$, respectively. From Eq. (10) [17] (Appendix 1):

$$\begin{aligned} P_+(\Delta s_{tot}(m)) &:= \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) + k_B \log \frac{P(m, d = +)}{P(m, d = -)} P(m, d = +)\right) \\ &= \exp\left(\frac{\Delta s_{tot}(m)}{k_B}\right) P_-(-\Delta s_{tot}(m)) \end{aligned} \tag{11}$$

Therefore,

$$\Delta s_{tot}(m)/k_B = \log \frac{P_+(\Delta s_{tot}(m))}{P_-(-\Delta s_{tot}(m))} \tag{12}$$

Using the general thermodynamic relation of entropy, $\Delta s = (W^{ex} - \Delta G)/T$, where $\Delta G$ denotes the free energy change without the contribution of the thermodynamic fluctuation and assuming that $W^{ex}$ (i.e., work done by the template DNA to RNAP) is equal to $-W$,

$$W(m) \leq -\Delta G(m) \tag{13}$$

$W(m)$ and $-\Delta G(m)$ represent the physical work and the free energy change, respectively, during the RNAP movement $m \to m + 1$. From the experimental data, the mean of $W(m) \sim 1.4\ k_BT$ (the mean value of work in Table 1), Eq. (14) was derived:

$$\Delta G(m) \geq -1.4\, k_B T \tag{14}$$

where the obtained lower limit value is approximately equal to the mean Gibbs free energy reported by Thomen et al. [8].

## 2.5 Fluctuation in RNAP movement

Here we introduced $t_{m+}$ and $t_{m-}$, which represent the time necessary for RNAP movement $m \to m + 1$ and $m + 1 \to m$, respectively. The experimental data has shown the average value of $t_{m+}/t_{m-} = 2.8–4.0$ [11, 18, 19]. Using this value, we estimated the probability ratio $P_+(\Delta s_{tot}\ (m))\ /P_-\ (-\Delta s_{tot}(m))$:

$$\frac{P_+(\Delta s_{tot}(m))}{P_-(-\Delta s_{tot}(m))} = t_{m+}/t_{m-} \cong 2.8 - 4.0 \tag{15}$$

Accordingly, the movement trajectory on the template DNA could be simulated under the condition that the RNAP moves randomly forward with a probability of $P_+(\Delta s_{tot}) = 0.8$ and backwards with a probability of$P_-(-\ \Delta s_{tot}) = 0.2$. This simulation is compatible with the simulation of RNAP movement on the template DNA reported in experimental studies [11, 18, 19] (Fig. 3A, B; Table 1; Appendix 2).

