Molecular motors that digest their track to rectify Brownian motion: processive movement of exonuclease enzymes

2009 J. Phys.: Condens. Matter 21 375108

(http://iopscience.iop.org/0953-8984/21/37/375108)

J. Phys.: Condens. Matter **21** (2009) 375108 (8pp) doi:10.1088/0953-8984/21/37/375108

Molecular motors that digest their track to rectify Brownian motion: processive movement of exonuclease enzymes

Ping Xie

Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

E-mail: pxie@aphy.iphy.ac.cn

Received 19 May 2009, in final form 4 August 2009
Published 21 August 2009
Online at stacks.iop.org/JPhysCM/21/375108

Abstract

A general model is presented for the processive movement of molecular motors such as λ-exonuclease, RecJ and exonuclease I that use digestion of a DNA track to rectify Brownian motion along this track. Using this model, the translocation dynamics of these molecular motors is studied. The sequence-dependent pausing of λ-exonuclease, which results from a site-specific high affinity DNA interaction, is also studied. The theoretical results are consistent with available experimental data. Moreover, the model is used to predict the lifetime distribution and force dependence of these paused states.

1. Introduction

Motor proteins or molecular motors are active enzymes that play fundamental roles in many active processes in living cells. A processive molecular motor is referred to a single protein molecule that binds to a molecular track and takes many discrete steps along the track before dissociating. The processive motions of many motor proteins such as kinesin [1, 2], dynein [3], myosin-V [4], myosin-VI [5] and DNA helicase [6, 7] along their tracks (microtubules, actin filaments, DNA molecules) are 'powered' by the hydrolysis of NTP, typically of ATP. There are still other motor proteins that do not make use of the hydrolysis of NTP but make use of the digestion of their molecular tracks to rectify Brownian motion along the tracks. One example is the collagenase that diffuses unidirectionally along collagen fibrils [8]. Another important class of these molecular motors are DNA exonuclease enzymes such as lambda exonuclease (λ-exo) [9–19], RecJ exonuclease [20–24], exonuclease I (ExoI) [12, 25, 26], exonuclease X (ExoX) [27], etc, that translocate processively along DNA.

DNA exonuclease enzymes represent an important class of enzymes that play key roles in DNA replication, repair and recombination. Of these, λ-exo is a highly processive enzyme (with a processivity of more than 3000 base pairs) that digests one of the double-stranded DNA (dsDNA) in the 5′–3′ direction in bacteriophage λ [9–12]. The digestion requires Mg^{2+} ions, but does not require ATP or GTP, and proceeds along DNA at rates of 2–3.5 nm s^{-1} [9, 10, 13, 14]. The structure shows that it is a trimeric protein, with the three identical subunits forming a symmetrical toroid, each subunit having a nuclease active site [17]. The channel formed by λ-exo is tapered such that dsDNA may enter one side but only single-stranded DNA (ssDNA) may exit from the other side. The complete enclosure of its DNA substrate is considered to be responsible for a significant increase of its processivity [17, 28]. RecJ digests ssDNA in the 5′–3′ direction with a processivity of about 1000 bases [20–24], while ExoI digests ssDNA in the 3′–5′ direction with a processivity of more than 900 bases [12, 26, 28]. As in the case of λ-exo, the digestions by RecJ and ExoI require Mg^{2+} ions, but do not require ATP. Structural studies show that both RecJ and ExoI are monomeric and have a roughly semicircular form, with a groove through the center of molecules [22, 26]. Only one nuclease active site is present in the groove of each enzyme. The narrow groove for both RecJ and ExoI may be characteristic of the exonucleases specific for ssDNA, which provides an explanation for their high processivity [22, 26, 28].

In the literature, a lot of works have been devoted to modeling the processive motion of those molecular motors that are 'powered' by the hydrolysis of NTP [29–39]. Moreover, several models, which include the burnt bridges Brownian ratchet model [8, 40–43] and the asymmetric potential Brownian ratchet model [44], have been proposed

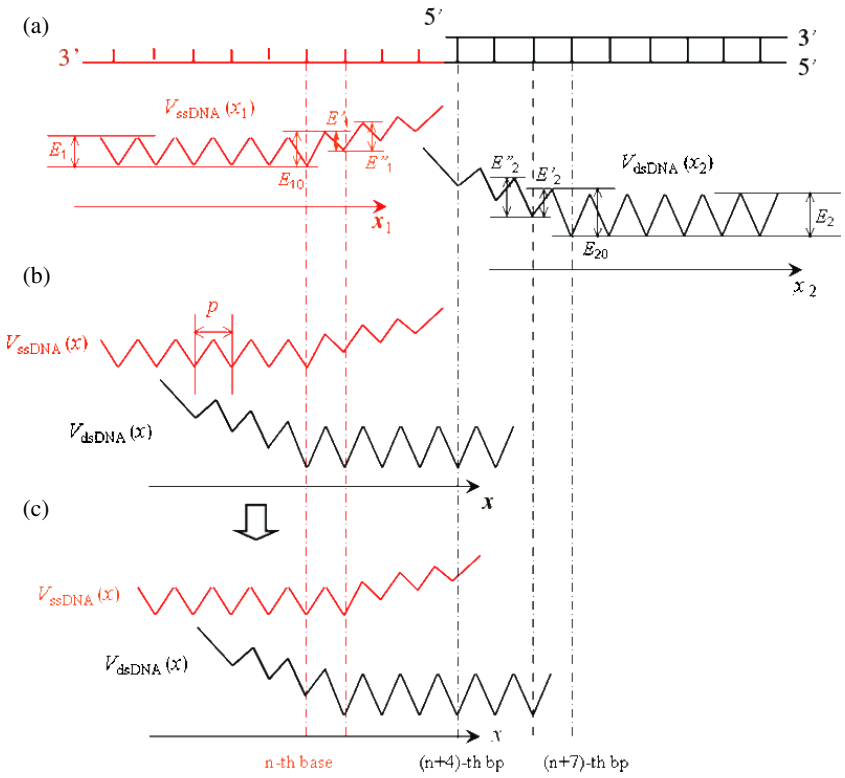

Figure 1. Interaction potentials V_{ssDNA} between λ-exo and 3′–5′ ssDNA and V_{dsDNA} between λ-exo and dsDNA (see the text for a detailed description). The base pair distance $p = 0.34$ nm.

