



# The diffusion constant of a labeled protein sliding along DNA

Isabelle Bonnet, P. Desbiolles

## ► To cite this version:

Isabelle Bonnet, P. Desbiolles. The diffusion constant of a labeled protein sliding along DNA. European Physical Journal E: Soft matter and biological physics, 2011, 34 (3), 10.1140/epje/i2011-11025-8 . hal-02466069

**HAL Id: hal-02466069**

**<https://hal.science/hal-02466069>**

Submitted on 4 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The diffusion constant of a labeled protein sliding along DNA

I. Bonnet<sup>a</sup> and P. Desbiolles<sup>b</sup>

Laboratoire Kastler Brossel, ENS, UPMC-Paris 6, CNRS UMR 8552, 24 rue Lhomond, 75005 Paris, France

Received 1 December 2010 and Received in final form 4 February 2011

Published online: 11 March 2011 – © EDP Sciences / Società Italiana di Fisica / Springer-Verlag 2011

**Abstract.** Long ago inferred by biochemists, the linear diffusion of proteins along DNA has recently been observed at a single-molecule level using fluorescence microscopy. This imaging technique requires labeling the protein of interest with a fluorophore, usually an organic nanosized dye that is not supposed to impact the dynamics of the protein. Yet individual proteins can also be tracked using much larger labels, like quantum dots or beads. We investigate here the impact of such a large label on the protein diffusion along DNA. Solving a Fokker-Planck equation, we estimate the diffusion constant of a protein-label complex diffusing in a periodic potential that mimics the DNA-protein interaction, the link between the protein and the label being modeled as a Hookean spring. Our results indicate that the diffusion constant can generally be calculated by considering that the motion of the protein in the DNA potential is decoupled from the Brownian motion of the label. Our conclusions are in good agreement with the experimental results we obtained with the restriction enzyme EcoRV, assuming a rotation-coupled diffusion of the enzyme along DNA.

## 1 Introduction

Many biological functions depend on interactions between DNA and site-specific DNA-binding proteins, which must search through large nonspecific regions of DNA to find their target, usually a few base-pair DNA sequence. Without consuming energy, some of these proteins are able to rapidly find their target DNA, occasionally even faster than allowed by three-dimensional diffusion [1]. Since long this “facilitated diffusion” has been connected to a translocation of proteins along nonspecific DNA [2]. Although the characterization of the microscopic processes involved in this translocation has been widely discussed over the past forty years [3–11], it is only recently that facilitated diffusion has been addressed using single-molecule approaches [12,13]. Among them, fluorescence-based techniques have been frequently used because they allow the direct visualization of the interaction between a protein and a DNA molecule [14]. Several teams have observed and characterized the diffusive motion of single fluorescently labeled proteins interacting with noncognate DNA [15–22], thus confirming the long-dating assumption that some proteins can slide along the nonspecific parts of DNA to accelerate their target search [23]. Furthermore, single-molecule fluorescence microscopy gives access to quantitative information regarding the dynamics of DNA-protein

interactions. For instance, the diffusion constant  $D$  associated to the Brownian motion of a protein along the DNA can be deduced from the analysis of single-protein trajectories. Recently, Blainey *et al.* [24], by comparing the linear diffusion constants  $D$  of proteins of various sizes, found that  $D$  decreases as the protein radius increases in a way consistent with a helical motion, which indicates that, for these proteins, sliding is a rotation-coupled translation. To detect the proteins by fluorescence microscopy, Blainey *et al.* used organic dyes whose sizes were much smaller than those of the proteins. It is therefore unlikely that the dynamics of the proteins was impacted by the size of the fluorescent dyes, and thus the labeled proteins could be considered as rigid bodies for the estimation of the linear diffusion constant and comparison with experimental data [24].

However, the fluorescent label used to track the protein motion by fluorescence microscopy can be as large as or even larger than the protein [20,25]. For instance, we recently investigated the linear diffusion of the restriction enzyme EcoRV, fluorescently labeled with a quantum dot (QD), the size of which was about 4 times the size of EcoRV. We found a diffusion constant  $D_{\text{QD}}$  for the EcoRV-QD complex which was about one third of the diffusion constant of EcoRV labeled with an organic dye [21,26]. This value for  $D_{\text{QD}}$  is much larger than that expected by considering the protein-QD complex as a rigid body diffusing along DNA, whether or not the complex rotates while sliding (see sect. 4).

To account for our experimental results, here we investigate the impact of the label size on the diffusion constant

<sup>a</sup> Present address: Developmental Biology and Genetics, Institut Curie, CNRS UMR 3215 and INSERM U 934, 26 rue d’Ulm, 75248 Paris cedex 05, France.

<sup>b</sup> e-mail: pierre.desbiolles@lkb.ens.fr



of a protein sliding along DNA. We do not consider the protein-label complex as a rigid body, but we assume that the label is attached to the protein by a flexible link which allows rapid Brownian fluctuations of the label. To derive the diffusion constant  $D$  of such a protein-label complex, we follow an approach developed by Zeldovich *et al.* [27], who described the motion of a molecular motor to which a large cargo is elastically attached. We similarly model the link between the protein and its label as a Hookean spring, and we solve the Fokker-Planck equation for the protein-label complex, modeling the DNA-protein interaction potential as a sinusoidal potential. This approach allows us to determine the dependence of  $D$  on the size of the label, a dependence which strongly differs from that predicted when considering the protein-label complex as a rigid body.

