Nucleotide Specificity versus Complex Heterogeneity in Exonuclease Activity Measurements

Jörg Enderlein
Institut für Neurowissenschaft und Biophysik 1, Forschungszentrum Jülich, Jülich, Germany

ABSTRACT A recent publication reported on measurements of Exonuclease I activity using a real-time fluorescence method that measures the time required by molecules of Exonuclease I to hydrolyze single-stranded DNA that was synthesized to have two fluorescently labeled nucleotides. The observed fluorescence-intensity curves were interpreted as a sign of strong heterogeneity of the activity of Exonuclease I. Here, I propose a different model, which assumes that Exonuclease I activity is nucleotide-dependent, and that a fluorescent label bound to a nucleotide significantly slows its cleavage rate. The presented model fits the observed data equally well, but can be used to make specific predictions upon observable sequence dependence of measured fluorescence-intensity curves.

INTRODUCTION

In recent years, several single-molecule experiments have demonstrated the amazing fact that the enzymatic activity of various proteins can vary considerably from molecule to molecule, or even fluctuate in time for one and the same molecule. Prominent examples are the monitoring of the enzymatic activity of single molecules of cholesterol oxidase (1), flavin reductase (2), alkaline phosphatase (3), or β-galactosidase (4). This heterogeneity is attributed to static structural heterogeneity or dynamic structural heterogeneity caused by fluctuations of a molecule between different structurally similar subconformations. Thus, it suggests that we should look for catalytic heterogeneity in other proteins also.

A recent article by Werner et al. (5) claims to have found a broad heterogeneity in the catalytic activity of Exonuclease I (Exo I). Their experiment is schematically depicted in Fig. 1. A large number of identical DNA single strands (56 nucleotides) is bound, at one end, to a polymer bead. The DNA strands are fluorescently labeled at two specific sites in the nucleotide sequence (positions 5 and 38 from the free end). The bead is incubated with Exo I in the absence of Mg^{2+}, so that the Exo I can bind to the DNA but is not able to cleave nucleotides. Then, the bead is suspended into a fluid flow and kept there by an optical tweezer. After addition of Mg^{2+} to the flow, Exo I starts cleaving single nucleotides in a processive way, and the cleavage of the labeled nucleotides is monitored by laser-induced fluorescence downstream.

The observed fluorescence intensity as a function of time consists of two broadened peaks caused by the cleavage of the first (position 5) and second (position 38) labeled nucleotides. By measuring the time between the peak maxima, one can estimate the average cleavage rate of Exo I. A surprising observation in Werner et al. (5) was that the fluorescence peak widths were much broader than expected. The authors used a simple-chain first-order kinetic model with uniform Exo I activity for fitting the data. Thus, the underlying kinetics was described by the simple reaction scheme

$$\frac{dp_n(t)}{dt} = -kp_n(t) + kp_{n-1}(t). \quad (1)$$

where $p_n(t)$ is the probability that the n^{th} nucleotide is the last nucleotide of the strand, and k is the assumedly uniform cleavage rate constant independent of the nature of nucleotide or local sequence. This model can be solved analytically (see, e.g., Gardiner (6)), and assuming that the observed fluorescence intensity $I(t)$ is proportional to the cleavage rate of the 5th plus the 38th nucleotide, one finds

$$I(t) \propto k[p_5(t) + p_{38}(t)] = k\left[\frac{(kt)^4}{4!} + \frac{(kt)^{37}}{37!}\right]\exp(-kt) \quad (2)$$

This leads to peak widths that are much more narrow than observed in the experiment, which Werner et al. interpreted as a sign of a strong heterogeneity of Exo I hydrolysis rates. Thus, they used the expression in Eq. 2 together with a truncated Gaussian distribution of hydrolysis rates k to fit the measured data and found that a Gaussian distribution with a peak at ~100 nucleotides (nt)/s, a width of ~65 nt/s, and truncated for values <10 nt/s fits the measurements best. Besides yielding an average value for the cleavage rate of ~174 nt/s (at ~31°C), which is significantly lower than that measured by more conventional techniques (275 nt/s at 37°C) (7), the distribution width found is exceptionally broad.

Surprisingly, Werner et al. completely neglected the possibility that the hydrolysis activity of Exo I could be nucleotide-specific, or that at least there could be a different cleavage rate for nucleotides with and without bound fluorescent labels. Nucleotide specificity was reported, e.g., for the activity of bacteriophage λ-exonuclease digestion of λ-phage DNA (8).

RESULTS

We will first consider a simple modification to the model (Eq. 1). Let us assume that the cleavage rate constants for labeled and unlabeled nucleotides are different. Thus, one now has

Submitted August 23, 2006, and accepted for publication November 13, 2006.