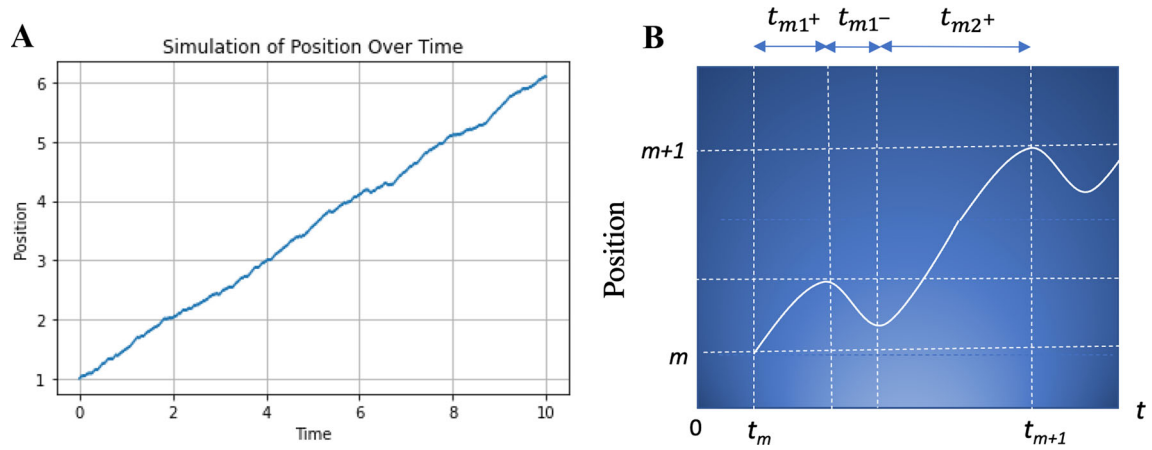


**Fig. 3** RNAP movement trajectory along the DNA template strand. **A** An example of the simulation plot. The algorithm is described in the Appendix 2, where $t_{m+}/t_{m-}$ was set to 4.0. The vertical axis represents the RNAP position ($1 \leq m \leq 6$). The horizontal axis represents the unit of time. **B** RNAP translocation trajectory. The vertical axis represents the position $m$ and $m$+1. The ratio $P_+(m)/P_-(m)$ was estimated from ${t_m}^+/{t_m}^- = \Sigma_{i=1}{}^2\ {t_{mi}}^+/\Sigma_{mi=1}{}^1\ {t_{mi}}^-$

Using Eq. (15),

$$W(m) \leq T\Delta s_{tot}(m) =: -\Delta G'(m) + k_B T \log \frac{P_+(\Delta s_{tot}(m))}{P_-(-\Delta s_{tot}(m))} \tag{16}$$

Considering the mean of $W(m) \sim 1.4\ k_B T$ (Table 1) and Eqs. (15, 16 ), Eq. (17) was obtained:

$$\Delta G'(m) \sim 0\, k_B T \tag{17}$$

where the obtained value is approximately equal to the lower Gibbs free energy value reported by Wang et al. and Guo et al. [9, 10].

### 2.6 Stochastic thermodynamics of reading genetic information

Furthermore, the contribution of the template DNA sequence genetic information was evaluated (Fig. 2; T7 promoter sequence used in experimental studies is shown in Table S1) [2, 20]. In a pre-transcriptional state, the information of $N_m$ is given by $s(N_m) = -\log P(N_m)$, where $P(N_m)$ denotes the frequency of each type of dNTP—alternatively, $s(N_m) = s_{Nm}/k_B$. When the transcription error is negligible, the conditional entropy when type $N_m$ is transcribed into $R_m$, complementary to $N_m$, by the RNAP is given by the conditional entropy that is defined by $s(R_m|N_m)$): $= -\log P(N_m)) = -\log P(R_m)$. Here, the mutual information for RNAP recognition of $N_m$ as $R_m$ is given by $h(R_m; N_m) := s(R_m) - s(R_m|N_m) = -\log P(R_m) - [-\log P(N_m)] = 0$. As a post-transcriptional state, $N_m$ does not change, and $R_m$ has already been determined. Therefore, $s(R_m|N_m) = 0$ gives the conditional entropy. At this point, mutual information $h(R_m, N_m)$ is given by $h(R_m, N_m) = s(R_m) - s(R_m|N_m) = -\log P(R_m) - 0 = s(R_m)$. Mutual information $-\log P(R_m)$ is fed back to the template DNA, and$s(R_m)$ becomes zero when $R_m$ is determined (Table S2). In this case, mutual information can be expressed using entropy $s_{Nm}$:

$$h(R_m, N_m) = s(N_m) = s_{Nm}/k_B \tag{18}$$

$$W(m) \leq -\Delta G(m) = T(\Delta s_0(m) + \Delta s_{tot}(m)) \tag{19}$$

$\Delta s_0$ denote the information gain from the genetic information. Considering the contribution of the mutual entropy to $\Delta G(m)$, Eq. (20) was obtained (Appendix 3):

$$\begin{aligned}\Delta s_{tot}(m) &= -k_B \log \frac{P_+(\Delta s_{tot}(m) + k_B \log P(N_m))}{P_-(-\Delta s_{tot}(m) - k_B \log P(N_m))} + k_B \log P(N_m) \\ &= -k_B \log \frac{P_+(\Delta s_{tot}(m) + k_B \log P(N_m))}{P_-(-\Delta s_{tot}(m) - k_B \log P(N_m))} - k_B T h(R_m, N_m)\end{aligned} \tag{20}$$