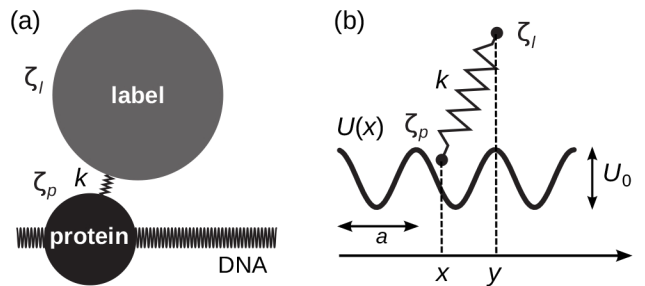
We then compare the diffusion constants  $D$  predicted by our model with the experimental results we obtained with EcoRV, a type-II restriction enzyme the linear diffusion of which we recently characterized [21,26]. In our experiments, we labeled EcoRV with three different fluorescent labels of increasing size, and we measured the diffusion constants  $D$  using single-molecule fluorescence microscopy. We show here that our model accounts for our experimental values for  $D$ , if we assume that the protein rotates when sliding along the DNA.

The paper is organized as follows. In sect. 2, we introduce the model and we derive the analytical expression of the linear diffusion constant  $D$  of the protein-label complex. Then we calculate  $D$  by solving both analytically and numerically the Fokker-Planck equation that describes our system. We derive a simple expression of  $D$ , for which we give a physical interpretation. In sect. 3, the dependence of  $D$  on the label size predicted by our model is compared with experimental data obtained with EcoRV. In sect. 4, we discuss alternative models, showing however that they fail to interpret our data.

## 2 Estimation of the linear diffusion coefficient of a labeled protein

### 2.1 A model for a protein sliding along DNA

We consider a protein-label complex sliding along DNA (fig. 1(a)). We assume that only the protein interacts with the DNA, and that the label is attached to the protein by a flexible linker, *e.g.* a PEG linker of a few tens of ethylene glycol units [24]. In the following we suppose that the typical time  $\tau_S$  the protein-label complex needs to slide over one DNA base-pair is large compared to the relaxation time  $\tau_R$  of the linker. Typically, for a PEG linker,  $\tau_R \sim 50$  ns [24], to be compared to  $\tau_S = a^2/2D \sim 5$   $\mu$ s, with  $a = 0.34$  nm (length of one DNA base-pair) and  $D \sim 10^{-2}$   $\mu$ m<sup>2</sup>s<sup>-1</sup>, as measured for EcoRV [21]. Under the assumption  $\tau_S \gg \tau_R$ , the dynamics of the label can be studied, as a first approximation, by considering that the protein is immobile. We then describe the linker as a polymer with a mean end-to-end length  $z_0$  that is within



**Fig. 1.** Model of a labeled protein diffusing along DNA. (a) The protein and the label are modeled as beads with viscous friction coefficient  $\zeta_p$  and  $\zeta_l$ . The protein slides along the DNA and is attached to the label by a linker modeled as a linear elastic spring with a spring constant  $k$ . (b) The DNA-protein interaction is modeled as a sinusoidal potential  $U(x)$  with an amplitude  $U_0$  and a periodicity  $a = 1$  DNA base-pair. The positions of the protein and of the label are denoted by  $x$  and  $y$ , respectively.

the nanometer range, a length that is also the typical amplitude of the Brownian fluctuations of the label around its equilibrium position. The elasticity of such a linker is entropic in nature, thus nonlinear, but for simplicity we model the linker as a Hookean spring with a spring constant  $k \sim k_B T / z_0^2$  (about 1 pN/nm for  $z_0 = 2$  nm), with  $T$  denoting the temperature and  $k_B$  the Boltzmann's constant. In addition, the fast relaxation of the linker allows us to well define the position of the label and to consider that  $\tau_S$  is the smallest characteristic time involved in our calculations. Since the assumption  $\tau_S \gg \tau_R$  is crucial in our approach, we now investigate its validity. First, since  $\tau_S$  scales as  $D^{-1}$ , this assumption fails for rapidly diffusing proteins ( $D > 1$   $\mu$ m<sup>2</sup>s<sup>-1</sup>). Second, the relaxation time  $\tau_R$  of the linker-label complex can be estimated assuming that the label is a sphere of radius  $r_l$  attached to a flexible polymer whose second end is immobile. For a small label ( $r_l \ll z_0$ ),  $\tau_R$  is the relaxation time of the polymer:  $\tau_R \sim \eta z_0^3 / k_B T$ , where  $\eta$  is the viscosity of the solution ( $10^{-3}$  Pa s for water), thus  $\tau_S \gg \tau_R$  ( $z_0 = 2$  nm gives  $\tau_R \sim 2$  ns). For a large label ( $r_l \gg z_0$ ), the relaxation time is given by  $\tau_R \sim \zeta_l / k$ , where  $\zeta_l$  is the friction coefficient of the label. Assuming  $\zeta_l = 6\pi\eta r_l$ , one gets  $\tau_R \sim 6\pi\eta z_0^2 r_l / k_B T$ , and the relation  $\tau_S \gg \tau_R$  leads to  $r_l \ll 300$  nm for  $\tau_S \sim 5$   $\mu$ s. In conclusion, the hypothesis  $\tau_S \gg \tau_R$  is correct when  $D \ll 1$   $\mu$ m<sup>2</sup>s<sup>-1</sup> and  $r_l \ll 300$  nm, two conditions well satisfied in our experiments.

The diffusion of a protein along DNA is a base-pair reading interaction, and thus sequence dependent [28]. Yet for simplicity we neglect this dependence in the first instance, and we consider here sliding as a one-dimensional diffusion in a sinusoidal potential (fig. 1(b)):

$$U(x) = -\frac{U_0}{2} \cos(2\pi x/a),$$

where  $U_0$  is the amplitude of the potential,  $x$  the position of the protein interacting with the DNA, and  $a$  the length of one DNA base-pair. We discuss later the case of diffusion in a random potential, a more realistic description of the DNA-protein interaction.