Address reprint requests to Jörg Enderlein, Institut für Neurowissenschaft und Biophysik 1, Forschungszentrum Jülich, Jülich, Germany. E-mail: j.enderlein@fz-juelich.de.

0006-3495/07/03/1556/03 $2.00

doi: 10.1529/biophysj.106.095851

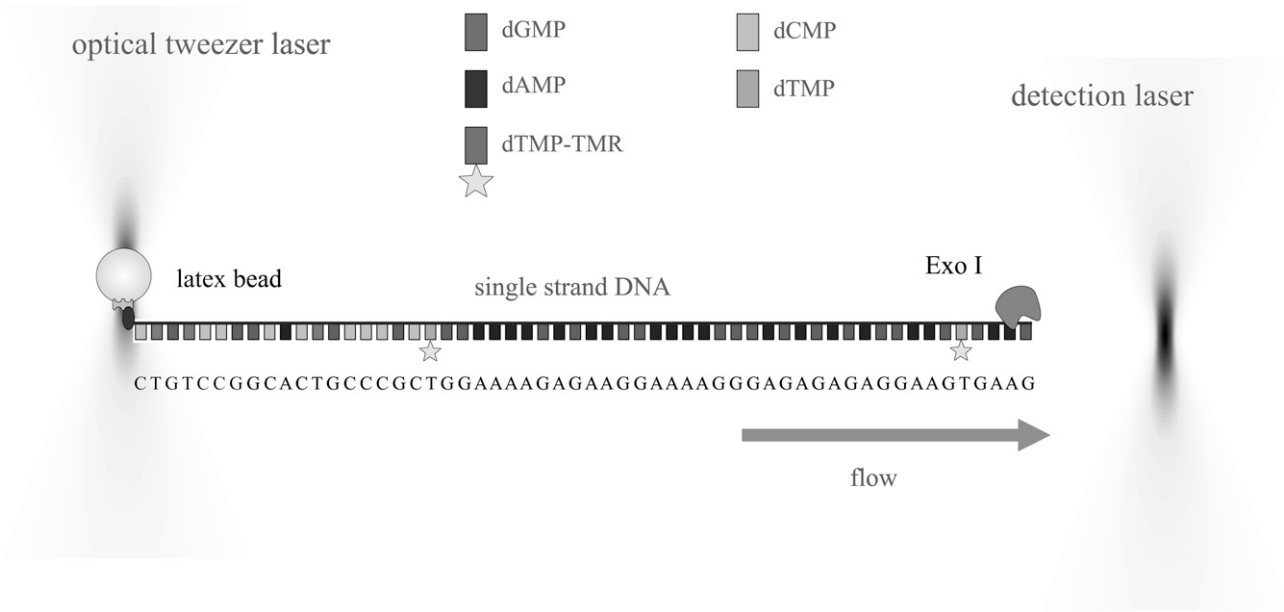

FIGURE 1 Principal scheme of the experiment. A single-strand DNA (57mer) is bound, via a biotin-streptavidin bond, to a latex bead that is kept within a fluid flow with an optical tweezer. Upon addition of Mg^{2+} to the flow, an attached Exo I starts to cleave off nucleotides from the strand. The 5th and 37th nucleotides (dTMP-TMR, position counted from the free end) are fluorescently labeled. After cleavage of these labeled nucleotides, they are transported by the fluid flow through a detection laser, and the resulting fluorescence is recorded.

$$\frac{dp_n(t)}{dt} = -k_n p_n(t) + k_{n-1} p_{n-1}(t), \tag{3}$$

where k_n takes a value k_u if the n^{th} nucleotide is unlabeled, and a value k_l if the n^{th} nucleotide is fluorescently labeled. The system of coupled differential equations (3) no longer has a simple analytic solution as Eq. 1, but can be solved numerically in a straightforward way using matrix exponentiation as implemented in many modern mathematics software packages such as Matlab or Mathematica. To check how well such a modified model would fit the data of Werner et al., I first calculated model data using Eq. 1 and assuming a Gaussian distribution of cleavage rates k, as reported by Werner et al. and described above. The resulting fluorescence-intensity curve comes very close to the observed intensities in the experiment (see Fig. 3 in Werner et al. (5)). The calculated model data curve is depicted by circles in Fig. 2. This figure also shows the best fit of Eq. 2, clearly demonstrating that the simple-model Eq. 1 indeed does not fit the observed data. The best-fit rate constant k of that model is equal to ~130 nt/s. The second solid-line curve in Fig. 2 shows the best fit of the modified system, Eq. 3, to the model data. The best-fit rate constants are now $k_u = 204$ nt/s and $k_l = 7.34$ nt/s. Although the fit quality is still not satisfactory, it demonstrates that assuming a strongly hindered cleavage rate for labeled nucleotides automatically leads to much broader peaks than assuming only a single uniform cleavage rate constant. Thus, peak broadening as observed in the experiments does not automatically imply broad distributions of cleavage rate constants.