(This figure is in colour only in the electronic version)

for the biased diffusion of collagenase along collagen fibrils via the digestion of collagen. In this work, we present a general rectified Brownian ratchet model for the processive translocation of DNA exonuclease enzymes such as λ-exo, RecJ and ExoI via the digestion of DNA.

2. Processive translocation of λ-exo

It is assumed that there exist both ssDNA-binding residues and dsDNA-binding residues in λ-exo. This assumption is similar to that adopted for nucleic acid polymerase enzymes such as DNA polymerase (DNAP) [45, 46], reverse transcriptase [47] and RNA polymerase (RNAP) [48, 49]. The presence of the interaction between λ-exo and 3′–5′ ssDNA is consistent with the experimental results by Sriprakash *et al* [15]. Moreover, the assumption of both the interaction of λ-exo with 3′–5′ ssDNA and the interaction with dsDNA are consistent with the experimental result by van Oijen *et al* [18], which showed that only a low fraction of dsDNA that lacks a free 3′ end was digested to ssDNA by the λ-exo enzyme. This is explained as follows. For dsDNA that lacks a free 3′ end, only the binding affinity E_2 between λ-exo and dsDNA is present. Thus the binding affinity of λ-exo for DNA, E_2, is smaller than the binding affinity, $E_1 + E_2$, for DNA with a free 3′ end, where E_1 is the binding affinity of λ-exo for ssDNA. This implies that the enzyme is easily detached from DNA that lacks a free 3′ tail, thus showing a low probability of digestion. Another point to note is that the lack of the binding affinity of λ-exo for the 5′ end of ssDNA, as assumed here, is consistent with the experimental results of Subramanian *et al* [14], which showed that the lower extent of digestion of 5′-OH DNA than 5′-phosphate DNA is due to a catalytic defect in the enzyme–substrate complex and is not due to an inability of λ-exo to bind the 5′-OH DNA.

Since no experimental data are available for the binding surface between λ-exo and DNA, without loss of generality, we consider that the ssDNA-binding residues cover four bases on the 3′–5′ ssDNA and the dsDNA-binding residues cover four base pairs on the dsDNA duplex. As is noted, taking other values for the binding surface has no effect on the analyses and results presented in this work. Consider the dsDNA with a 3′–5′ single-stranded tail, as schematically shown in figure 1(a) (upper diagram). Then, as the structure shows [17], the dsDNA-binding residues should be on the right side of the ssDNA-binding residues, with a zero interval, and the nuclease active site is located near the leftmost point of the dsDNA-binding residues. Thus, the interaction potential, $V_{ssDNA}(x_1)$, between the ssDNA-binding residues and the 3′–5′ ssDNA tail can be approximately shown in the middle diagram of figure 1(a), where E_1 is the binding affinity for all four bases of the ssDNA that the ssDNA-binding residues can cover, E'_1 is the binding affinity for only three bases that only part of the ssDNA-binding residues can cover and x_1 represents the coordinate of the leftmost point on the ssDNA-binding residues along the DNA. The interaction potential, $V_{dsDNA}(x_2)$, between the dsDNA-binding residues and the dsDNA can be approximately shown in the lower diagram of figure 1(a),

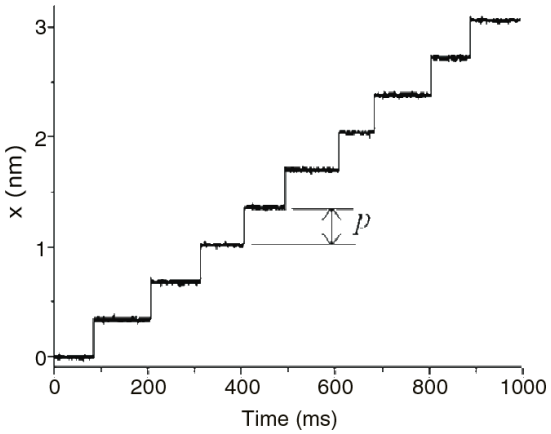

Figure 2. A typical numerical result for the displacement of λ-exo along DNA as a function of time, showing stepwise motion with the step size of one base pair. The binding affinities $E_1 + E_2' = 18k_BT$ and $E_1 + E_{20} = 25k_BT$.

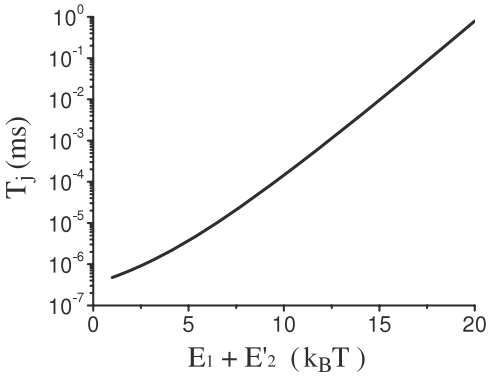

Figure 3. Results of jumping time T_j versus binding affinity $E_1 + E_2'$ for λ-exo.

where E_2 is the binding affinity for the sugar–phosphate backbones connecting all four base pairs on the dsDNA, E_2' is the binding affinity for the backbones connecting only three base pairs on the dsDNA and x_2 represents the coordinate of the rightmost point on the dsDNA-binding residues along the DNA. Note that the binding affinity E_1' for only three bases is smaller than the binding affinity E_1 for all four bases. Due to the same reason, the binding affinity E_2' is smaller than E_2.

