

Article

Harnessing Information Thermodynamics: Conversion of DNA Information into Mechanical Work in RNA Transcription and Nanopore Sequencing

Tatsuaki Tsuruyama ^{1,2,3,4} 

¹ Department of Discovery Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan; tsuruyam@ddm.med.kyoto-u.ac.jp; Tel.: +81-75-366-7417

² Department of Physics, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan

³ Department of Clinical Laboratory, Graduate School of Health Sciences, Kyoto Tachibana University, Kyoto 607-8175, Japan

⁴ Kitano Medical Institute, Kitano Hospita, Osaka 530-8480, Japan

Abstract: Recent advancements in information thermodynamics have revealed that information can be directly converted into mechanical work. Specifically, RNA transcription and nanopore sequencing serve as prime examples of this conversion, by reading information from a DNA template. This paper introduces an information thermodynamic model in which these molecular motors can move along the DNA template by converting the information read from the template DNA into their own motion. This process is a stochastic one, characterized by significant fluctuations in forward movement and is described by the Fokker–Planck equation, based on drift velocity and diffusion coefficients. In the current study, it is hypothesized that by utilizing the sequence information of the template DNA as mutual information, the fluctuations can be reduced, thereby biasing the forward movement on DNA and, consequently, reducing reading errors. Further research into the conversion of biological information by molecular motors could unveil new applications, insights, and important findings regarding the characteristics of information processing in biology.

Keywords: mutual information; transcription; RNA polymerase; free energy change; fluctuation theorem; nanopore sequencing



Citation: Tsuruyama, T. Harnessing Information Thermodynamics: Conversion of DNA Information into Mechanical Work in RNA

Transcription and Nanopore Sequencing. *Entropy* **2024**, *26*, 324. <https://doi.org/10.3390/e26040324>

Academic Editor: Yasar Demirel

Received: 25 March 2024

Revised: 4 April 2024

Accepted: 8 April 2024

Published: 11 April 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The integration of nonequilibrium thermodynamics into biophysics has markedly progressed, covering areas such as membrane transport [1–3], signal transduction [4,5], and the study of molecular motors [6–10]. This field has seen significant theoretical advancements, notably the discovery of the fluctuation theorem [11–14], which have enabled mechanical measurements to determine the forces propelling molecular motors [6] and to estimate torque in singular instances, including experiments on F1-adenosine triphosphatase [7–9]. RNA polymerase (RNAP), a molecular motor of considerable interest, translocates along DNA to synthesize RNA precursors, driven predominantly by the free energy change from the hydrolysis of ribonucleotide triphosphates (rNTP) [15,16]. Notably, reports on the free energy changes associated with RNAP translocation have varied widely, ranging from $-1.3 k_B T$ to $3.4 k_B T$ [17–19], a variance that has, yet, to be adequately explained [17–20], thus highlighting a knowledge gap in our understanding of RNAP mechanics.

Previously, we reported on an information thermodynamics model related to the movement of RNAP [21]. In that report, we detailed the kinetics of RNAP and described a mechanism through which mutual information, derived from the sequence information of the template DNA, is converted into work for RNAP's own movement. This framework within information thermodynamics explains that the extent to which fluctuating stochastic processes and contributions of mutual information are incorporated in the free energy change results in the variations in the value.

This article introduces an advanced model for the transcription process by RNA polymerase (RNAP), based on the foundation of experimental studies [22]. The model employs the Fokker–Planck equation, incorporating drift and diffusion coefficients characteristic of molecular motors, to elucidate the stochastic behavior of RNAP movement. Furthermore, this framework is extended to explore the thermodynamic properties of molecular motors in nanopore sequencing. Nanopore sequencing represents an innovative technology for deciphering long sequences of DNA, by passing individual DNA molecules through a nanopore embedded in a synthetic membrane. As each deoxyribonucleotide triphosphate (dNTP = dATP, dTTP, dCTP, dGTP) traverses the nanopore, it induces distinctive perturbations in the ionic current, enabling the identification of the specific dNTP type. This technique offers a means to directly sequence DNA and RNA molecules without the need for amplification via polymerase chain reaction (PCR) or complex chemical reactions. Fundamentally, nanopore sequencing is based on the principle of detecting changes in electrical conductivity caused by nucleic acid molecules passing through nanoscale pores embedded in an electrically insulating membrane. In nanopore sequencing machinery, a type of molecular motor, known as helicase, plays a role in unwinding double-stranded DNA and supplying it in a single-stranded form to the nanopore. In this process, the helicase moves along the DNA strand. The molecular motor bears a resemblance to RNAP moving along template DNA. Therefore, we propose a unified theoretical model to understand these mechanisms cohesively [23,24].

The commonality between nanopore sequencing and transcription processes involving molecular motors lies in the fact that during each step of movement between two dNTPs, the molecular motor is repeatedly reset to its pre-movement state (refer to Figure 1). This phenomenon originates from the repetitive structure of DNA, which is a double helix. Furthermore, although to varying degrees, both RNA polymerase (RNAP) and nanopore motor proteins exhibit fluctuations in their direction of movement, ultimately demonstrating a stochastic bias towards forward motion. To account for this bias, we previously reported that RNAP can function as a Maxwell’s demon, capable of converting information encoded in the DNA sequence into directed motion [23,24]. In this paper, we explore the possibility of extending this concept to the relative movement of motor molecules along template DNA in nanopore sequencing.