Our first goal is to calculate the 1D diffusion constant  $D$  of a labeled protein diffusing in the potential  $U(x)$  as a function of  $D_p$  and  $D_l$ , the 3D diffusion constants of the protein and of the label, *i.e.* the diffusion constants of objects freely diffusing in solution. Assuming that the protein and the label can be treated as spherical objects of radius  $r_p$  and  $r_l$ , their friction coefficients are  $\zeta_{p,l} = 6\pi\eta r_{p,l}$ , and  $D_{p,l}$  are related to  $\zeta_{p,l}$  by the Einstein relation  $D_{p,l} = k_B T / \zeta_{p,l}$ . The radius of most of the proteins that have been observed sliding along DNA is in the nanometer range (for instance we measured that the radius of EcoRV is about 4 nm [21]). Therefore, when considering an organic dye as a label, we expect  $D_l \gg D_p$ , as the dye radius is about 0.5 nm [29]. The label size, however, can be much larger than that of the protein: the radius of a commercial functionalized quantum dot is typically 15–20 nm (according to suppliers), and the beads used for tracking proteins by bright-field light microscopy are usually much larger than 20 nm, which implies that  $D_l \ll D_p$ . Apart from these two extreme regimes ( $D_l \gg D_p$  and  $D_l \ll D_p$ ), protein sliding can also be investigated using fluorescent proteins, such as fluorescently labeled streptavidin. In this case  $D_l \simeq D_p$  as the size of the label is similar to that of the protein. It is therefore essential to find an expression for  $D$  that is not restricted to limiting cases ( $D_l \gg D_p$  or  $D_l \ll D_p$ ) but valid whatever the label size.

## 2.2 General expression of the diffusion coefficient of a labeled protein

We follow an approach initially developed for studying motor proteins transporting cargos [27]. Though most of the results of this section can be derived from the original paper, we remind the reader the main steps of deriving  $D$ , using our own notations. The probability  $P(x, y, t)$  of finding the protein at position  $x$  with the label at position  $y$  at time  $t$  satisfies the Fokker-Planck equation:

$$\begin{aligned} \frac{\partial P}{\partial t} = & D_p \frac{\partial^2 P}{\partial x^2} + D_l \frac{\partial^2 P}{\partial y^2} + D_l \frac{\partial}{\partial y} \left[ P \frac{y-x}{z_0^2} \right] \\ & + D_p \frac{\partial}{\partial x} \left[ P \left( \frac{d\tilde{U}}{dx} + \frac{x-y}{z_0^2} \right) \right], \end{aligned} \quad (1)$$

with  $\tilde{U} = U/k_B T$ . Following [27], we look for a solution of this equation in the form

$$P = \frac{1}{\sqrt{Dt}} f(\tau, u, \tilde{z}),$$

with  $\tau = x/\sqrt{Dt}$ ,  $u = x/a$  and  $\tilde{z} = (x-y)/z_0$ . At long time, *i.e.* when the characteristic length  $\sqrt{Dt}$  over which the protein has diffused is much larger than the periodicity  $a$  of  $U(x)$ , the diffusive motion of the protein-label complex is described by the large-scale variation of  $f$  with  $\tau$ , while the variations of  $f$  with  $u$  and  $\tilde{z}$  reflect the modulation of the protein motion by the potential  $U(x)$  [27].

Thus, introducing the small parameter  $\epsilon = a/\sqrt{Dt}$ , the Fokker-Planck equation (1) can be written as

$$[\hat{L}_0 + \epsilon \hat{L}_1 + \epsilon^2 \hat{L}_2] f = 0. \quad (2)$$

Expressions of the operators  $\hat{L}_0$ ,  $\hat{L}_1$  and  $\hat{L}_2$  can be found in appendix A. Inserting a solution  $f$  in the form

$$f(\tau, u, \tilde{z}) = \sum_{i=0}^2 \epsilon^i g_i(\tau) h_i(u, \tilde{z}), \quad (3)$$

in relation (2), and stating the following relations between  $g_1(\tau)$ ,  $g_2(\tau)$  and  $g_0(\tau)$ :

$$g_1(\tau) = \frac{dg_0(\tau)}{d\tau}, \quad g_2(\tau) = \frac{d^2 g_0(\tau)}{d\tau^2},$$

one obtains, at second order in  $\epsilon$ ,  $g_0(\tau) = \exp(-\tau^2/4)$  and the following expression for the diffusion coefficient  $D$  of the protein-label complex:

$$D = D_p \left( 1 + \int_0^1 du \int_{-\infty}^{\infty} d\tilde{z} \left( \frac{d\tilde{U}}{du} + \frac{a}{z_0} \tilde{z} \right) h_1(u, \tilde{z}) \right). \quad (4)$$

Thus an estimation of  $h_1(u, \tilde{z})$  is necessary for calculating  $D$ . We propose two approaches to determine this probability: i) an analytical approach, which provides an approximate solution valid when the amplitude of the Brownian fluctuations is large compared to the DNA-protein potential periodicity ( $z_0 \gg a$ ), ii) a numerical approach, following a numerical scheme proposed in [27].