Moreover, the interaction between the enzyme and sequence-nonspecific DNA is likely via electrostatic force. This is consistent with the structure of RecJ, since RecJ mainly interacts with ssDNA through the DNA backbone and the residues near the presumed nuclease active site are positively charged [22]. The structure of *E. coli* ExoI also showed that the surface of the groove located near the presumed nuclease active site is positively charged [26], consistent with the electrostatic interaction proposal. If we assume that the electrostatic interaction distance between the enzyme and DNA is larger than the base pair distance $p = 0.34$ nm then it follows that E_{20} and E_{10} defined in figure 1 should be larger than E_2 and E_1, respectively, while E_2'' and E_1'' should be larger than E_2' and E_1', respectively. Note that, at the sequence-specific site of DNA, besides the electrostatic interaction between λ-exo and the DNA, the enzyme can also employ hydrogen bonding interactions with the DNA as well as stacking interactions with the nucleobases of the ssDNA, giving a much larger binding affinity (see section 3).

If we represent the position of λ-exo by its center-of-mass position, x, which is drawn to be coincident with the leftmost point on the ssDNA-binding residues, the potential $V_{\rm dsDNA}(x)$ between the dsDNA-binding residues and the dsDNA is shifted towards the $-x$ direction by 7 bases (or base pairs), as shown in the lower diagram of figure 1(b). After one base, i.e. the $(n+4)$th base, on the 5′–3′ strand is digested by the nuclease active site that locates at the $(n+4)$th base, the double strand that connects the $(n+4)$th base pair becomes 3′–5′ single stranded. Thus, the interaction potential $V_{\rm dsDNA}(x)$ between the dsDNA-binding residues and the dsDNA becomes the one shown in the lower diagram of figure 1(c), while the interaction potential $V_{\rm ssDNA}(x)$ between the ssDNA-binding residues and the 3′–5′ single strand is changed to that shown in the upper diagram of figure 1(c). The mathematical forms of $V_{\rm ssDNA}(x)$ and $V_{\rm dsDNA}(x)$ shown in figures 1(b) and (c) are given in the appendix.

The movement of λ-exo along DNA in the over-damped environment can be described by the following Langevin equation:

$$\Gamma\frac{{\rm d}x}{{\rm d}t} = -\frac{\partial V(x,t)}{\partial x} + \xi(t), \tag{1}$$

where $V(x) = V_{\rm ssDNA}(x) + V_{\rm dsDNA}(x)$, with $V_{\rm ssDNA}(x)$ and $V_{\rm dsDNA}(x)$ in the forms of figure 1(b) before the digestion of base $(n+4)$ on the 5′–3′ strand and in the forms of figure 1(c) after the digestion of base $(n+4)$. $\xi(t)$ is the fluctuating Langevin force, with $\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(t')\rangle = 2k_BT\Gamma\delta(t-t')$. The drag coefficient $\Gamma = 6\pi\eta r = 9.4\times 10^{-11}$ kg s^{-1}, where the viscosity $\eta = 0.01$ g cm^{-1} s^{-1} and λ-exo is considered as a sphere with radius $r = 5$ nm.

A typical result of the displacement of λ-exo as a function of time by numerically solving equation (1) using the stochastic Runge–Kutta method, as used elsewhere [33, 34], is shown in figure 2, where we have taken $E_1 + E_2' = 18k_BT$ and the nucleotide-digestion time T_d with a mean value of 100 ms [9, 10, 13, 14]. The results in figure 2 are insensitive to the value of $E_1 + E_2'$ provided that it is not very large (see figure 3). From figure 2, it is seen that the λ-exo translocates processively along DNA, with the step size of one base pair. In the model, a forward stepping corresponds to a jumping of the Brownian particle (i.e. the λ-exo) from the potential well with a shallower depth $E_1 + E_2'$ to that with a depth E_1+E_{20} corresponding to the deepest well of potential $V(x) = V_{\rm ssDNA}(x) + V_{\rm dsDNA}(x)$ (figure 1(c)). The transition of the potential well at a given position (e.g. nth base) on the DNA from that with the deepest depth $E_1 + E_{20}$ (figure 1(b)) to that with the shallower depth $E_1 + E_2'$ (figure 1(c)) is via the digestion of one base on the 5′–3′ strand. Thus, the dwell time of a forward step is the sum of the digestion time T_d and the jumping time, T_j, over the potential barrier $E_1 + E_2'$ to the deepest potential well with depth $E_1 + E_{20}$. The digestion time T_d is determined by the chemical reaction rate [13, 14]. Here we determine the jumping time T_j.

The corresponding Fokker–Planck form of equation (1) is $\frac{\partial P(x,t)}{\partial t} = \frac{1}{\Gamma}\frac{\partial}{\partial x}[\frac{\partial V(x)}{\partial x}P(x,t)] + D\frac{\partial^2 P(x,t)}{\partial^2 x}$, where $D = k_B T/\Gamma$. From this Fokker–Planck equation the mean first-passage time for the enzyme to jump from the shallower potential well at the nth base (figure 1(c)) to the deepest potential well at the next $(n+1)$th base can be calculated by $T_j = \frac{1}{D}\int_0^p dy \exp[V(y)/(\Gamma D)]\int_0^y \exp[-V(z)/(\Gamma D)]\,dz$ [50], where the origin of the coordinate is taken at the position of the nth base in figure 1(c) and, for simplicity, the backward jumping from the potential well at the nth base to the previous shallower potential well at the $(n-1)$th base is not taken into account. By integration with $V(x) = V_{ssDNA}(x) + V_{dsDNA}(x)$ given in figure 1(c) (see also the appendix) we obtain the mean jumping time

$$T_j \approx \frac{(p\Gamma)^2 D}{4(E_1+E_2')^2}\left[\exp\left(\frac{E_1+E_2'}{\Gamma D}\right)-1\right] - \frac{p^2\Gamma}{4(E_1+E_2')} + \frac{(p\Gamma)^2 D}{4(E_1+E_2')(E_1+E_{20})}\left[\exp\left(\frac{E_1+E_2'}{\Gamma D}\right) - \exp\left(-\frac{E_{20}-E_2'}{\Gamma D}\right)\right] + \frac{p^2\Gamma}{4(E_1+E_{20})}, \quad (2)$$

where, for simplicity, we have taken $\exp(-\frac{E_1+E_2'}{\Gamma D}) \approx 0$ because $E_1+E_2' \gg \Gamma D = k_B T$. Since $E_1+E_{20} > E_1+E_2'$, the third and fourth terms are smaller than the first and second terms, respectively. As a result, for an approximation, we can neglect the third and fourth terms in equation (2) and thus we have

$$T_j \approx \frac{(p\Gamma)^2 D}{4(E_1+E_2')^2}\left[\exp\left(\frac{E_1+E_2'}{\Gamma D}\right)-1\right] - \frac{p^2\Gamma}{4(E_1+E_2')}. \quad (3)$$