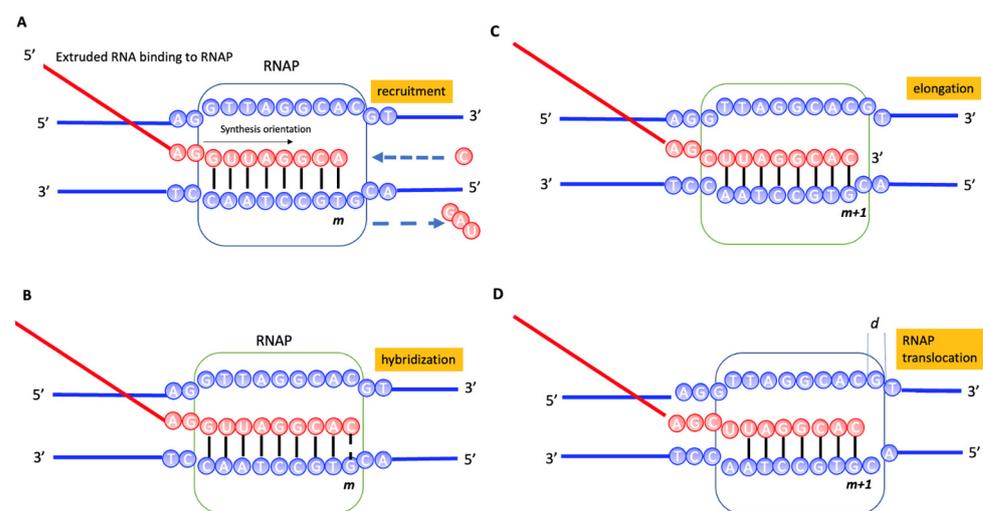


Figure 1. RNAP translocation model. The figure represents a transcription apparatus, where the thin rectangle symbolizes RNA polymerase (RNAP). The thick red line represents the RNA precursor, composed of ribonucleotides rNTP (rGTP, rCTP, rATP, rUTP) denoted by the letters G, C, A, U. The thick blue line signifies the template DNA, composed of deoxyribonucleotides dNTP (dGTP, dCTP, dATP, dTTP) represented by the letters G, C, A, T. Detailed naming for rNTP and dNTP is shown in the Abbreviation section. m and $m + 1$ denote the position of dNTPs. (A) Ribonucleotide recruitment:

RNA polymerase (RNAP) accurately reads dGTP on the template DNA. It then selectively incorporates rCTP, while excluding non-complementary rNTPs. (B) Ribonucleotide hybridization: the incorporated rCTP forms hydrogen bonds with dGTP at the RNA reading edge. (C) Elongation: at this stage, the incorporated rCTP is added to the growing RNA precursor. (D) Forward bias in RNAP motion: the incorporation of rCTP prevents any arbitrary backward movement of RNAP, inducing a forward bias in its motion. Thus, the nucleotide type, in this case, dGTP, is represented as mutual information, converting into the translocation work of RNAP.

2. Methods

2.1. Transcription System Model

The molecular mechanism of RNAP involves intricate processes where RNAP reads the DNA template and synthesizes RNA strands, commencing with the precise recognition of promoter sequences to initiate transcription. This is followed by the elongation phase, where RNAP exhibits remarkable processivity, adding ribonucleotides in a template-directed manner, while ensuring transcriptional fidelity through nucleotide selection and proofreading. The activity is finely modulated by various transcription factors and co-factors that influence initiation, elongation, and termination phases. Structural dynamics play a crucial role, with conformational changes within RNAP being pivotal throughout the transcription cycle [17,19,25–28]. The transcription proceeds in four steps, as follows: (1) recruitment of dNTP to be transcribed (Figure 1A), (2) hybridization of the recruited rNTP with the template DNA (Figure 1B), (3) RNA precursor elongation by adding the rNTP to the RNA precursor bound to the RNAP (Figure 1C), and (4) the entire RNAP translocation (Figure 1D).

Thus, the nucleotide type, in this case, dGTP, is represented as mutual information, converting into the translocation work of RNAP.

When analyzing the movement of RNAP with short template DNA, the master equation, which represents the time evolution of the probability of each state, proves to be effective. If the probability of RNAP moving from position x to position $x + 1$ in the next time step is P_f , and the probability of moving from position x to position $x - 1$ is P_b , the master equation can be described as follows. The temporal evolution of the probability $P(x,t)$ that RNAP is located at position x and time t is described by the following expression:

$$\frac{\partial P(x,t)}{\partial t} = P_f P(x-1,t) + P_b P(x+1,t) - (P_f + P_b) P(x,t) \quad (1)$$

Here, the first term on the right side represents an increase in probability due to moving from $x - 1$ to x , the second term represents an increase from $x + 1$ to x , and the third term represents a decrease from x to other positions. $P_f(x,t)$ and $P_b(x,t)$ represent the probability distribution functions for RNA polymerase moving forward and backward, respectively, at any time t and position x . Meanwhile, in the continuous limit for long template DNA, the probability $P(x,t)$ that RNAP is located at position x is as follows:

$$\frac{\partial P(x,t)}{\partial t} = D \frac{\partial^2 P(x,t)}{\partial x^2} - \mu \frac{\partial P(x,t)}{\partial x} \quad (2)$$

where D is the diffusion coefficient and μ is the drift velocity. If the drift velocity and the diffusion coefficient are constant, the probability density function follows a normal distribution. This process is similar to Brownian motion and if the initial position is known as $x = 0$, the distribution of RNAP moving forward at any time t and position x is:

$$P_f(x,t) \propto \exp\left(-\frac{(x - \mu t)^2}{4Dt}\right) \quad (3)$$

The coordinate of the leading edge reached by RNAP is denoted by $L = Nd$ and the probability density function for the rare event of reverse motion is given as follows:

$$P_b(x, t) \propto \exp\left(-\frac{(x - L - \mu t)^2}{4Dt}\right) \quad (4)$$

In a report measuring the fluctuational movement of RNAP on the template DNA, the ratio of forward to backward fluctuations was found to be approximately 1.3 [18]. In Equations (3) and (4), which models movements akin to Brownian motion, we conducted simulations by varying the drift velocity, μ , and the diffusion coefficient, D , to achieve a forward-to-backward fluctuation ratio of 1.3, as shown in Figure 2.