Substitution of Eq. (20) into Eq. (19) gives:

$$\begin{aligned}W(m) &\leq -\Delta G(m) \\ &= -\Delta G''(m) + k_B T \log \frac{P_+(\Delta s_{tot}(m) + k_B \log P(N_m))}{P_-(-\Delta s_{tot}(m) - k_B \log P(N_m))} + k_B T h(R_m, N_m)\end{aligned} \tag{21}$$

where $-\Delta G''(m) = T\Delta s_0\ (m)$. Accordingly, by substituting the mean experimental value of $W\ (m) \sim 1.4\ k_BT$, $P_+(\Delta s_{tot}(m) + k_B \log P(N_m))\ /P_-(-\Delta s_{tot}(m) - k_B \log P(N_m)) = t_{m+}/t_{m-} = 2.8$–$4.0$ [8, 10], and $-s_{Nm} = h\ (R_m, N_m) = -\log P(N_m) \sim 1.4$ (Table 1 and Table S1) into Eq. (21), Eq (22) was obtained:

$$1.4\,k_BT \leq -\Delta G''(m) + k_BT \log 4.0 + 1.4\,k_BT \tag{22}$$

and therefore

$$\Delta G''(m) \leq 1.4\,k_BT, \tag{23}$$

where the obtained lower limit value is approximately equal to the mean Gibbs free energy value of $1.6\,k_BT$ reported by Wang et al. (Fig. 1). In conclusion, the discrepancy in the values of free energy changes in enzyme transfer stems from the different definitions of free energy.

## 3 Discussion

It has been found that free energy exhibits various values based on whether fluctuations in the direction of movement and the contribution of mutual information are included. The definition of mutual information is theoretically arbitrary and depends on the model. Several experiments support the conversion of mutual information to physical work $W(m)$ [21]. However, there is currently no evidence of any biochemical reactions that can convert mutual information from the genetic information into physical work.

The information of the dNTP type of the template DNA to be transcribed is fed back to the DNA by RNAP, preventing RNAP fluctuation along the trajectory of the DNA.This is achieved by anchoring the RNAP through the hybridization between the incorporated rNTP and the transcribed dNTP. As a result, RNAP moves forward predominantly in the $3' \rightarrow 5'$ direction, indicating the work done on RNAP by DNA. This process resembles the translocation of Brownian particles in the presence of a periodic potential, as described in the mutual information–work conversion model [21. In this model, a Maxwell's demon can move a particle against its potential by repeating the process of obtaining information about the direction of movement of the particle [10]. From the current model, RNAP can be regarded a Maxwell's demon capable of converting mutual information from DNA sequence into physical work.

The conversion of information into the translocation work reminds us of the Szilard engine [22]. In the current model, the RNAP demon reads the dNTP type of nucleic acid in the template or determines whether the recruited rNTP is complementary to the template DNA. When the rNTP is complementary, RNAP recruits it to the active catalytic center and catalyzes its binding to RNA, forming a new hybridization and facilitating RNAP translocation. In contrast, if the rNTP is not complementary, it will not be recruited and RNAP translocation will not occur [12].

It is noteworthy that in template DNA transcription, RNAP-transcribed information is retained in RNA molecules in another form of genetic information. Landauer argued that when information obtained from the measurement is stored in the memory of the demon, it is necessary for the demon to erase the information in the subsequent formation step. This argument has widely been accepted as a solution to Maxwell's demon paradox. However, considering that RNAP retains the genetic information as RNA, it is unnecessary to erase the memory according to Landauer's principle [23]. RNAP can repeat the cycle of reading the dNTP information and convert it into self-moving work, which is a molecular motor that can convert genetic information into work. Besides, the current model claims that RNAP possesses the unique property of being able to move itself owing to its own work. This observation seems to violate the second law of thermodynamics, and RNAP appears to be a self-driven molecular machine that cannot function unless it can convert the work of mutual information. The fact that RNAP has developed such a mechanism is intriguing in terms of molecular evolution.

Frequent fluctuations between at least $3' \rightarrow 5'$ and $5' \rightarrow 3'$ are approximately 3:1 to 4:1 and may be influenced by the fact that there are four types of nucleic acid nucleotides. It has been experimentally reported that there is a difference in free energy required for the movement for each type [6, 10, 24, 25].