Using equation (3), the calculated results of T_j versus E_1+E_2' are shown in figure 3. It is seen that, even for a very large value of $E_1+E_2' = 20k_B T$ (for example, for $E_1 = 10k_B T$ and $E_2' = 10k_B T$), T_j is smaller than 1 ms, which is much smaller than $T_d = 100$ ms. This implies that the dwell time is, in general, mainly determined by the digestion time T_d, which is consistent with the experimental data [18].

If there is an external force, F, acting on the λ-exo, then E_1+E_2' in equation (3) is replaced by $E_1+E_2'+Fp/2$, where it is defined that a backward force has a positive value. The calculated results of T_j versus F are shown in figure 4. As expected, T_j increases with the increase of F. However, T_j increases slowly with the increase of F. For example, when F is increased from 0 to 10 pN, T_j increases only by about 1.44-fold. Thus, the dwell time is still mainly determined by the digestion time T_d even under a backward load as large as 10 pN.

3. Sequence-dependent pause of λ-exo

In the above studies, for simplicity, we consider that the electrostatic interactions between λ-exo and DNA are homogeneous along the DNA, i.e. the binding affinities, E_1 and E_2, are considered to be DNA-sequence-independent. However, at the sequence-specific site of the DNA, besides the electrostatic interactions between λ-exo and the DNA,

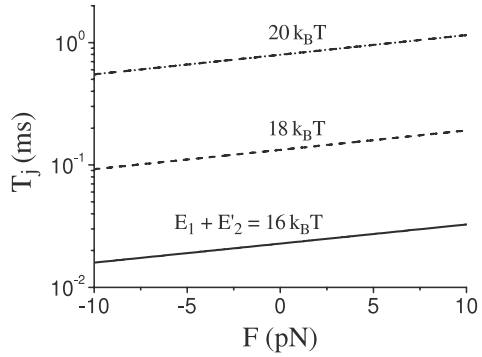

Figure 4. Results of jumping time T_j versus external force F acted on λ-exo for different values of binding affinity E_1+E_2'.

the enzyme can also employ hydrogen bonding interactions with the DNA as well as stacking interactions with the nucleobases of the ssDNA. Although direct support of this argument is not available from the crystal structure of λ-exo in the absence of DNA [17], it can be inferred from the available structure of restriction endonuclease *Bam*HI bound to nonspecific DNA [51] and the available structure bound to sequence-specific DNA [52]. Consequently, if λ-exo is positioned at a specific sequence, there exists a much larger value of E_1+E_2' than that at nonspecific sequences. That implies that the λ-exo will take a much longer time to jump over this larger barrier to the next binding position. Using equation (2) the calculated results of T_j versus E_1+E_2' for different values of E_1+E_{20} are shown in figure 5(a). It is seen that, even for E_1+E_{20} varying in a wide range (from $20k_B T$ to $30k_B T$), T_j varies in a narrow range. At $E_1+E_2' = 28.5k_B T$, $T_j \approx 4$ s, which is close to the value determined experimentally [19] and is much larger than the digestion time $T_d \approx 100$ ms. Thus, the sequence-dependent pausing time is mainly determined by the jumping time T_j, which is in contrast to the dwell time as discussed in section 2. When an external force F is acting on the λ-exo, E_1+E_2' and E_1+E_{20} in equation (2) are replaced by $E_1+E_2'+Fp/2$ and $E_1+E_{20}-Fp/2$, respectively. The calculated results of T_j versus F with $E_1+E_2' = 28.5k_B T$ and $E_1+E_{20} = 25k_B T$ are shown in figure 5(b). It is seen that the external force influences the sequence-dependent pausing time. For example, when a forward load is increased from 0 to 20 pN, the mean pausing time is reduced from 4.1 to 1.9 s; while when a backward load is increased from 0 to 20 pN, the mean pausing time is increased from 4.1 to 9.2 s.

Moreover, it is noted that different DNA sequences should have slightly different strengths of hydrogen bonding and stacking interactions with λ-exo, giving slightly different values of E_1+E_2'. To see the effect of different sequences on the pausing lifetime, we numerically solve equation (1) by using different values of E_1+E_2'. In figures 6(a)–(c) we show the calculated lifetime distributions of the pause for $E_1+E_2' = 28k_B T$, $28.5k_B T$ and $29k_B T$, respectively. It is seen that the lifetime distribution for a given E_1+E_2' has the single-exponential form. In other words, the pausing lifetime is single exponential for a given sequence. This is in agreement with the experimental result of Perkins *et al* [19]. Moreover,