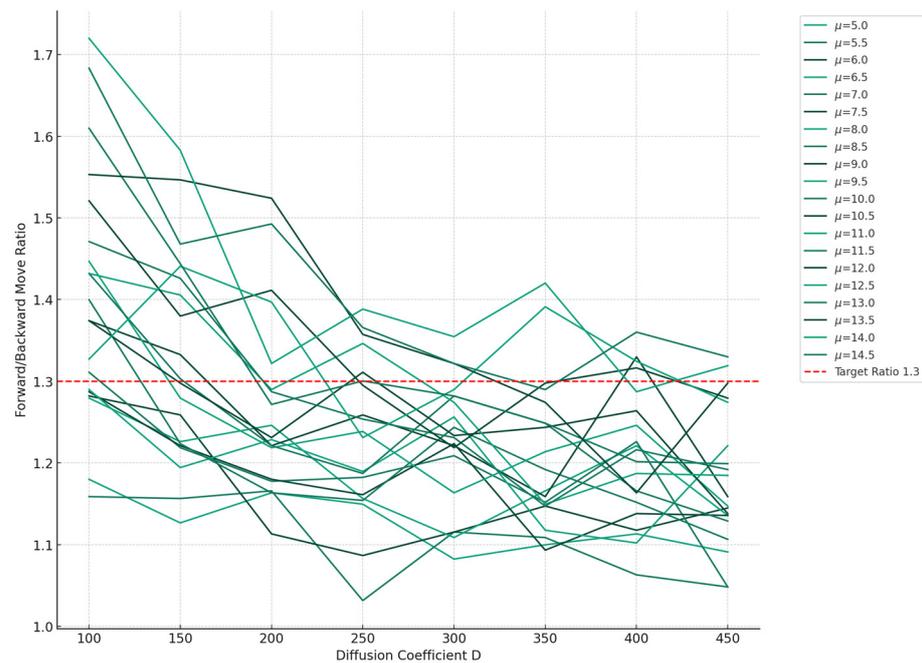


Figure 2. Plot of the ratio of forward to backward movement of RNAP against the diffusion coefficient. The drift velocity was varied in increments of 0.5, ranging from 5.0 to 14.5. The red dashed line represents the ratio of forward-to-backward movement = 1.3, based on experimental data. The closest approximation to a movement ratio of 1.3 was observed at a drift velocity (μ) of 13.0 and a diffusion coefficient (D) of 300, where the movement ratio was approximately 1.30.

By comparing plots from an experimental study [18] (see Figure 2) and simulation (Figure 3), it was found that in Equation (3), μ/D is approximately $1/30$, with its logarithmic value being 0.03.

The low ratio indicates that the thermodynamic force promoting motion is small, suggesting that random fluctuations dominate the movement. Thus, the thermodynamic driving force is not optimized for precise transcription. Consequently, considering the contribution of drift from mutual information derived from the nucleotide sequence of the template DNA implies that the information from the sequence acts as a “guide” for the movement direction of RNAP, reducing fluctuations in movement and directing RNAP to proceed forward during the reading of the DNA template, thereby instructing more accurate transcription by RNAP. As a result, the fidelity of transcription is enhanced.

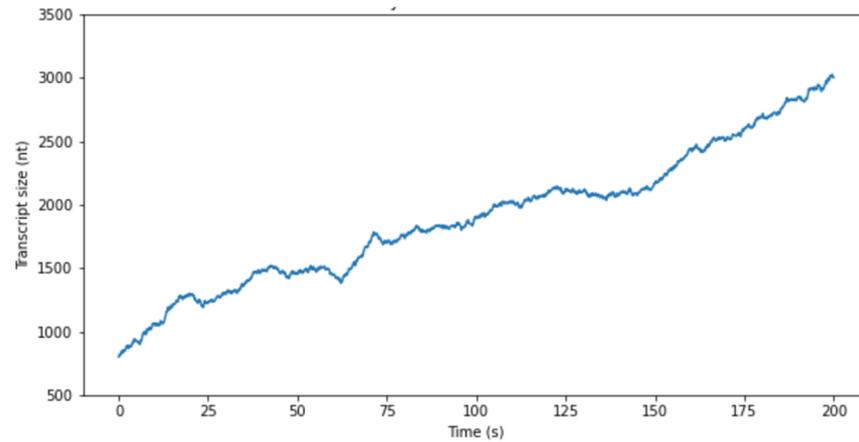


Figure 3. A plot of RNAP movement on the template T7 promoter of the λ phage. The model was predicated on a stochastic random walk paradigm. The abscissa delineates the duration of movement, whereas the ordinate reflects the quantum of base pairs traversed along the template DNA. Computations for the plot were underpinned by previously documented experimental outcomes, wherein RNA Polymerase (RNAP) was observed to traverse approximately 3 kB of DNA within a span of 200 s [18]. Within the framework of this simulation, the ratio of μ to D was established at 13/300.

To calculate the forces involved in motion, we use the probability density functions based on Equations (3) and (4), define the transition probabilities as R_f and R_b according to the fluctuation theorem, assume that the basic interval d is approximately 0.39 nm, the transcription position at m th bases with $x = md$ ($1 \leq m \leq N$), and, further, taking into account experimental data that the ratio of forward to backward motions is about 1.3, $\log R_f/R_b \sim 0.8$, and the entropy change involved in motion can be calculated as follows, by taking the logarithm of the ratio of probability density functions according to the fluctuation theorem:

$$\Delta S(x, t) = \log \frac{R_f(x, t)P_f(x, t)}{R_b(x, t)P_b(x, t)} \tag{5}$$

Substituting Equations (3) and (4) into Equation (5) and taking the limit as t approaches infinity, we have:

$$\Delta S(x, t) = k_B \left(\frac{\mu L}{2D} + \log \frac{R_f(x, t)}{R_b(x, t)} \right) \tag{6}$$

And we introduced the entropy change per unit length of the template DNA:

$$\Delta s(x, t) = k_B \left(\frac{\mu}{2D} + L^{-1} \log \frac{R_f(x, t)}{R_b(x, t)} \right) \tag{7}$$