A possible modification for Eq. 3 is to assume different cleavage rates for different nucleotides. In that case, for the sequence shown in Fig. 1, one has to consider three rate constants: the two rate constants, k_G and k_A, for cleaving a guanine and an adenine, respectively, and $k'_T \equiv k_l$ for cleaving the labeled thymine. Then, a least-squares fit returns the values $k_G = 403$ nt/s, $k_A = 201$ nt/s, and $k'_T = 6.95$ nt/s. However, fit quality is not significantly improving and similar to using the simpler model of Eq. 3.

Using this insight, a more refined model can be proposed for the observed data. This model assumes that the exonuclease cleavage rate is dependent on the nature of the nucleotides adjacent to the cleaved bond. Then, one has

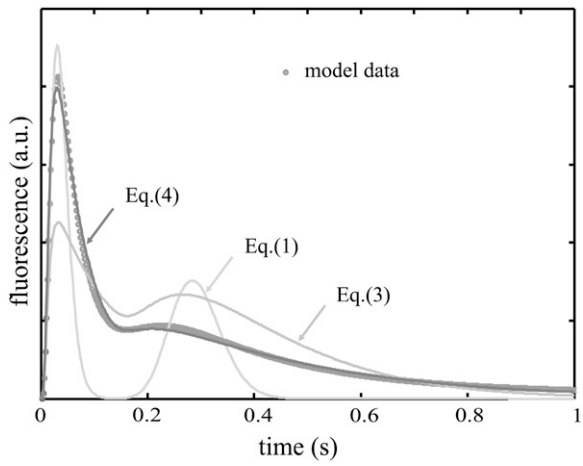

FIGURE 2 Fits of the different models to the model data (*open circles*). For details, see text.

$$\frac{dp_{\mathrm{n}}(t)}{dt} = -k_{\mathrm{s(n),s(n+1)}}p_{\mathrm{n}}(t) + k_{\mathrm{s(n-1),s(n)}}p_{\mathrm{n-1}}(t), \qquad (4)$$

where $k_{\alpha,\beta}$ is the cleavage rate constant for a bond between nucleotide α and nucleotide β, and $s(n)$ is the nucleotide at the n^{th} position. Notice that, in general, $k_{\alpha,\beta}$ must not to be equal to $k_{\beta,\alpha}$, reflecting the unidirectionality of the cleavage process. Thus, for calculating the observed fluorescence intensity one has now to consider seven different cleavage rate constants: $k_{\mathrm{G,G}}$, $k_{\mathrm{G,A}}$, $k_{\mathrm{A,G}}$, $k_{\mathrm{A,A}}$, $k'_{\mathrm{G,T}}$, $k'_{\mathrm{T,G}}$, and $k'_{\mathrm{T,C}}$. A nucleotide symbol with a star denotes a nucleotide with a fluorescent label attached. The best fit of this generalized model to the model data is shown in Fig. 2 also. The obtained cleavage rate constants are $k_{\mathrm{G,G}} = 373$ nt/s, $k_{\mathrm{G,A}} = 248$ nt/s, $k_{\mathrm{G,A}} = 247$ nt/s, $k_{\mathrm{A,A}} = 236$ nt/s, $k'_{\mathrm{G,T}} = 207$ nt/s, $k'_{\mathrm{T,G}} = 15.7$ nt/s, and $k'_{\mathrm{T,C}}$ 3.24 nt/s. Taking into account that there are, between the first and last labeled nucleotides, five GG-bonds, nine AA-bonds, 10 GA-bonds, and 10 AG-bonds, the average cleavage rate constant for unlabeled nucleotides in the given sequence is ~270 nt/s, a value surprisingly close to that reported by Brody et al. (7). Moreover, the fit results suggest that GG-bonds are cleaved faster than AA-bonds, and the cleavage rate constants for GA- and AG-bonds are close and intermediate between those for AA- and GG-bonds. Also, the label on a nucleotide seems to slow down the cleavage of an adjacent nucleotide, as shown by the value of $k'_{\mathrm{G,T}}$ which is smaller than those of $k_{\mathrm{G,G}}$ and $k_{\mathrm{G,A}}$, whereas the presence of a label significantly reduces the cleavage rate of the labeled nucleotide itself, as shown by the small values of $k'_{\mathrm{T,G}}$ and $k'_{\mathrm{T,C}}$.