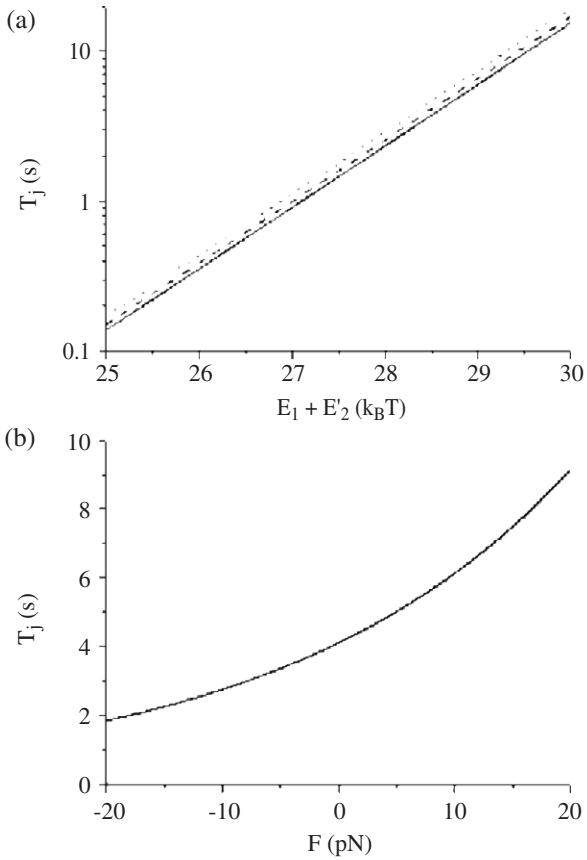

Figure 5. Results of mean pausing lifetime for λ-exo. (a) Mean pausing lifetime versus binding affinity $E_1 + E'_2$ under no external force. Dotted line is for binding affinity $E_1 + E_{20} = 20k_BT$, dashed line for binding affinity $E_1 + E_{20} = 25k_BT$ and solid line for binding affinity $E_1 + E_{20} = 30k_BT$. (b) Mean pausing lifetime versus external force F, with binding affinities $E_1 + E'_2 = 28.5k_BT$ and $E_1 + E_{20} = 25k_BT$.

as noted from equation (2), the mean pausing lifetime for the sequence with a smaller $E_1 + E'_2$ is shorter than that for the sequence with a larger $E_1 + E'_2$. In figure 6(d) we show the pausing-lifetime distribution by counting all values of lifetime for the three values of $E_1 + E'_2$, with equal counting numbers for the three values of $E_1 + E'_2$. The distribution can be well fitted by a linear sum of two exponentials with time constants of 4 and 6 s (amplitudes 70% and 30%, respectively). This statistical characteristic is similar to that for RNAP obtained by Herbert *et al* [53]. The future experiment is hoped to test this prediction for λ-exo. Similarly, based on the present model it is expected that the pausing efficiency (or probability) for the sequence with a small $E_1 + E'_2$ that gives a short mean lifetime should be smaller than that for the sequence with a large $E_1 + E'_2$ that gives a long mean lifetime. This is also consistent with the experimental results [19].

4. Processive translocation of RecJ

As both RecJ and ExoI have very similar structures [22, 26], we propose that the processive translocation of RecJ along 5′–3′ ssDNA and that of ExoI along 3′–5′ ssDNA have a similar mechanism. Here we take RecJ as the example to illustrate the mechanism.

Since RecJ only interacts with 5′–3′ ssDNA, it is assumed that there exist only the ssDNA-binding residues. As shown from the available structure [22], the ssDNA-binding residues cover four bases on 5′–3′ ssDNA. Moreover, as is shown experimentally [24], RecJ has a specific interaction with the 5′ end of ssDNA, which is different from λ-exo that has no specific interaction with the 5′ end of DNA (see section 2). Thus, when the ssDNA-binding residues of RecJ bind the four bases of 5′–3′ ssDNA that do not include the 5′ end, the binding affinity should be smaller than that when the 5′ end is included. As a result, the interaction potential, $V(x)$, between RecJ and 5′–3′ ssDNA can be approximately shown in figure 7(a), where E is the binding affinity for the four bases of the ssDNA that include the 5′ end, E' is the binding affinity for the three bases of the ssDNA that include the 5′ end, E'' is the binding affinity for the four bases of the ssDNA that do not include the 5′ end and x represents the coordinate of the rightmost point on the ssDNA-binding residues along the ssDNA. After one base, i.e. the $(n-3)$th base, is digested by the enzyme, the potential $V(x)$ becomes the one shown in figure 7(b).

Therefore, a forward stepping of RecJ corresponds to a jumping of the Brownian particle from the shallower potential well of depth E' at the nth base to the deepest potential well of depth E at the next $(n+1)$th base (figure 7(b)) and the mean jumping time can be still approximately calculated by equation (3) but with E' instead of $E_1 + E'_2$. The transition of the potential well at a given position (e.g. nth base) on the ssDNA from that with the deepest depth E (figure 7(a)) to that with the shallower depth E' (figure 7(b)) is via the digestion of the 5′-end base on the ssDNA. Similar to figure 2 for λ-exo, RecJ translocates processively along the ssDNA with the step size of one base and the results of the mean jumping time T_j versus E' are shown in figure 3, where, however, the title of the horizontal axis, $E_1 + E'_2$, is replaced by E'.

It should be noted that the ExoI and RecJ exonucleases have different DNA footprints. The ssDNA-binding residues of ExoI cover 12 bases on the 3′–5′ ssDNA [26], while the ssDNA-binding residues of RecJ cover 4 bases on the 5′–3′ ssDNA. In this work, both exonucleases are proposed to have the same translocation mechanism.

5. Discussion

It is interesting to note that the processive translocation of λ-exo along DNA is via the Brownian ratchet mechanism which is rectified by the digestion of the 5′–3′ strand of dsDNA, while the processive translocation of DNAP along the DNA template is via the Brownian ratchet mechanism which is rectified by the synthesis of a matched base complementary to the ssDNA template. Although the two enzymes have ‘opposite’ biological functions, based on the model presented here for λ-exo (see section 2) and the model presented previously for DNAP [45, 46], they show similar ratchet mechanisms for the unidirectional translocation. For the former the digestion of one base on the 5′–3′ strand induces its interaction potential with DNA, $V(x) = V_{\mathrm{ssDNA}}(x) + V_{\mathrm{dsDNA}}(x)$, changing from