In the steady state, the first term on the right-hand side is determined to be constant, whereas the second term accounts for the effects of fluctuations. The second term on the right-hand side corresponds to the heat absorbed by the heat bath in contact with the system when the RNAP moves from m to $m + 1$, provided that the system is in or not far from equilibrium, as described by the local detailed balance formula. In this model, the movement along the template DNA is divided by the distance d between dNTPs, assuming that RNAP reads m nucleotides, reaching $x = md$. We can define the free energy change required for the transition from per L :

$$\Delta g_m = -k_B T \Delta s(m) \tag{8}$$

In the above, $\Delta s(m) = \Delta s(x, t)$. Taking the average of the left hand, the equation can be expressed as:

$$\Delta G_m = \langle \Delta g_m \rangle = -k_B T \left\langle \frac{\mu}{2D} + L^{-1} \log \frac{R_f(m)}{R_b(m)} \right\rangle \tag{9}$$

In the above, $R_f(x, t) = R_f(m)$ and $R_b(x, t) = R_b(m)$. From Equation (9),

$$\langle \Delta \Delta G_m \rangle = -k_B T \left(\left\langle \frac{\mu}{2D} + L^{-1} \log \frac{R_f(m)}{R_b(m)} \right\rangle - \left\langle \frac{\mu}{2D} + L^{-1} \log \frac{R_f(m+1)}{R_b(m+1)} \right\rangle \right) \leq -k_B T \left\langle L^{-1} \log \frac{R_f(m)R_b(m+1)}{R_f(m+1)R_b(m)} \right\rangle < 0 \tag{10}$$

In the equation above, we assume that the movement of RNAP does not progress as m increases, and we used equations $R_f(m) > R_f(m+1)$ and $R_b(m+1) > R_b(m)$. This Equation (10), under the assumption that the transition probability $R(m)$ is independent of m , can explain previous reports suggesting that $\langle \Delta \Delta G \rangle$ approaches a value close to zero. However, it cannot account for reports that indicate a positive value for $\langle \Delta \Delta G \rangle$ [17–19].

2.2. Information Thermodynamics of Gaining Mutual Information

In this study, we propose a model wherein RNAP utilizes the base sequence information from the template DNA as mutual information for its translocation. The mutual information arises as RNAP, positioned between $(m - 1)d$ and md , acquires information about whether the incoming rNTPs can hybridize with dNTPs at position m , also denoting the degree of polymerization of dNTPs at the N position. If an rNTP is capable of hybridizing, it is incorporated into the transcription complex; otherwise, it is not incorporated (refer to Figure 1A). For instance, in the transcription initiation from the extensively studied T7 promoter DNA sequence, the probability of the appearance of dGTP is estimated to be 0.22 and the probability that the corresponding rCTP is incorporated is also estimated to be 0.22 (refer to Table 1). The step where RNAP identifies the type of rNTP is akin to a selection measurement where one out of the four types of rNTPs is chosen, reminiscent of Maxwell’s demon performing a binary measurement of particle velocities exceeding a specific threshold [25]. Here, P_m denotes the probability of the rNTP to be selected, when RNAP is at position $m = \text{dATP, dUTP, dCTP, or dGTP}$. In this case, the mutual information is given by:

$$I_m = -k_B \sum_m P_m \log P_m \tag{11}$$

$$i_m = -k_B \log P_m \tag{12}$$

In the information thermodynamic framework, the RNAP translocation step is considered as the feedback step from the dNTP in the template DNA (Figure 1D). The mutual information feedback is defined as follows:

$$I_m = -k_B \sum_m P'_m \log P'_m \tag{13}$$

$$i_m = -k_B \log P'_m \tag{14}$$

Here, P'_m denotes the probability that the transcribed dNTP to be positioned at m when RNAP is fed back. The stochastic entropy production in the entire system can be written as:

$$s^{tot}_m = i_m - i'_m + \Delta s(m) \tag{15}$$

and summing for total step m . Therefore, when considering the differential in entropy, it follows from the entropy production that $\Delta S^{tot}_m \geq 0$, we have [5,26]:

$$\Delta s(m) \geq -(I_m - I'_m) \tag{16}$$

Table 1. Frequency of each mononucleotide in the T7 promoter of the λ phage. $1 \leq m \leq N$. If thymine (T) is replaced with uracil (U), the frequency of the occurrence of nucleotides would be the same for both DNA and RNA.

	G	A	T(U)	C	Total
Probability; $P(m)$	0.22	0.38	0.16	0.25	1.00
$-P(m) \log P(m)$	0.33	0.37	0.29	0.35	1.34
$-\log P(m)$	1.51	0.96	1.83	1.38	1.42

2.3. Free Energy Change during RNAP Translocation

Finally, we consider the upper limit of the movement work performed on RNAP. The change in ensemble free energy, represented as the average of $\Delta G(m + 1) - \Delta G(m)$, is denoted as $\langle \Delta \Delta G_m \rangle$. Let w_m represent the work done to DNA at $x(t)$ by RNAP during the feedback step of the mutual information. Following Equation (16), $\langle w_m - \Delta \Delta G_m \rangle \beta \geq \langle s_m \rangle - I'_m \geq - (I_m - I'_m) - I'_m = -I_m$ is established. If the work done to RNAP by DNA is $w_m^{ext} = -w_m$, where w_m^{ext} represents the external work to RNAP by DNA, then:

$$w_m^{ext} \leq -\langle \Delta \Delta G_m \rangle + k_B T I_m \tag{17}$$

An essential point to note is that, during each cycle of reading a single base pair, the only change in the transcription complex is the movement of RNAP along the template DNA by one base pair, with no other changes occurring within the transcription system. Consequently, it can be stated that the change in free energy due to factors other than RNAP's movement is essentially zero. Thus, the energy required for RNAP to move must be supplied by some means. Our contention is that the mutual information derived from the base sequence of the template DNA should serve as the source of energy for RNAP's movement. This theory relies profoundly on the fact that DNA is structured as a repeating double helix and models the movement of RNAP almost as a Markov process. This implies that the type of base read by RNAP does not affect the subsequent reading process.