## 2.3 Analytical expression of the diffusion coefficient

The differential equation that drives  $h_1(u, \tilde{z})$  is obtained from eq. (2) at first order in  $\epsilon$ . It involves  $h_0(u, \tilde{z})$ , which can be derived from eq. (2) at the lowest order in  $\epsilon$ :  $\hat{L}_0 f_0 = 0$ , with  $f_0 = g_0(\tau) h_0(u, \tilde{z})$ . This equation can be written as

$$\left[ \hat{L}_{0,u} + \frac{a}{z_0} \hat{L}_{0,uz} + \left( \frac{a}{z_0} \right)^2 \frac{D_p + D_l}{D_p} \hat{L}_{0,z} \right] h_0 = 0, \quad (5)$$

where the operator  $\hat{L}_0$  has been split into three sub-operators  $\hat{L}_{0,u}$ ,  $\hat{L}_{0,z}$  and  $\hat{L}_{0,uz}$  (given in appendix A), respectively acting on  $u$  only,  $\tilde{z}$  only and coupling  $u$  and  $\tilde{z}$ . Equation (5) has an analytical solution in the form  $h_0(u, \tilde{z}) = l_0(u) k_0(\tilde{z})$  (see appendix A for the expressions of  $l_0(u)$  and  $k_0(\tilde{z})$ ). Since  $\hat{L}_{0,u} l_0(u) = 0$ ,  $l_0(u)$  describes the diffusion of an unlabeled protein in the potential  $\tilde{U}$ . Similarly,  $\hat{L}_{0,z} k_0(\tilde{z}) = 0$ , and thus  $k_0(\tilde{z})$  describes the Brownian fluctuations of the label around a static protein. Consequently, as  $\hat{L}_0 f_0 = 0$ , one gets  $\hat{L}_{0,uz} h_0 = 0$ , as in this case the label motion is independent of that of the protein.

At the first order in  $\epsilon$ , eq. (2) reads  $g_1(\tau)\hat{L}_0 h_1 + \hat{L}_1 g_0 h_0 = 0$ , which can be written as

$$\left[ \hat{L}_{0,u} + \frac{a}{z_0} \hat{L}_{0,uz} + \left( \frac{a}{z_0} \right)^2 \frac{D_p + D_l}{D_p} \hat{L}_{0,z} \right] h_1(u, \tilde{z}) = \left( \pi \tilde{U}_0 \sin(2\pi u) + \tilde{z} \frac{a}{z_0} \right) l_0(u) k_0(\tilde{z}). \quad (6)$$

We did not find an analytical solution of this equation. Yet without the coupling term  $\hat{L}_{0,uz} h_1$ , a solution of (6) is given by

$$h_1^{\text{nc}}(u, \tilde{z}) = l_0(u) k_1(\tilde{z}) + l_1(u) k_0(\tilde{z}), \quad (7)$$

where  $l_1(u)$  and  $k_1(\tilde{z})$  are solutions of

$$\begin{aligned} \hat{L}_{0,u} l_1(u) &= \pi \tilde{U}_0 \sin(2\pi u) l_0(u), \\ \frac{a}{z_0} \frac{D_p + D_l}{D_p} \hat{L}_{0,z} k_1(\tilde{z}) &= \tilde{z} k_0(\tilde{z}). \end{aligned}$$

The expressions of  $l_1(u)$  and  $k_1(\tilde{z})$  are given in appendix A. Then we looked for an approximate solution of (6) with the following ansatz:

$$h_1^*(u, \tilde{z}) = A l_0(u) k_1(\tilde{z}) + B l_1(u) k_0(\tilde{z}), \quad (8)$$

where  $A$  and  $B$  are constants. Inserting (8) in (6), we found  $A$  and  $B$  that depend only on  $D_p$ ,  $D_l$  and  $\tilde{U}_0$  (see appendix A). The ansatz function  $h_1^*$  is then solution of

$$\hat{L}_0 h_1^*(u, \tilde{z}) = \left( \pi \tilde{U}_0 \sin(2\pi u) + \tilde{z} \frac{a}{z_0} \right) l_0(u) k_0(\tilde{z}) + \phi(u, \tilde{z}),$$

which is similar to (6) with the additional term  $\phi(u, \tilde{z})$

$$\phi(u, \tilde{z}) = \tilde{z} \frac{a}{z_0} \frac{D_l}{D_p + D_l I_0^2(\tilde{U}_0/2)} (l_0(u) - 1) k_0(\tilde{z}),$$

where  $I_0$  is the modified Bessel function of the first kind. Thus, the probability  $h_1^*(u, \tilde{z})$  is a good approximation of  $h_1(u, \tilde{z})$  if  $|\phi(u, \tilde{z})|$  is small compared to the absolute value of the right-hand side of eq. (6). This assumption is well satisfied when the 3D diffusion constant of the protein is large compared to that of the label ( $D_p \gg D_l$ ), also for low amplitude potentials ( $\tilde{U}_0 \ll 1$ ), as in this case  $l_0(u) \sim 1$ , and for large amplitude potentials ( $\tilde{U}_0 \gg 1$ ), as  $I_0^2(\tilde{U}_0/2) \gg 1$ . For a potential of intermediate amplitude, the assumption  $a \ll z_0$  roughly ensures that  $|\phi(u, \tilde{z})|$  is small compared to the term proportional to  $\tilde{U}_0$  on the right-hand side of eq. (6), and thus that  $h_1^*(u, \tilde{z})$  remains a good approximation of  $h_1(u, \tilde{z})$ . As demonstrated in appendix B, this approximation fails when  $a/z_0 \geq 1$ .

Assuming  $a \ll z_0$ , an analytical expression of the linear diffusion constant  $D$  of the protein-label complex can thus be found inserting  $h_1^*(u, \tilde{z})$  in eq. (4). We found

$$D = \frac{D_p D_l}{D_p + D_l I_0^2(\tilde{U}_0/2)}.$$

This expression for  $D$  can be interpreted by rewriting it in the following form:

$$\frac{1}{D} = \frac{1}{D_l} + \frac{I_0^2(\tilde{U}_0/2)}{D_p}. \quad (9)$$

Setting  $D = k_B T / \zeta$ , which defines an effective friction constant  $\zeta$  for the protein-label complex, eq. (9) gives  $\zeta = \zeta_l + I_0^2(\tilde{U}_0/2) \zeta_p$ . This relation can be interpreted as follows:  $\zeta$  is the sum of the friction constant of the label in solution and that of the protein diffusing in the potential. This result is valid under the assumptions that the amplitude  $z_0$  of the Brownian fluctuations of the label is large compared to the size of the protein diffusion step  $a$  and that the relaxation time  $\tau_R$  of the protein-label linker is small compared to the time  $\tau_S$  the protein needs to slide over  $a$ . Since a friction coefficient is proportional to viscosity, one can also interpret this result by considering that the protein diffuses in a medium of viscosity  $I_0^2(\tilde{U}_0/2) \eta$ , while the label, although following the protein in its diffusive motion, still moves in a medium of viscosity  $\eta$ .