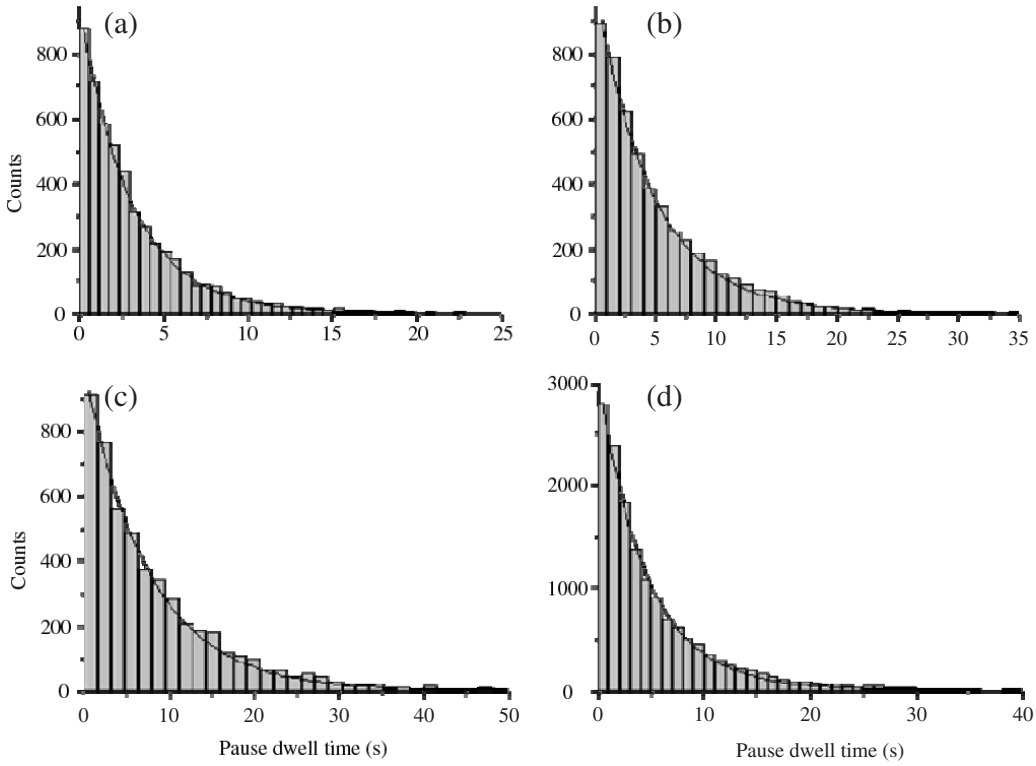

Figure 6. Calculated lifetime distributions of pausing for λ-exo, with exponential fits, for binding affinity $E_1 + E_{20} = 20k_BT$. (a)–(c) Distributions with binding affinities $E_1 + E_2' = 28k_BT$, $28.5k_BT$ and $29k_BT$, fitted with a single exponential, $C\exp(-t/\tau)$, of time constants $\tau \approx 2.9$ s, 4.7 s and 7.6 s, respectively. (d) Distribution by counting all values of lifetime for the three values of binding affinity $E_1 + E_2'$, with equal counting numbers for the three values of $E_1 + E_2'$. The distribution is fitted by a linear sum of two exponentials, $C[P_1\exp(-t/\tau_1) + P_2\exp(-t/\tau_2)]$, with time constants of $\tau_1 = 4$ and $\tau_2 = 6$ s (amplitudes $P_1 = 70\%$ and $P_2 = 30\%$, respectively).

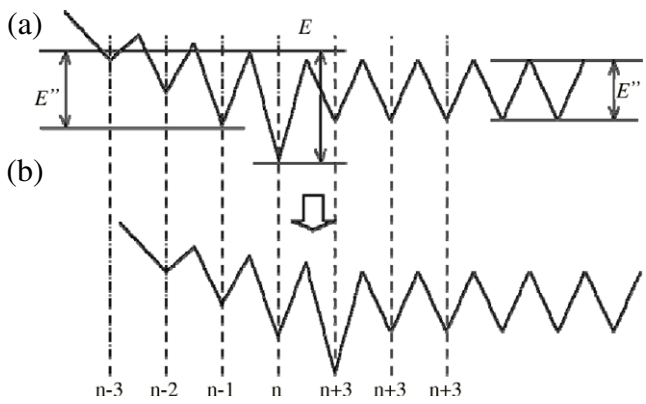

Figure 7. Interaction potential $V(x)$ between RecJ and 5′–3′ ssDNA (see the text for the detailed description). Since RecJ has an ssDNA-binding site, the interaction potential $V(x)$ between the site and ssDNA is similar to either the interaction potential $V_{\text{ssDNA}}(x)$ between the ssDNA-binding site of λ-exo and 3′–5′ ssDNA or the interaction potential $V_{\text{dsDNA}}(x)$ between the dsDNA-binding site of λ-exo and dsDNA (see figure 1). However, since RecJ has a specific interaction with the 5′ end of ssDNA, the potential $V(x)$ has a larger depth when the ssDNA-binding site of RecJ covers four bases and includes the 5′ end than that when the ssDNA-binding site covers four bases but does not include the 5′ end (comparing $V(x)$ given here with either $V_{\text{ssDNA}}(x)$ or $V_{\text{dsDNA}}(x)$ given in figure 1).

that shown in figure 1(b) to that shown in figure 1(c). Thus a forward stepping corresponds to a jumping of the Brownian particle (i.e. the λ-exo) from the potential well with shallower depth $E_1 + E_2'$ to that with the deepest depth $E_1 + E_{20}$. In contrast, for the DNAP the synthesis of a matched base complementary to the ssDNA template induces its interaction potential with DNA, $V(x) = V_{\text{ssDNA}}(x) + V_{\text{dsDNA}}(x)$, changing from that shown in figure 1(c) to that shown in figure 1(b). Thus a forward stepping corresponds to a jumping of the Brownian particle (i.e. the DNAP) from the potential with shallower depth $E_1' + E_2$ to that with deeper depth $E_{10} + E_2$.

An interesting behavior for λ-exo is the occurrence of sequence-dependent pausing during its processive translocation. The similar behavior has also been observed for RNAP [53]. It is thus expected that the similar pausing behaviors for both λ-exo and RNAP may share the same mechanism, i.e. the short pauses result from the sequence-dependent high binding affinities [48]. Moreover, different sequences should have slightly different binding affinities, resulting in different short pausing durations.