This conceptualization of free energy accounts for the variations in free energy values reported in experimental studies [18,19,21]. The mutual information per nucleotide read by RNAP from the DNA template is calculated as $I \approx -k_B T \log (1/4) \approx 1.4 k_B T$. Subsequently, we introduce the conversion efficiency parameter, α , which quantifies the proportion of the information gained converted into the mechanical work of forward translocation. The value of α ranges between 0 and 1, where $\alpha = 1$ signifies the theoretical maximum efficiency, denoting the complete conversion of information into mechanical work. The work required for the movement of RNAP, considering the effects of mutual information, can be modeled as follows:

$$w_m^{ext} = -\langle \Delta \Delta G_m \rangle + \alpha_m k_B T I_m \tag{18}$$

Taking into account the efficiency of mutual information, I , on the movement of RNAP, and based on experimental data, the left side is found to be $1.38 k_B T$ [17–20], thus estimating α to be approximately 0.84. This indicates a high efficiency in converting mutual information into work, implying that the utilization of mutual information in the movement of RNA polymerase is highly effective, underscoring the importance of the role of information in the transcription process. The coefficient α_m is influenced by various environmental factors, making theoretical estimation challenging. Therefore, it should be determined experimentally.

Here, if we denote the entire right side of Equation (18) as $-\langle \Delta \Delta G_m' \rangle$, then $\langle \Delta \Delta G_m' \rangle - \langle \Delta \Delta G_m \rangle$ becomes equal to $-k_B T I$. In this way, by considering the free energy based on $\Delta \Delta G_m'$ and taking into account the mutual information I obtained by RNAP from the template DNA, it can be said that the free energy increases by $k_B T I$.

2.4. A Proposal for the Information Thermodynamics of Nanopore Sequencing

Employing the principles of information thermodynamics to technologies like nanopore sequencing could improve the accuracy of sequence identification and enrich our comprehension of the process [27]. The data analysis involved in nanopore sequencing necessitates the discernment of valuable information amidst substantial noise. Consequently, a better grasp of the interplay between the quantity of information and the associated noise, guided by the principles of information thermodynamics, has the potential to refine data analysis algorithms.

The configuration of the motor protein and the nanopore located on the membrane within the sequencing apparatus is depicted in Figure 4 [23,24,27,28]. An ionic current traverses the nanopore due to a consistent voltage applied across the membrane, generating a positive charge on the trans side. A motor protein that has helicase activity facilitates the initial unwinding of a double-stranded DNA molecule (or an RNA–DNA hybrid duplex), a critical precursor step, since the nanopore is only capable of accommodating single-stranded DNA. After this unwinding, a single-stranded DNA or RNA, which bears a negative charge, is progressively propelled through the nanopore by the applied voltage. As nucleotides pass through the nanopore, they induce a specific alteration in the current, which enables the determination of the individual nucleotide type at an approximate rate of 450 bases per second [23,29–31]. This process corresponds to the drift velocity, μ .

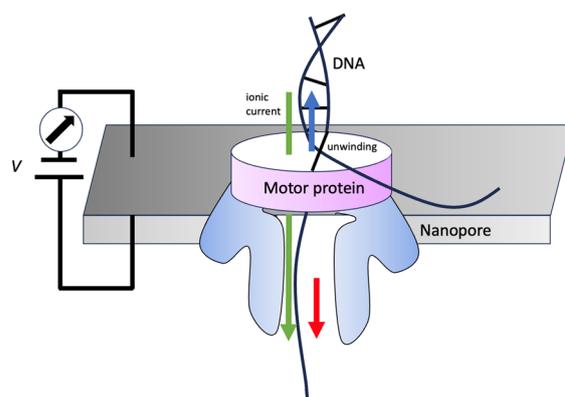


Figure 4. The motor protein helicase in nanopore sequencing. The upward blue arrow indicates the relative movement of the motor protein along the DNA, while the downward red arrow represents the actual movement of the DNA through the nanopore. A downward green arrow represents ionic current.

It is crucial to emphasize that the movement mechanism to the nanopore reflects the relative motion between the DNA motor protein and the DNA itself. This is a fundamental aspect observed in both the RNAP transcription system and the nanopore sequencing system, wherein the regulated passage of nucleic acids and the sequential identification of nucleotide bases are vital operations. These systems exemplify the sophisticated strategies employed to manipulate biological molecules [30]. Significantly, after the recognition of a single nucleotide, these proteins are reset to their pre-recognition state.

A comprehensive account is necessary to develop the information thermodynamics for nanopore sequencing systems, which encompasses electrochemical effects [21]. This account must detail the variations in the directional movement of nanopore motor molecules along DNA. In this nanopore motor molecule, the relative motion of the molecular motor is directed forward due to the stable potential across the membrane, allowing for the omission of fluctuations in the forward and backward directions. Consequently, the term including μ/D in Equation (6) can be omitted, as it can be approximated as follows:

$$\Delta S \approx k_B \log \frac{R_f(m)}{R_b(m)} \quad (19)$$

$$\Delta s = k_B L^{-1} \log \frac{R_f(m)}{R_b(m)} \quad (20)$$

For the sake of simplicity in the following discussion, we will ignore the dependency of the position m on the template DNA in nanopore molecular motors. Furthermore, from Equation (11), $\langle \Delta \Delta G \rangle$ is given as:

$$\langle \Delta \Delta G \rangle \approx -k_B T \left\langle L^{-1} \log \frac{R_f(m)R_b(m+1)}{R_f(m+1)R_b(m)} \right\rangle \quad (21)$$

The external work performed on the motor molecule, indicated by the electrochemical work term, eV , using the logarithm of ϕ that represents the electrochemical factors to the Gibbs' free energy [32], is given as follows:

$$eV = -\langle \Delta \Delta G \rangle + \alpha_1 k_B T I + k_B T \log \phi \quad (22)$$