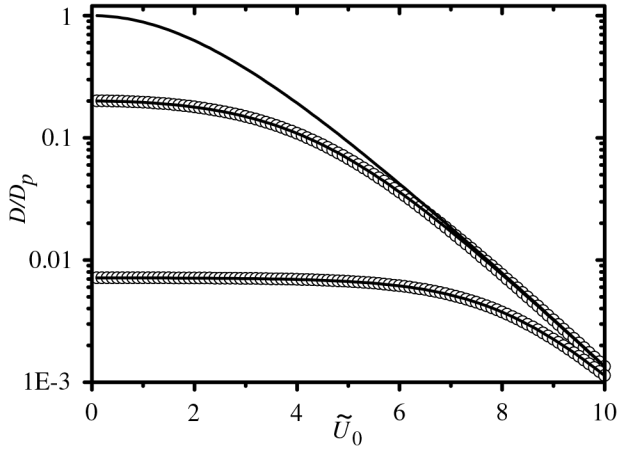
We now consider some cases for which the expression of  $D$  given by eq. (9) is consistent with known results. For an unlabeled protein diffusing along the DNA, our model predicts  $D = D_p / I_0^2(\tilde{U}_0/2)$ , which is the exact solution of the Fokker-Planck equation in the case of the periodic potential  $U(x)$  [30]. When  $\tilde{U}_0 \ll 1$ , the friction coefficient of the protein-label complex is  $\zeta = \zeta_l + \zeta_p$ , which is also the expected result. Finally, for potentials of large amplitude ( $\tilde{U}_0 \gg 1$ ),  $I_0^2(\tilde{U}_0/2) \gg 1$ , thus  $D \simeq D_p / I_0^2(\tilde{U}_0/2)$ , and the diffusion coefficient of the protein-label complex does not depend on the size of the label. Even if not explicit, this assumption has been frequently made in many single-molecule experiments. For instance, when the diffusion of membrane receptors labeled with quantum dots was evaluated in living cells using single-molecule fluorescence microscopy [31], the diffusive behavior of a receptor was considered to be not impacted by the size of the QD, though bigger than that of the receptor. This assumption is in fact justified as the viscosity of a cell membrane is much larger than that of the buffer surrounding the cell, whose viscosity is close to that of water [32]. Though in this case the receptors diffuse in a 2D space, while our analysis is related to linear diffusion, the conclusions should qualitatively be the same.

## 2.4 Numerical estimation of the diffusion coefficient and comparison with analytical data

The diffusion constant  $D$  of a labeled protein can also be numerically estimated by solving eq. (6) following a numerical scheme proposed by [27]. Briefly, the function  $h_1(u, \tilde{z})$  is expanded in Fourier series

$$h_1(u, \tilde{z}) = \sum_{m=-M}^{+M} \sum_{n=-N}^{+N} C_{m,n} e^{-i 2\pi m u} e^{-i 2\pi \epsilon_L n \tilde{z}}, \quad (10)$$

with  $\epsilon_L = z_0 / L_z$ , where  $L_z$  is the size of the box over which the dependence of  $h_1$  on  $\tilde{z}$  is evaluated and  $C_{m,n}$



**Fig. 2.** Diffusion constants  $D$  (analytical result, solid lines) and  $D^n$  (numerical result, circles) of a protein-label complex as a function of the amplitude  $\tilde{U}_0$  of the sinusoidal potential, for  $a/z_0 = 0.2$ . We plot the ratios  $D/D_p$  and  $D^n/D_p$ , where  $D_p$  is the 3D diffusion constant of the protein. Upper curve: unlabeled protein. Middle curves: protein attached to a label whose size is 4 times the protein size. Analytical solution  $D$  (solid line) and numerical solution  $D^n$  (circles) coincide whatever the amplitude  $\tilde{U}_0$ . Lower curves: protein attached to a label whose size is 139 times the protein size. Similarly, analytical solution  $D$  (solid line) and numerical solution  $D^n$  (circles) coincide whatever the amplitude  $\tilde{U}_0$ . For large values of  $\tilde{U}_0$ ,  $D/D_p$  tends to the upper curve (unlabeled protein), whatever the size of the label.

are complex Fourier coefficients. The expression (10) of  $h_1$  is inserted in eq. (6), and  $2(N+1)(M+1)$  linear algebraic equations on  $C_{m,n}$  are obtained by identifying terms with equal frequencies. The coefficients  $C_{m,n}$  are then obtained by inverting the related complex matrix using Mathematica [33], with  $M = 10$ ,  $N = 15$  and  $\epsilon_L = 0.1$ . The boundary conditions for the variables  $u$  and  $\tilde{z}$  were treated similarly as in [27], as well as the normalization of  $h_1$ . A numerical value  $D^n$  of the diffusion constant is then obtained by calculating numerically the integral in eq. (4).

In fig. 2 we compare the numerical  $D^n$  with the analytical expression of  $D$  given by eq. (9). We plot both  $D/D_p$  and  $D^n/D_p$  as a function of  $\tilde{U}_0$  for an unlabeled protein and for a protein attached to a label whose size is 4 or 139 times the size of the protein, as these three cases will be discussed later (see sect. 3). Figure 2 shows that  $D$  coincides with  $D^n$  whatever the label size, which validates our analytical approach. Numerical results shown in fig. 2 were obtained with  $a/z_0 = 0.2$ , and we verified that these results do not depend on the ratio  $a/z_0$  as long as  $a/z_0 \ll 1$  (see fig. 4 in appendix B). When  $a/z_0 \geq 1$ , the numerical  $D^n$  departs from the analytical result  $D$  and, when  $a/z_0 \gg 1$ , the protein-label complex behaves like a solid (see appendix B).