DISCUSSION AND CONCLUSION

It should be noted that the fit quality of our generalized model to the model data, as shown in Fig. 2, is comparable with the fit quality of the distributed rate constant model as used by Werner et al., which can be seen by comparing our Fig. 2 with Fig. 3 in Werner et al. (5). Thus, both models consistently fit the observed data similarly well. Because the measurements in Werner et al. (5) were performed only on the single sequence shown in Fig. 1, no decision can be made between both models using the existing data. Moreover, the proposed model as described by Eq. 4 assumes, in general, $4 \times 4 = 16$ independent rate constants for bond cleavage between unlabeled nucleotides, and, at maximum, another 32 rate constants for bond cleavage between all possible combinations of a labeled and an unlabeled nucleotide (although it would be advisable to use only one type of nucleotide for labeling, which then reduces the number of rate constants accordingly). However, in contrast to the approach by Werner et al., the model presented here can make testable predictions. Because the cleavage rate is strongly sequence-dependent, different sequences will lead to different but predictable fluorescence-intensity curves. As an example, I calculated the expected fluorescence-intensity curves when all guanines from positions 7 through 36 are replaced by adenines (polyA sequence), and vice versa (polyG sequence), using the fit results for the cleavage rate constants of the previous section. The resulting time courses of the fluorescence intensity are depicted in Fig. 3, together with that for the original sequence. Thus, by performing measurements like those in Werner et al. (5) on different sequences, one may, in the end, uniquely untangle sequence specificity of exonuclease activity from an actual distribution of rate constants.

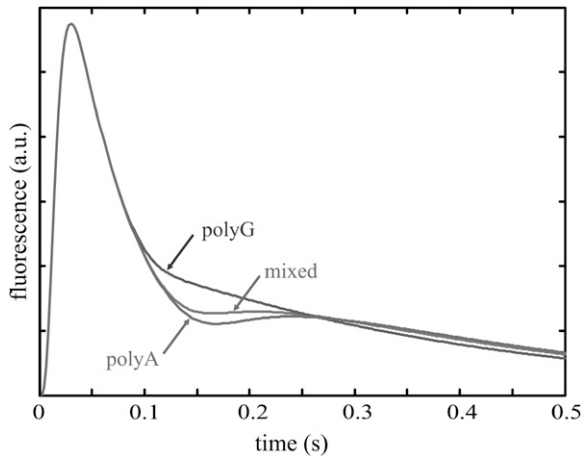

FIGURE 3 Expected fluorescence-intensity curves if all guanines from positions 7 through 36 were to be replaced by adenines (polyA sequence), and vice versa (polyG sequence). The fit result of Eq. 4 for the original sequence is also shown. For better visibility of the differences the time axis is shorter than in Fig. 2.

I thank Hong Cai for suggesting that I work on the matter presented here.

REFERENCES

1. Lu, H. P., L. Xun, and X. S. Xie. 1998. Single-molecule enzymatic dynamics. *Science*. 282:1877–1882.
2. Yang, H., G. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie. 2003. Protein conformational dynamics probed by single-molecule electron transfer. *Science*. 302:262–266.
3. Craig, D. B., E. A. Arriaga, J. C. Y. Wong, H. Lu, and N. J. Dovichi. 1996. Studies on single alkaline phosphatase molecules: reaction rate and activation energy of a reaction catalyzed by a single molecule and the effect of thermal denaturation. The death of an enzyme. *J. Am. Chem. Soc.* 118:5245–5253.
4. Craig, D. B., and N. J. Dovichi. 1998. *Escherichia coli* β-galactosidase is heterogeneous with respect to the activity of individual molecules. *Can. J. Chem.* 76:623–626.
5. Werner, J. H., H. Cai, R. A. Keller, and P. M. Goodwin. 2005. Exonuclease I hydrolyzes DNA with a distribution of rates. *Biophys. J.* 88:1402–1412.
6. Gardiner, C. W. 1985. Handbook of Stochastic Methods. Springer, New York.
7. Brody, R., K. Doherty, and P. Zimmerman. 1986. Processivity and kinetics of the reaction of exonuclease-I from *Escherichia coli* with polydeoxyribonucleotides. *J. Biol. Chem.* 261:7136–7143.
8. Van Oijen, A. M., P. C. Blainey, D. J. Crampton, C. C. Richardson, T. Ellenberger, and X. S. Xie. 2003. Single-molecule kinetics of λ exonuclease reveal base dependence and dynamic disorder. *Science*. 301:1235–1238.