α_1 denotes a conversion parameter. For simplicity, the mutual information and the electrochemical factor ϕ are assumed to be independent of m . Experimental observations reveal that when a load voltage of 180 mV is applied during nanopore sequencing, it corresponds to a reading error rate of approximately 38%, as supported by references [27,30,32–34]. Consequently, Equation (22) determines the coefficient α_1 as $1 - 0.38 = 0.62$. At this membrane potential load, a single nucleotide is observed to simultaneously translocate through the nanopore. Under a higher load voltage of 300 mV [32], potentially leading to reducing the fluctuation in movement independent of m , i.e., $R_f(m) \approx R_f(m+1)$ and $R_b(m) \approx R_b(m+1)$. This results in $\langle \Delta \Delta G \rangle \approx 0$ in Equation (21). (See Equation (10)). When the potential V is set to 180 mV or 300 mV, the values of $eV/k_B T$ are calculated to be 7.0 ($=v_1$) or 11.7 ($=v_2$), respectively [21]. Equation (22) can be rewritten as:

$$v_1 k_B T = \alpha_1 k_B T I + k_B T \log \phi \quad (23)$$

At a higher applied load voltage of 300 mV, it is observed that two nucleobases can pass through the pore. This underscores the significant impact of voltage conditions on the accuracy of nucleobase translocation, when the conversion parameter is set to α_2 , as demonstrated in the equation below:

$$v_2 k_B T = \alpha_2 k_B T I_m + k_B T \log \phi' \quad (24)$$

From Equations (23) and (24), the following result is yielded:

$$(v_2 - v_1) k_B T = k_B T (\alpha_2 - \alpha_1) I - k_B T \log \phi' / \phi \quad (25)$$

For example, we can obtain the ratio ϕ' / ϕ by substituting $v_1 = 7.0$, $v_2 = 11.0$, $I \approx -\log(1/4) \approx 1.4$, $\alpha_1 = 0.62$, and $\alpha_2 = 1.0$ into Equation (25):

$$\phi' / \phi \sim e^{-4.2} \quad (26)$$

As a result, the environmental factor decreased by a factor of $e^{-4.2} = 0.014$. This suggests that the application of a 300 mV voltage is effective in sufficiently reducing environmental factor.

Next, it has been reported that the introduction of polyethylene glycol (PEG) into the reaction system can significantly reduce reading errors by crowding effects [33]. To delineate this phenomenon, we first describe the relationship between work, free energy, and mutual information from Equation (22), as follows:

$$v_1 k_B T = \alpha_3 k_B T I + k_B T \log \phi'' \quad (27)$$

Herein, the addition of PEG to the measurement system results in the environmental factor being amended from ϕ to ϕ'' . From Equations (23) and (27), we derive:

$$0 \sim (\alpha_3 - \alpha_1) k_B T I + k_B T \log \frac{\phi''}{\phi} \quad (28)$$

And we obtain Equation (29) by substituting $\alpha_1 = 0.62$, $\alpha_3 = 1.0$, and $I = 1.4$ into Equation (28):

$$-0.53 k_B T \approx k_B T \log \frac{\phi''}{\phi} \quad (29)$$

This ratio indicates that the addition of PEG reduces the influence of environmental factors or noise by a factor of ϕ''/ϕ , which is equivalent to $e^{-0.53} \approx 0.59$. Thus, the method based on information thermodynamics presented herein enables a quantitative estimation of crowding effect that affects the motion of motor proteins on template DNA. In conclusion, the fundamental equation of Nanopore sequencing is given by Equation (30):

$$eV = \alpha k_B T I + k_B T \log \phi \quad (30)$$

3. Discussion

This article presented information on the thermodynamics of transcription and nanopore sequencing systems. The informational thermodynamics of RNA polymerase (RNAP) enabled the conceptualization of the transcriptional complex as a so-called Szilard Engine model, capable of converting information into work [5,33,34]. Various experiments have supported the conversion of mutual information into physical work [35,36]. For example, in a specific experimental setup, periodic electron potentials were observed to interact with Brownian particles. The introduction of an appropriate feedback controller into the system prevented the particles from falling back, thereby enabling their transition to higher potential levels [36]. As a result, mutual information could be converted into physical work [36–40].

In contrast to the spontaneous release of energy through the hydrolysis of high-energy phosphate bonds during the formation or elongation reactions between ribonucleotides and deoxyribonucleotides, experimental observations have demonstrated that the movements of RNA polymerase (RNAP) and nanopore motor proteins are governed by stochastic processes. Therefore, the concept of utilizing nucleic acid sequence information to diminish fluctuations in forward motion holds significant implications. The acquisition of information from nucleic acid sequences can reduce the randomness associated with the movements of RNAP, thereby decreasing unnecessary backtracking and the incorporation of incorrect bases during transcription, which, in turn, enhances transcription efficiency. This reduction in energy wastage has the potential to improve the overall efficiency of biological systems.

In the domain of information thermodynamics, the mechanical work derived from “information acquisition” is proportional to the amount of information. The work derived using the mutual information can be expressed as $k_B T I$ and the efficiency of this utilization has been denoted by α . The further advancement and refinement of this theory would greatly benefit from detailed and quantitative experimental validation. Specifically, exploring the interactions between enzyme kinetics and nucleotide recognition within nanopores is expected to reveal new insights into the fidelity and efficiency of sequencing.

However, it is important to note that the discussion presented here is not universally applicable to all molecular motors. This discourse is not about perpetual machines that can indefinitely extract mutual information from a thermal bath and convert it into work. Instead, it emphasizes that, in the cases of RNAP and nanopore molecular motors, the focus is on their interaction with evolutionarily refined DNA as the carrier of information and their exposure to the informational chemical bath constituted by deoxyribonucleotides. Furthermore, as mentioned earlier and at the risk of being repetitive, it is essential to note that DNA possesses repetitive structures and with each reading of a single nucleotide

sequence by RNAP or nanopore molecules, the thermodynamic state is reset, rendering the system Markovian. This aspect is a crucial condition for the modeling presented herein.