Finally, fig. 2 shows that for large values of  $\tilde{U}_0$  the ratio  $D/D_p$  tends to the value associated to an unlabeled protein, whatever the label size. It confirms that, for an interaction potential with a large amplitude, the dynamics of the complex is governed by that of the protein.

## 2.5 Diffusion coefficient of a labeled protein diffusing in a random potential

The sinusoidal potential  $U(x)$  we chose to mimic the DNA-protein interaction, although unrealistic, has the advantage of making the calculations possible. Yet the DNA-protein potential is sequence-dependent, and thus its amplitude varies during the protein motion. A more realistic modeling of the DNA-protein interaction is a random potential with a roughness  $\sigma$  [28], the characteristic length over which the potential varies along the DNA remaining  $a$  (one DNA base-pair). Setting  $\tilde{\sigma} = \sigma/k_B T$ , the linear diffusion coefficient of the protein in such a random potential is given by [28]:  $D = D_p/F(\tilde{\sigma})$ , with  $F(\tilde{\sigma}) = (1 + \tilde{\sigma}^2/2)^{-1/2} \exp(7\tilde{\sigma}^2/4)$ . This expression for  $D$  is similar to that obtained for a sinusoidal potential,  $F(\tilde{\sigma})$  replacing  $I_0^2(\tilde{U}_0/2)$ . Again, one can interpret this expression of  $D$  as the diffusion coefficient of a protein diffusing in a medium of viscosity  $F(\tilde{\sigma})\eta$ . Therefore we assume that, in the case of a labeled protein diffusing in a random potential, eq. (9) can be rewritten as

$$\frac{1}{D} = \frac{1}{D_l} + \frac{F(\tilde{\sigma})}{D_p}, \quad (11)$$

which will be used to interpret our experimental data.

## 3 Results

### 3.1 Experimental values of the linear diffusion constant of EcoRV

Our modeling effort was motivated by the results of recent single-molecule experiments we performed in order to characterize the facilitated diffusion of EcoRV. The fluorescence microscopy set-up and the protocol for these experiments are detailed in [21]. Through these experiments, we showed, by direct observation of the enzyme interacting with the DNA, that EcoRV can slide along the DNA. In addition, we measured the linear diffusion constant  $D$  of EcoRV labeled with three different fluorescent labels of increasing sizes:

- i) An organic dye called Cy3B (PA63131, GE Healthcare). Cy3B is a bright version of the standard Cy3 dye, which was attached to the enzyme through a reaction between a Cy3B-maleimide and a EcoRV cysteine at position 58. We measured the hydrodynamic radius of EcoRV-Cy3B using Fluorescence Correlation Spectroscopy (FCS) and found  $r(\text{EcoRV} - \text{Cy3B}) = 3.9 \pm 0.1 \text{ nm}$  [21]. Since the molecular weight of Cy3B ( $\sim 0.8 \text{ kDa}$ ) is small compared to that of EcoRV (58 kDa), we state that the radius of EcoRV-Cy3B is also the radius of unlabeled EcoRV, and thus that the linear diffusion constant  $D$  we measured for the EcoRV-Cy3B complex coincides with that of EcoRV.

**Table 1.** Experimental diffusion constants  $D$  of EcoRV (radius  $r_p = 3.9 \pm 0.1$  nm) attached to three labels of increasing radius  $r_l$ .

	EcoRV-Cy3B	EcoRV-savCy3	EcoRV-QD
$r_l$ (nm)	$\sim 0.5$	$3.9 \pm 0.2$	$15.4 \pm 0.3$
$D$ ( $10^{-2} \mu\text{m}^2 \text{s}^{-1}$ )	$1.1 \pm 0.2$	$1.2 \pm 0.1$	$0.32 \pm 0.02$

- ii) A protein, streptavidin-Cy3 (016-160-084, Jackson), called from now as savCy3, which is streptavidin labeled with the organic dye Cy3. To attached this label to EcoRV, the enzyme was first biotinylated through a reaction between a Maleimide-PEG<sub>2</sub>-Biotin reagent (Pierce) and a cysteine at position 58 [21]. The strong affinity between biotin and streptavidin provided an efficient labeling of EcoRV. According to the provider, the length of the PEG spacer arm in the maleimide reagent is about 3 nm, which ensures that the protein and the label are linked by a flexible molecule, as stated in our model. We measured the hydrodynamic radius of savCy3 by FCS and found  $r(\text{savCy3}) = 3.9 \pm 0.2$  nm, which is also the radius of EcoRV. This result was predictable as the molecular weight of streptavidin (53 kDa) is similar to that of EcoRV (58 kDa).
- iii) 605 nm streptavidin-coated quantum-dots from Invitrogen (Q10001MP), designed from now on as QD. QD was attached to EcoRV in the same way savCy3 was. We measured the hydrodynamical radius of QD by FCS and found  $r(\text{QD}) = 15.4 \pm 0.3$  nm, in agreement with the value given by the provider.

The linear diffusion constants  $D$  for EcoRV labeled with these fluorophores was determined as previously described [21], using a PIPES buffer at  $p\text{H} = 6.8$ . The experimental values of  $D$  labeled with Cy3B [21], savCy3 (see appendix C) and QD [26] are listed in table 1.

The diffusion constant of EcoRV-savCy3 is thus similar to that of EcoRV-Cy3B, although the size of streptavidin is similar to that of EcoRV and much larger than that of Cy3B. The value of  $D$  for EcoRV labeled with a QD, the size of which is 4 times that of EcoRV, is only one third of that of EcoRV. We next confront these experimental data with the results predicted by our model.