A limitation of this model is that the dNTP type and direction of movement were considered independent factors in Eq. (1). Accordingly, the RNAP state probability function, $P\ (m, d, N)$, is not necessarily described using the products, $P(m)$, $P(d)$, and $P(N)$. As per experimental data, the torque effect on DNA by RNAP transfer is not completely reset, and the above assumption of the Markov process does not necessarily hold true [24]. However, since the simulation based on the simple assumption of the fluctuation ratio between $m \rightarrow m + 1$ movement and $m+1 \rightarrow m$ movement in Fig. 3 does not significantly differ from the experimental results, Eq. (1) will not be significantly different from the actual equation of state when ignoring the difference in RNAP movement due to the difference in dNTP. In addition, in the future, detailed analyses using models adhering to the solid phase in the cytoplasmic fluid should be performed [25].

In conclusion, RNAP is capable of directly converting genetic code information in the template DNA into movement of RNAP itself on the DNA. This provides a new perspective on transcriptional regulation and information thermodynamics.

## 4 Numerical data estimation

Mathematica™ version 11 (Wolfram, Champaign, Il) and Jupyter Notebook 6.3.0 were used.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1140/epjp/s13360-023-04191-y.

**Acknowledgements** This research was funded by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Synergy of Fluctuation and Structure: Quest for Universal Laws in Nonequilibrium Systems, P2013-201 Grant-in-Aid for Scientific Research on Innovative Areas, MEXT, Japan). I would like to express my gratitude to Professor Masayuki Imai from Tohoku University and Professor Takahiro Sagawa from the University of Tokyo for their valuable advice. This research was financially provided by the Radiation Effects Research Foundation (RERF) provided for the publication of the article. The views of author do not reflect those of RERF.

**Data availability** All data are available in the main text or the supplementary materials.



## Appendix 1

Equation (11) is given as follows:

$$\begin{aligned}
P_+(\Delta s_{tot}(m)) &= \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) - k_B \log \frac{P(m,d=+)}{P(m,d=-)}\right) P(m,d=+) \\
&= \exp\left(\frac{\Delta s_{tot}(m)}{k_B}\right) \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) - k_B \log \frac{P(m,d=-)}{P(m,d=+)}\right) P(m,d=-) \\
&= \exp\left(\frac{\Delta s_{tot}(m)}{k_B}\right) \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) + k_B \log \frac{P(m,d=-)}{P(m=1,d=+)}\right) P(m,d=-) \\
&= \exp\left(\frac{\Delta s_{tot}(m)}{k_B}\right) \int_{3'}^{5'} d(-)\delta\left(-\Delta s_{tot}(m) - k_B \log \frac{P(m,d=-)}{P(m,+)}\right) P(m,d=-) \\
&= \exp\left(\frac{\Delta s_{tot}(m)}{k_B}\right) P_-\left(-\frac{\Delta s_{tot}(m)}{k_B}\right)
\end{aligned} \tag{24}$$

Therefore,

$$\Delta s_{tot}(m)/k_B = \log \frac{P_+(\Delta s_{tot}(m))}{P_-(-\Delta s_{tot}(m))} \tag{25}$$

## Appendix 2

Simulation algorithm:

The probability of RNAP transitioning from the current position $X(t) = m$ to the next position $m < x \leq m + 1$ is given by the two equations (i) and (ii):

(i)
$$P\left(X(t+1) = x | X(t) = m, \ \sigma_f^2\right) = P_+(\Delta s_{tot}(m)) \frac{1}{\sqrt{2\pi\ \sigma_f^2}} \exp\left\{\frac{(-(x-(m\ +\ 1))^2}{2\ \sigma_f^2}\right\} \tag{i}$$
$$+P_-(\Delta s_{tot}(m)) \frac{1}{\sqrt{2\pi\ \sigma_f^2}} \exp \frac{(-(x-(m-\ 1))^2}{2\ \sigma_f^2} \ (if\ x > 1)$$

(ii)
$$P\left(X(t+1) = x | X(t) = y, \ \sigma_f^2\right) = P_+(\Delta s_{tot}(m)) \frac{1}{\sqrt{2\pi\ \sigma_f^2}} \exp\left\{\frac{(-(x-(m+\ 1))^2}{2\ \sigma_f^2}\right\} (if\ x = 1) \tag{ii}$$
$$P\left(X(t+1) = x | X(t) = m, \ \sigma_f^2\right) = 0\ (otherwise)$$

As an argument, $\sigma_f$ specifies the standard deviation of the fluctuations. Fluctuations at each time point occur randomly. In the current simulation, the fluctuation was calculated using the Python library as follows: fluctuation = np.random.normal(fluctuation_mean, fluctuation_std).