Conclusion: Intracellular molecular motors moving along nucleic acid DNA possess the ability to acquire mutual information from the base sequence and this may be involved in the motor's own movement. The mechanism by which information is efficiently converted into mechanical work could be a characteristic of biological systems. The model introduced here contributes to the development of methods to mitigate sequencing leader errors, considering factors such as drift speed, diffusion coefficient, and environmental factors in nanopore sequencing, as well as transcription by RNA polymerase.

Funding: This study was supported 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).

Institutional Review Board Statement: Not Applicable.

Data Availability Statement: All datasets used in the present study are listed in the references.

Acknowledgments: I would like to express my gratitude to Kenichi Yoshikawa of Kyoto University and Doshisha University for their valuable advice.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

RNAP	RNA polymerase
m	Position of the dNTP in the DNA ($1 \leq m \leq N$)
L	The length of the template DNA, equal to Nd
$P(x,t)$	Probability that RNAP is located at any time, t , and position, x
$P_f(x,t)$	Probability distribution of RNAP moving forward at any time, t , and position, x
$P_b(x,t)$	Probability distribution of RNAP moving backward at any time, t , and position, x
D	The diffusion coefficient of RNAP movement
μ	Drift velocity of RNAP movement
R_f	Transitional probability in RNAP movement in the forward orientation
R_b	Transitional probability in RNAP movement in the backward orientation
I_m	Mutual information gained by RNAP from dNTP at m
i_m	Stochastic mutual information gained by RNAP from dNTP at m
S^{tot}_m	Total entropy production in the entire system
s^{tot}_m	Total stochastic entropy production in the entire system
$\Delta S(m)$	The entropy produced in the RNAP movement from m to $m + 1$ position per length of the template DNA, L
I'_m	Mutual information feedback by dNTP at m
i'_m	Stochastic mutual information feedback by dNTP at m
w^{ext}	Work done to RNAP at m by DNA
ΔG_m	Gibbs' free energy change before and after RNAP movement
ΔS	The entropy produced in the nanopore motor protein movement from m to $m + 1$ position
Δs	The entropy produced in the nanopore motor protein movement from m to $m + 1$ position per length of the template DNA, L
e	Electron charge
V	Voltage across the membrane in nanopore sequencing
ϕ	Electrochemical factors of the entropy
k_B	Boltzmann constant
dNTP	Deoxyribonucleotide
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
rNTP	Ribonucleoside Triphosphates
rATP	Ribonucleoside Triphosphate Adenine

rCTP	Ribonucleoside Triphosphate Cytidine
rGTP	Ribonucleoside Triphosphate Guanine
rUTP	Ribonucleoside Triphosphate Uracil

References

- Berezhkovskii, A.M.; Bezrukov, S.M. Counting translocations of strongly repelling particles through single channels: Fluctuation theorem for membrane transport. *Phys. Rev. Lett.* **2008**, *100*, 038104. [[CrossRef](#)] [[PubMed](#)]
- Hasegawa, Y. Quantum Thermodynamic Uncertainty Relation for Continuous Measurement. *Phys. Rev. Lett.* **2020**, *125*, 050601. [[CrossRef](#)] [[PubMed](#)]
- Van Vu, T.; Hasegawa, Y. Geometrical Bounds of the Irreversibility in Markovian Systems. *Phys. Rev. Lett.* **2021**, *126*, 010601. [[CrossRef](#)] [[PubMed](#)]
- Sagawa, T.; Kikuchi, Y.; Inoue, Y.; Takahashi, H.; Muraoka, T.; Kinbara, K.; Ishijima, A.; Fukuoka, H. Single-cell *E. coli* response to an instantaneously applied chemotactic signal. *Biophys. J.* **2014**, *107*, 730–739. [[CrossRef](#)] [[PubMed](#)]
- Ito, S.; Sagawa, T. Maxwell’s demon in biochemical signal transduction with feedback loop. *Nat. Commun.* **2015**, *6*, 7498. [[CrossRef](#)] [[PubMed](#)]
- Andrieux, D.; Gaspard, P. Fluctuation theorems and the nonequilibrium thermodynamics of molecular motors. *Phys. Rev. E.* **2006**, *74 Pt 1*, 011906. [[CrossRef](#)] [[PubMed](#)]
- Hayashi, K.; Ueno, H.; Iino, R.; Noji, H. Fluctuation theorem applied to F1-ATPase. *Phys. Rev. Lett.* **2010**, *104*, 218103. [[CrossRef](#)] [[PubMed](#)]
- Lau, A.W.; Lacoste, D.; Mallick, K. Nonequilibrium fluctuations and mechanochemical couplings of a molecular motor. *Phys. Rev. Lett.* **2007**, *99*, 158102. [[CrossRef](#)] [[PubMed](#)]
- McMillan, D.G.; Watanabe, R.; Ueno, H.; Cook, G.M.; Noji, H. Biophysical Characterization of a Thermoalkaliphilic Molecular Motor with a High Stepping Torque Gives Insight into Evolutionary ATP Synthase Adaptation. *J. Biol. Chem.* **2016**, *291*, 23965–23977. [[CrossRef](#)]
- Seifert, U. Stochastic thermodynamics of single enzymes and molecular motors. *Eur. Phys. J. E Soft Matter* **2011**, *34*, 26. [[CrossRef](#)]
- Crooks, G.E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E.* **1999**, *60*, 2721–2726. [[CrossRef](#)]
- Gaspard, P. Fluctuation theorem for nonequilibrium reactions. *J. Chem. Phys.* **2004**, *120*, 8898–8905. [[CrossRef](#)] [[PubMed](#)]
- Sagawa, T.; Ueda, M. Fluctuation theorem with information exchange: Role of correlations in stochastic thermodynamics. *Phys. Rev. Lett.* **2012**, *109*, 180602. [[CrossRef](#)] [[PubMed](#)]
- Verley, G.; Lacoste, D. Fluctuation theorems and inequalities generalizing the second law of thermodynamics out of equilibrium. *Phys. Rev. E.* **2012**, *86 Pt 1*, 051127. [[CrossRef](#)] [[PubMed](#)]
- Minetti, C.A.; Remeta, D.P.; Miller, H.; Gelfand, C.A.; Plum, G.E.; Grollman, A.P.; Breslauer, K.J. The thermodynamics of template-directed DNA synthesis: Base insertion and extension enthalpies. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 14719–14724. [[CrossRef](#)]
- Dickson, K.S.; Burns, C.M.; Richardson, J.P. Determination of the free-energy change for repair of a DNA phosphodiester bond. *J. Biol. Chem.* **2000**, *275*, 15828–15831. [[CrossRef](#)] [[PubMed](#)]
- Thomen, P.; Lopez, P.J.; Bockelmann, U.; Guillerez, J.; Dreyfus, M.; Heslot, F. T7 RNA polymerase studied by force measurements varying cofactor concentration. *Biophys. J.* **2008**, *95*, 2423–2433. [[CrossRef](#)]
- Wang, M.D.; Schnitzer, M.J.; Yin, H.; Landick, R.; Gelles, J.; Block, S.M. Force and velocity measured for single molecules of RNA polymerase. *Science* **1998**, *282*, 902–907. [[CrossRef](#)] [[PubMed](#)]
- Guo, Q.; Sousa, R. Translocation by T7 RNA polymerase: A sensitively poised Brownian ratchet. *J. Mol. Biol.* **2006**, *358*, 241–254. [[CrossRef](#)]
- Bai, L.; Fulbright, R.M.; Wang, M.D. Mechanochemical kinetics of transcription elongation. *Phys. Rev. Lett.* **2007**, *98*, 068103. [[CrossRef](#)]
- Tsuruyama, T. RNA polymerase is a unique Maxwell’s demon that converts its transcribing genetic information to free energy for its movement. *Eur. Phys. J. Plus* **2023**, *138*, 604. [[CrossRef](#)]
- Da, L.T.; Chao, E.; Shuai, Y.; Wu, S.; Su, X.D.; Yu, J. T7 RNA polymerase translocation is facilitated by a helix opening on the fingers domain that may also prevent backtracking. *Nucleic Acids Res.* **2017**, *45*, 7909–7921. [[CrossRef](#)] [[PubMed](#)]
- Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [[CrossRef](#)]
- Deamer, D.W.; Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **2002**, *35*, 817–825. [[CrossRef](#)] [[PubMed](#)]
- Szilard, L. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behav. Sci.* **1964**, *9*, 301–310. [[CrossRef](#)] [[PubMed](#)]
- Sagawa, T.; Ueda, M. Second law of thermodynamics with discrete quantum feedback control. *Phys. Rev. Lett.* **2008**, *100*, 080403. [[CrossRef](#)]
- Mathé, J.; Visram, H.; Viasnoff, V.; Rabin, Y.; Meller, A. Nanopore unzipping of individual DNA hairpin molecules. *Biophys. J.* **2004**, *87*, 3205–3212. [[CrossRef](#)]