### 3.2 Comparison with the model

The comparison between the experimental diffusion constants and the results provided by our model requires to estimate  $F(\tilde{\sigma})$ ,  $D_p$  and  $D_l$  (eq. (11)). For this we assume, as in [21], that the enzyme tracks the DNA helix when sliding along the DNA, *i.e.* that the motion of EcoRV is a rotation-coupled linear diffusion, an assumption that has been recently reinforced for other DNA-binding proteins [24]. The value  $F(\tilde{\sigma}) \sim 74$ , which corresponds to a roughness  $\sigma = 1.65 k_B T$  of the DNA-EcoRV nonspecific interaction potential, is deduced from  $D_{\text{Cy3B}} = 1.1 \pm 0.2 \cdot 10^{-2} \mu\text{m}^2 \text{s}^{-1}$  [21]. For estimating  $D_p$ ,

the friction coefficient  $\zeta_p$  of the free enzyme must be modified to take into account the rotation of the protein-label complex [34, 35]

$$\zeta_p^{\text{rot}} = \zeta_p \left( 2\pi \frac{r_p}{h} \right)^2 \left[ \frac{4}{3} + \left( \frac{r_{OC}}{r_p} \right)^2 \right],$$

where  $r_p$  is the radius of the protein,  $h = 3.4$  nm, and  $r_{OC}$  the distance between the center of mass of protein and the DNA axis. The crystallographic structure of the nonspecific DNA-EcoRV complex [36] suggests that  $r_{OC}$  is small compared to  $r_p$ , and thus can be omitted. From now we note  $D_p^{\text{rot}} = k_B T / \zeta_p^{\text{rot}}$  the diffusion constant of an unlabeled enzyme that spins while diffusing freely along the DNA. We also estimate  $D_l$  by similarly modifying the friction coefficient  $\zeta_l$  of the label to account for a rotation-coupled diffusion of the label at the distance  $r_{OC} = r_p + r_l$  from the DNA axis, where  $r_l$  is the radius of the label, and we note  $D_l^{\text{rot}} = k_B T / \zeta_l^{\text{rot}}$  the diffusion constant of a label that spins while diffusing.

The diffusion coefficient  $D$  of a protein-label complex tracking the DNA helix is then derived from eq. (11)

$$\frac{D}{D_p^{\text{rot}}} = \frac{1}{D_p^{\text{rot}}/D_l^{\text{rot}} + F(\tilde{\sigma})}, \quad (12)$$

where the ratio  $D_p^{\text{rot}}/D_l^{\text{rot}}$  is given by

$$\frac{D_p^{\text{rot}}}{D_l^{\text{rot}}} = \frac{\zeta_l^{\text{rot}}}{\zeta_p^{\text{rot}}} = \left( \frac{r_l}{r_p} \right)^3 \left[ 1 + \frac{3}{4} \left( \frac{r_p + r_l}{r_l} \right)^2 \right].$$

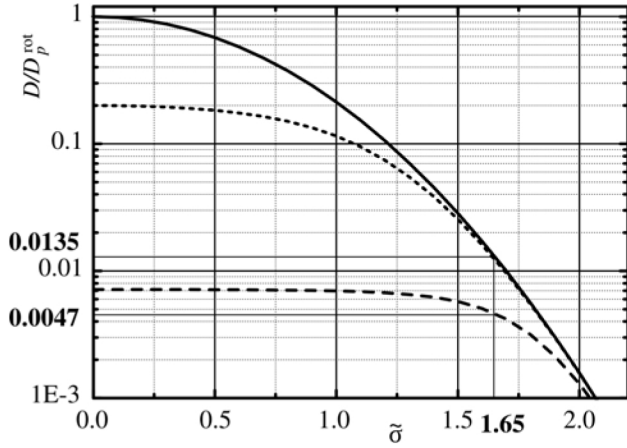
When EcoRV is labeled with Cy3B,  $r_l \ll r_p$ , and thus  $D_{\text{Cy3B}}/D_p^{\text{rot}} = 1/F(\tilde{\sigma} = 1.65) \simeq 0.0135$  (see fig. 3). When EcoRV is labeled with savCy3,  $r_l = r_p$ , then  $D_p^{\text{rot}}/D_l^{\text{rot}} = 4$ . In our model (sect. 2), this amounts to considering that the size of the label is 4 times the size of the protein (see fig. 2). We conclude that  $D_{\text{savCy3}}/D_p^{\text{rot}} = 1/(4 + F(\tilde{\sigma})) \simeq 0.0128$  (see fig. 3), a value which is close to  $D_{\text{Cy3B}}/D_p^{\text{rot}} = 0.0135$ . Thus, our model predicts that the diffusion constant of EcoRV labeled with streptavidin is similar to that of EcoRV labeled with a small organic dye, in agreement with our experimental results.

When EcoRV is labeled with a QD,  $r_l = 4r_p$ , then  $D_p^{\text{rot}} = 139 D_l^{\text{rot}}$ . In our model, this amounts to considering that the size of the label is 139 times the size of the protein (see fig. 2). We conclude that  $D_{\text{QD}}/D_p^{\text{rot}} = 1/(139 + F(\tilde{\sigma})) \simeq 0.0047$  (see fig. 3). Thus, our model predicts  $D_{\text{QD}} \simeq 0.35 D_{\text{Cy3B}}$ , a result which is close to the experimental one ( $D_{\text{QD}} \simeq D_{\text{Cy3B}}/3$ ).

## 4 Discussion

Here we consider alternative models for the enzyme diffusion, and we show that these models do not account for the diffusion constants we measured.