If $m$ is greater than 1, the next position is ahead of the current position, and the probability of transitioning to $x$ is a combination of two terms:

$P_+$ ($\Delta s_{tot}$) multiplied by the probability of transitioning from $m$ to $m + 1$, denoted as $P(x; m + 1, \sigma_f)$ and $P_-$ ($\Delta s_{tot}$) multiplied by the probability of transitioning from $m$ to $m$ -1, denoted as $P(x; m - 1, \sigma_f)$. If $m$ is equal to 1, the next position is the same as the current position, and the probability of transitioning to $m$ is given by $P_+$ ($\Delta s_{tot}$) multiplied by the probability of transitioning from $m$ to $m + 1$. If $m$ is less than or equal to 1, the probability of transitioning to $m$ is 0, indicating that the next position cannot be less than 1. In summary, the formula calculates the probability of transitioning from the current position $m$ to the next position $x$, considering the probabilities of moving forward and backward as well as the probabilities of specific transitions determined by the $\sigma_f$ parameter. The formula represents the conditional probability of the next position $X(t + 1)$ being $x$, given that the current position $X$(t) is $m$.

The Python code is as follows:

```
import numpy as np

import matplotlib.pyplot as plt

# Initial conditions.

t0 = 0 # Initial time

x0 = 1 # Initial position (greater than or equal to 1)

p_plus = 0.75 # Probability of moving forward

p_minus = 0.25 # Probability of moving backwards

T = 10 # Total simulation time.
dt = 0.01 # Time interval
# Calculation of the number of transitions.
transition_count = int(T/dt)
# Lists to store time and position during the simulation.
time_points = np.linspace(t0, T, transition_count + 1)
positions = [× 0]
#Simulation execution.
for t in range(transition_count):
u = np.random.uniform(0, 1) # Generate a uniform random number between 0 and 1
# Randomly move forward or backward.
if u<p_plus:d
x_dt = 1
else:
dx_dt = − 1
# Calculate the next position and add it to the list.
x_next = positions[− 1] + dx_dt * dt
if x_next< 1: # 1 Do not take values less than 1
x_next = 1
positions.append(x_next)
# If x is greater than or equal to 10, end the simulation.
if x_next> = 10:
break
# Plotting the results.
plt.plot(time_points, positions)
plt.xlabel('Time')
plt.ylabel('Position')
plt.title('Simulation of Position Over Time')
```

```
plt.grid(True)
plt.show()
```

## Appendix 3

Equation (20) is given as follows:

$$
\begin{aligned}
P_+(\Delta s_{tot}(m) - k_B \log P(N_m)) &= \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) - k_B \log P(N_m) + k_B \log \frac{P(m,+;N_m)}{P(m,-;N_m)}\right) P(m,+;N_m) \\
&= \exp\left(\frac{\Delta s_{tot}(m) - k_B \log P(N_m)}{k_B}\right) \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) - k_B \log P(N_m) - k_B \log \frac{P(m,-;N_m)}{P(m,+;N_m)}\right) P(n,-;N_m) \\
&= \exp\left(\frac{\Delta s_{tot}(m) - k_B \log P(N_m)}{k_B}\right) \int_{3'}^{5'} d(+)\delta\left(\Delta s_{tot}(m) - k_B \log P(N_m) + k_B \log \frac{P(m,-;N_m)}{P(m,+;N_m)}\right) P(n,-;N_m) \\
&= \exp\left(\frac{\Delta s_{tot}(m) - k_B \log P(N_m)}{k_B}\right) \int_{3'}^{5'} d(-)\delta\left(-\Delta s_{tot}(m) + k_B \log P(N_m) + k_B \log \frac{P(m,-;N_m)}{P(m,+;N_m)}\right) P(n,-;N_m) \\
&= \exp\left(\frac{\Delta s_{tot}(m) + k_B \log P(N_m)}{k_B}\right) P_-(-\Delta s_{tot}(m) - k_B \log P(N_m))
\end{aligned}
$$