28. Ling, X.S. DNA sequencing using nanopores and kinetic proofreading. *Quant. Biol.* **2020**, *8*, 187–194. [[CrossRef](#)]
29. Angevine, C.E.; Robertson, J.W.F.; Dass, A.; Reiner, J.E. Laser-based temperature control to study the roles of entropy and enthalpy in polymer-nanopore interactions. *Sci. Adv.* **2021**, *7*, eabf5462. [[CrossRef](#)]
30. Chen, Y.; Nie, F.; Xie, S.Q.; Zheng, Y.F.; Dai, Q.; Bray, T.; Wang, Y.X.; Xing, J.F.; Huang, Z.J.; Wang, D.P.; et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* **2021**, *12*, 60. [[CrossRef](#)]
31. Craig, J.M.; Laszlo, A.H.; Brinkerhoff, H.; Derrington, I.M.; Noakes, M.T.; Nova, I.C.; Tickman, B.I.; Doering, K.; de Leeuw, N.F.; Gundlach, J.H. Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11932–11937. [[CrossRef](#)] [[PubMed](#)]
32. Ben McNally, M.W.; Meller, A. Electro-mechanical unzipping of individual DNA molecules using synthetic sub-2 nm pores. *Nano Lett.* **2008**, *8*, 3418–3422. [[CrossRef](#)] [[PubMed](#)]
33. Sagawa, T.; Ueda, M. Nonequilibrium thermodynamics of feedback control. *Phys. Rev. E* **2012**, *85 Pt 1*, 021104. [[CrossRef](#)] [[PubMed](#)]
34. Szilárd, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z. für Phys.* **1929**, *53*, 840–856. [[CrossRef](#)]
35. Hoang, T.M.; Pan, R.; Ahn, J.; Bang, J.; Quan, H.T.; Li, T. Experimental Test of the Differential Fluctuation Theorem and a Generalized Jarzynski Equality for Arbitrary Initial States. *Phys. Rev. Lett.* **2018**, *120*, 080602. [[CrossRef](#)]
36. Toyabe, S.; Sagawa, T.; Ueda, M.; Muneyuki, E.; Sano, M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **2010**, *6*, 988–992. [[CrossRef](#)]
37. Sagawa, T.; Ueda, M. Minimal energy cost for thermodynamic information processing: Measurement and information erasure. *Phys. Rev. Lett.* **2009**, *102*, 250602. [[CrossRef](#)]
38. Sagawa, T.; Ueda, M. Generalized Jarzynski equality under nonequilibrium feedback control. *Phys. Rev. Lett.* **2010**, *104*, 090602. [[CrossRef](#)]
39. Hasegawa, Y.; Van Vu, T. Fluctuation Theorem Uncertainty Relation. *Phys. Rev. Lett.* **2019**, *123*, 110602. [[CrossRef](#)]
40. Van Vu, T.; Vo, V.T.; Hasegawa, Y. Entropy production estimation with optimal current. *Phys. Rev. E* **2020**, *101*, 042138. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.