The crystal structure of λ-exo shows that it consists of three identical subunits, with each subunit having a nuclease active site [17]. Thus, in the present model, it is considered that each subunit should have an ssDNA-binding site and a dsDNA-binding site. At any time, only one of the three subunits is interacting with DNA. As proposed before [17, 28], the toroidal form of the trimer is responsible for a significant increase of its processivity. Although RecJ and ExoI are monomeric, the single polypeptide chain forms an asymmetric semicircular structure that can enclose the DNA substrate [22, 26], which can provide high processivity [28].

However, the non-complete enclosure of DNA by RecJ and ExoI can only give a lower processivity than the complete enclosure by λ-exo. Thus, it is understandable that the λ-exo has evolved a trimeric and toroidal structure to ensure a much higher processivity of more than 3000 base pairs than that of about 1000 bases in RecJ and ExoI.

It is mentioned that, in this work, we only consider the interaction of single exonuclease molecules with DNA. However, the exonuclease activity *in vivo* involves the cooperation of the exonuclease enzyme with other proteins. For instance, single-stranded DNA-binding proteins (SSBs) are known to interact directly with core-catalytic domains of RecJ and ExoI, increasing the processivity and stimulating the exonuclease activities of the enzymes [54–56]. Thus, to study the dynamics of these exonucleases *in vivo*, besides the interactions between the enzymes and DNA, as studied in this work, the interactions between the enzymes and SSBs must also be taken into account, which will be studied in the future.

In conclusion, we studied the processive translocation of λ-exo, RecJ and ExoI along DNA. We proposed that the three enzymes share the similar rectified ratchet mechanism for their processive motion, with only slight differences for different enzymes. The experimentally observed sequence-dependent pausing of λ-exo can be explained well by assuming a higher site-specific binding affinity for DNA at the specific sequence. To test the argument, it is hoped to experimentally determine the effect of external force on the mean sequence-dependent pausing lifetime (see figure 5(b)) and the statistical characteristic of the pausing-lifetime distribution (see figure 6(d)).

Appendix

Taking the origin of the coordinate at the position of the nth base, the mathematical forms of $V_{\mathrm{ssDNA}}(x)$ and $V_{\mathrm{dsDNA}}(x)$ shown in figure 1(b) in the range of $-3p/2 < x \leqslant 3p/2$ can be written as follows:

$$V_{\mathrm{ssDNA}}(x) = -\frac{2E_1}{p}x - 2E_1,$$

$$V_{\mathrm{dsDNA}}(x) = -\frac{2E_2''}{p}x + E_{20} - 2E_2'' - E_2' \quad \text{when } -\frac{3p}{2} < x \leqslant -p$$

$$V_{\mathrm{ssDNA}}(x) = \frac{2E_1}{p}x + 2E_1, \quad V_{\mathrm{dsDNA}}(x) = \frac{2E_2'}{p}x + E_{20} + E_2' \quad \text{when } -p < x \leqslant -\frac{p}{2}$$

$$V_{\mathrm{ssDNA}}(x) = -\frac{2E_1}{p}x, \qquad V_{\mathrm{dsDNA}}(x) = -\frac{2E_{20}}{p}x \quad \text{when } -\frac{p}{2} < x \leqslant 0$$

$$V_{\mathrm{ssDNA}}(x) = \frac{2E_{10}}{p}x, \qquad V_{\mathrm{dsDNA}}(x) = \frac{2E_2}{p}x \quad \text{when } 0 < x \leqslant \frac{p}{2}$$

$$V_{\mathrm{ssDNA}}(x) = -\frac{2E_1'}{p}x + E_{10} + E_1',$$

$$V_{\mathrm{dsDNA}}(x) = -\frac{2E_2}{p}x + 2E_2 \quad \text{when } \frac{p}{2} < x \leqslant p$$

$$V_{\mathrm{ssDNA}}(x) = \frac{2E_1''}{p}x + E_{10} - E_1' - 2E_1'',$$

$$V_{\mathrm{dsDNA}}(x) = \frac{2E_2}{p}x - 2E_2 \quad \text{when } p < x \leqslant \frac{3p}{2}.$$

Similarly, the mathematical forms of $V_{\mathrm{ssDNA}}(x)$ and $V_{\mathrm{dsDNA}}(x)$ shown in figure 1(c) in the range of $-p/2 < x \leqslant 5p/2$ are written as follows:

$$V_{\mathrm{ssDNA}}(x) = -\frac{2E_1}{p}x, \qquad V_{\mathrm{dsDNA}}(x) = -\frac{2E_2''}{p}x + E_{20} - E_2' \quad \text{when } -\frac{p}{2} < x \leqslant 0$$

$$V_{\mathrm{ssDNA}}(x) = \frac{2E_1}{p}x, \qquad V_{\mathrm{dsDNA}}(x) = \frac{2E_2'}{p}x + E_{20} - E_2' \quad \text{when } 0 < x \leqslant \frac{p}{2}$$

$$V_{\mathrm{ssDNA}}(x) = -\frac{2E_1}{p}x + 2E_1, \quad V_{\mathrm{dsDNA}}(x) = -\frac{2E_{20}}{p}x + 2E_{20} \quad \text{when } \frac{p}{2} < x \leqslant p$$

$$V_{\mathrm{ssDNA}}(x) = \frac{2E_{10}}{p}x - 2E_{10}, \qquad V_{\mathrm{dsDNA}}(x) = \frac{2E_2}{p}x - 2E_2 \quad \text{when } p < x \leqslant \frac{3p}{2}$$

$$V_{\mathrm{ssDNA}}(x) = -\frac{2E_1'}{p}x + E_{10} + 3E_1',$$

$$V_{\mathrm{dsDNA}}(x) = -\frac{2E_2}{p}x + 4E_2 \quad \text{when } \frac{3p}{2} < x \leqslant 2p$$

$$V_{\mathrm{ssDNA}}(x) = \frac{2E_1''}{p}x + E_{10} - E_1' - 4E_1'',$$

$$V_{\mathrm{dsDNA}}(x) = \frac{2E_2}{p}x - 4E_2 \quad \text{when } 2p < x \leqslant \frac{5p}{2}.$$