Therefore,

$$
(\Delta s_{tot}(m) - k_B \log P(N_m))/k_B = -\log \frac{P_+(\Delta s_{tot}(m) + k_B \log P(N_m))}{P_-(-\Delta s_{tot}(m) - k_B \log P(N_m))}
$$

and

$$
\Delta s_{tot}(m) = -\log \frac{P_+(\Delta s_{tot}(m) + k_B \log P(N_m))}{P_-(-\Delta s_{tot}(m) - k_B \log P(N_m))} + k_B \log P(N_m)
$$

## References


1. A.M. Berezhkovskii, S.M. Bezrukov, Phys. Rev. Lett. **100**(3), 038104 (2008)
2. S. Ito, T. Sagawa, Nat. Commun. **6**, 7498 (2015)
3. A.W. Lau, D. Lacoste, K. Mallick, Phys. Rev. Lett. **99**(15), 158102 (2007)
4. J. Ma, C. Tan, X. Gao, R.M. Fulbright Jr., J.W. Roberts, M.D. Wang, Proc. Natl. Acad. Sci. U. S. A. **116**(7), 2583 (2019)
5. H. Yoon, A. Warshel, Proteins **84**(11), 1616 (2016)
6. L.T. Da, E. Chao, Y. Shuai, S. Wu, X.D. Su, J. Yu, Nucleic Acids Res. **45**(13), 7909 (2017)
7. C.A. Minetti, D.P. Remeta, H. Miller, C.A. Gelfand, G.E. Plum, A.P. Grollman, K.J. Breslauer, Proc. Natl. Acad. Sci. U. S. A. **100**(25), 14719 (2003)
8. P. Thomen, P.J. Lopez, U. Bockelmann, J. Guillerez, M. Dreyfus, F. Heslot, Biophys. J. **95**(5), 2423 (2008)
9. M.D. Wang, M.J. Schnitzer, H. Yin, R. Landick, J. Gelles, S.M. Block, Science **282**(5390), 902 (1998)
10. Q. Guo, R. Sousa, J. Mol. Biol. **358**(1), 241 (2006)
11. L. Bai, R.M. Fulbright, M.D. Wang, Phys. Rev. Lett. **98**(6), 068103 (2007)
12. S. Wu, J. Wang, X. Pu, L. Li, Q. Li, Biophys. J. **114**(8), 1755 (2018)
13. B. Duan, J. Yu, J. Phys. Chem. B **121**(15), 3777 (2017)
14. T.W. Turowski, E. Petfalski, B.D. Goddard, S.L. French, A. Helwak, D. Tollervey, Mol. Cell. **79**(3), 488 (2020)
15. T. Hatano, S.-I. Sasa, Phys. Rev. Lett. **86**(16), 3463 (2001)
16. Y. Hasegawa, Phys. Rev. Lett. **125**(5), 050601 (2020)
17. Y. Hasegawa, T. Van Vu, Phys. Rev. Lett. **123**(11), 110602 (2019)
18. E.A. Abbondanzieri, W.J. Greenleaf, J.W. Shaevitz, R. Landick, S.M. Block, Nature **438**(7067), 460 (2005)
19. L. Bai, A. Shundrovsky, M.D. Wang, J. Mol. Biol. **344**(2), 335 (2004)
20. T. Sagawa, M. Ueda, Phys. Rev. Lett. **109**(18), 180602 (2012)
21. S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, M. Sano, Nat. Phys. **6**(12), 988 (2010)
22. L. Szilard, Behav. Sci. **9**(4), 301 (1964)
23. L. R, IBM Journal of Research and Development, **5** (3), 183 (1961).
24. J. Ma, L. Bai, M.D. Wang, Science **340**(6140), 1580 (2013)
25. S. A. R. Z. Asghar, Ali N, Soft Computing (2023).