References

[1] Okada Y and Hirokawa N 2000 *Proc. Natl Acad. Sci. USA* **97** 640
[2] Yildiz A, Tomishige M, Vale R D and Selvin P R 2004 *Science* **303** 676
[3] Mallik R, Carter B C, Lex S A, King S J and Gross S P 2004 *Nature* **427** 649
[4] Mehta A D, Rock R S, Rief M, Spudich J A, Mooseker M S and Cheney R E 1999 *Nature* **400** 590
[5] Rock R S, Rice S E, Wells A L, Purcell T J, Spudich J A and Sweeney H L 2001 *Proc. Natl Acad. Sci. USA* **98** 13655
[6] Bianco P R, Brewer L R, Corzett M, Balhorn R, Yeh Y, Kowalczykowski S C and Baskin R J 2001 *Nature* **409** 374
[7] Kim D-E, Narayan M and Patel S S 2002 *J. Mol. Biol.* **321** 807
[8] Saffarian S, Collier I E, Marmer B L, Elson E L and Goldberg G 2004 *Science* **361** 108
[9] Radding C M 1966 *J. Mol. Biol.* **18** 235
[10] Little J W 1967 *J. Biol. Chem.* **242** 679
[11] Carter D M and Radding C M 1971 *J. Biol. Chem.* **246** 2502
[12] Thomas K R and Olivera B M 1978 *J. Biol. Chem.* **253** 424
[13] Mitsis P G and Kwagh J G 1999 *Nucleic Acids Res.* **27** 3057
[14] Subramanian K, Rutvisuttinunt W, Scott W and Myers R S 2003 *Nucleic Acids Res.* **31** 1585

[15] Sriprakash K S, Lundh N, Huh M M-O and Radding C M 1975 *J. Biol. Chem.* **250** 5438
[16] Takahashi N and Kobayashi I 1990 *Proc. Natl Acad. Sci. USA* **87** 2790
[17] Kovall R and Matthews B W 1997 *Science* **277** 1824
[18] van Oijen A M, Blainey P C, Crampton D J, Richardson C C, Ellenberger T and Xie X S 2003 *Science* **301** 1235
[19] Perkins T T, Dalal R V, Mitsis P G and Block S M 2003 *Science* **301** 1914
[20] Lovett S T and Kolodner R D 1989 *Proc. Natl Acad. Sci. USA* **86** 2627
[21] Rajman L A and Lovett S T 2000 *J. Bacteriol.* **182** 607
[22] Yamagata A, Kakuta Y, Masui R and Fukuyama K 2002 *Proc. Natl Acad. Sci. USA* **99** 5908
[23] Courcelle C T, Chow K-H, Casey A and Courcelle J 2006 *Proc. Natl Acad. Sci. USA* **103** 9154
[24] Han E S, Cooper D L, Persky N S, Sutera V A, Whitaker R D, Montello M L and Lovett S T 2006 *Nucleic Acids Res.* **34** 1084
[25] Lehman I R and Nussbaum A L 1964 *J. Biol. Chem.* **239** 2628
[26] Breyer W A and Matthews B W 2000 *Nat. Struct. Biol.* **7** 1125
[27] Viswanathan M and Lovett S T 1999 *J. Biol. Chem.* **274** 30094
[28] Breyer W A and Matthews B W 2001 *Protein Sci.* **10** 1699
[29] Astumian R D and Bier M 1994 *Phys. Rev. Lett.* **72** 1766
[30] Jülicher F, Ajdari A and Prost J 1997 *Rev. Mod. Phys.* **69** 1269
[31] Fox R F and Choi M H 2001 *Phys. Rev.* E **63** 051901
[32] Xie P, Dou S-X and Wang P-Y 2007 *J. Mol. Biol.* **366** 976
[33] Xie P, Dou S-X and Wang P-Y 2007 *Biochim. Biophys. Acta* **1767** 1418
[34] Xie P 2008 *Biochim. Biophys. Acta* **1777** 1195
[35] Lipowsky R and Liepelt S 2008 *J. Stat. Phys.* **130** 39
[36] Kafri Y, Lubensky D K and Nelson D R 2004 *Biophys. J.* **86** 3373
[37] Betterton M D and Jülicher F 2003 *Phys. Rev. Lett.* **91** 258103
[38] Xie P 2006 *Biochim. Biophys. Acta* **1764** 1719
[39] Xie P 2007 *J. Theor. Biol.* **249** 566
[40] Elston T C and Peskin C S 2000 *SIAM J. Appl. Math.* **60** 842
[41] Mai J, Sokolov I M and Blumen A 2001 *Phys. Rev.* E **64** 011102
[42] Antal T and Krapivsky P L 2005 *Phys. Rev.* E **72** 046104
[43] Saffarian S, Qian H, Collier I, Elson E and Goldberg G 2006 *Phys. Rev.* E **73** 041909
[44] Qian J, Xie P, Dou S-X and Wang P-Y 2006 *J. Theor. Biol.* **243** 322
[45] Xie P 2007 *Arch. Biochem. Biophys.* **457** 73
[46] Xie P 2009 *J. Theor. Biol.* **259** 434
[47] Xie P 2009 *Virus Res.* doi:10.1016/j.virusres.2009.03.022
[48] Xie P 2008 *Biosystem* **93** 199
[49] Xie P 2009 *Biochim. Biophys. Acta* **1789** 212
[50] Gardiner C W 1983 *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Berlin: Springer)
[51] Viadiu H and Aggarwal A K 2000 *Mol. Cell* **5** 889
[52] Newman M, Strzelecka T, Dorner L F, Schildkraut I and Aggarwal A K 1995 *Science* **269** 656
[53] Herbert K M, La Porta A, Wong B J, Mooney R A, Neuman K C, Landick R and Block S M 2006 *Cell* **125** 1083
[54] Genschel J, Curth U and Urbanke C 2000 *Biol. Chem.* **381** 183
[55] Han E S, Cooper D L, Persky N S, Sutera V A Jr, Whitaker R D, Montello M L and Lovett S T 2006 *Nucleic Acids Res.* **34** 1084
[56] Sharma R and Rao D N 2009 *J. Mol. Biol.* **385** 1375