First, we maintain the assumption that the linear diffusion of EcoRV is a rotation-coupled motion, and we consider the protein-label complex as a rigid body. The



**Fig. 3.** Diffusion constant  $D$  of a EcoRV-label complex spinning along DNA as a function of the roughness  $\tilde{\sigma}$  of a random potential. We plot the ratio  $D/D_p^{\text{rot}}$ , where  $D_p^{\text{rot}}$  is the 3D diffusion constant of the protein modified to take into account the rotation of EcoRV. Solid line: unlabeled enzyme with a rotation-coupled linear diffusion. Dotted line: enzyme attached to a fluorescent streptavidin. For  $\tilde{\sigma} = 1.65$ , the diffusion constant  $D_{\text{savCy3}}$  is close to that of the unlabeled enzyme or, equivalently, to that of EcoRV labeled with Cy3B. Dashed line: enzyme attached to a QD. For  $\tilde{\sigma} = 1.65$ , the diffusion constant  $D_{\text{QD}}$  is about one third of the diffusion constant of the unlabeled enzyme.

friction coefficient of the complex diffusing in the DNA-protein interaction potential is then  $\zeta = (\zeta_p^{\text{rot}} + \zeta_l^{\text{rot}})F(\tilde{\sigma})$ , and the diffusion coefficient is now given by

$$\frac{D}{D_p^{\text{rot}}} = \frac{1}{(D_p^{\text{rot}}/D_l^{\text{rot}} + 1)F(\tilde{\sigma})}. \quad (13)$$

This expression of  $D$  differs greatly from that given by eq. (12). Together with  $D_{\text{Cy3B}}/D_p^{\text{rot}} = 1/F(\tilde{\sigma})$ , eq. (13) predicts  $D_{\text{savCy3}}/D_{\text{Cy3B}} = 1/5$ , which is not consistent with our experimental data ( $D_{\text{savCy3}}/D_p^{\text{rot}} \simeq D_{\text{Cy3B}}/D_p^{\text{rot}}$ ). The result is even worse when considering EcoRV labeled with a QD. In this case, eq. (13) predicts  $D_{\text{QD}}/D_{\text{Cy3B}} = 1/140$ , while experimentally we observed  $D_{\text{QD}} \simeq D_{\text{Cy3B}}/3$ . Therefore, considering the protein-label complex as a rigid body that spins during linear diffusion does not account for our experimental data.

Second, we suppose that EcoRV does not rotate when sliding along the DNA. Under this assumption,  $F(\tilde{\sigma}) = D_p/D_{\text{Cy3B}} \simeq 5 \cdot 10^3$ , and thus  $\tilde{\sigma} \simeq 2.3$ . If we consider the protein-label complex as a rigid body, the friction coefficient of the complex diffusing in the DNA-protein interaction potential is  $\zeta = (\zeta_p + \zeta_l)F(\tilde{\sigma})$ , and we expect  $D_{\text{savCy3}}/D_{\text{Cy3B}} = 1/2$  and  $D_{\text{QD}}/D_{\text{Cy3B}} = 1/5$ . Again, these predictions are not consistent with our experimental data. If we consider that the label motion is decoupled from that of the protein, the friction coefficient of the complex is  $\zeta = \zeta_p F(\tilde{\sigma}) + \zeta_l \simeq \zeta_p F(\tilde{\sigma})$ , as  $F(\tilde{\sigma}) \gg \zeta_l/\zeta_p$ . Thus the diffusion constant of the labeled enzyme should not

**Table 2.** Linear diffusion constant of the protein-label complex predicted by different models. Experimentally,  $D_{\text{savCy3}}/D_{\text{Cy3B}} \simeq 1$  and  $D_{\text{QD}}/D_{\text{Cy3B}} \simeq 1/3$ .

Link	Rotation	$D_{\text{savCy3}}/D_{\text{Cy3B}}$	$D_{\text{QD}}/D_{\text{Cy3B}}$
elastic	yes	0.95	0.35
rigid	yes	1/5	1/140
rigid	no	1/2	1/5
elastic	no	$\sim 1$	$\sim 1$

depend on the label size, which, again, is not consistent with our experimental data.

The diffusion constants predicted by all the models discussed here are listed in table 2, which shows that the model which assumes a rotation-coupled diffusion of EcoRV and a label elastically attached to the enzyme best accounts for our experimental data.

Other models could be considered, such as rotation-coupled diffusion of EcoRV alternating with a translational diffusion along the DNA, which would however introduce additional parameters [37]. Finally, the random potential we used to describe the interaction between the protein and the DNA should be refined, as EcoRV is a sequence-specific enzyme which could encounter, during sliding, localized potential wells that mimicks its cognate site and thus may impact on the linear diffusion [38]. However, provided that an effective friction coefficient for an unlabeled site-specific protein can be defined in the presence of such wells, our approach should be relevant to describe the impact of a large label on the protein diffusion. Our goal here was to show that our data are consistent with the predictions of a model in which the motion of the label and that of the protein are decoupled, yet more data would be needed to demonstrate unambiguously that EcoRV rotates when diffusing along the DNA.

## 5 Conclusion

We calculated the diffusion constant of a labeled protein sliding along DNA according to the size of the label, assuming that the motion of the label and that of the protein are decoupled. Our results are in agreement with our previous experimental data concerning the diffusion of the type-II restriction enzyme EcoRV labeled with fluorophores of different sizes, assuming a rotation-coupled diffusion of the enzyme along DNA. Our results may be useful in interpreting experiments performed at the single-molecule level, especially those using large labels to study the diffusive motion of molecules.

We thank Jean-François Joanny, François Graner and Mathieu Coppey for their careful reading and suggestions. We acknowledge support from Centre National de la Recherche Scientifique, École normale supérieure, Egide and CNano IdF-CNRS. The authors belong to the CNRS consortium CellTiss.



## Appendix A. Detailed equations

The function  $f(\tau, u, \tilde{z})$ , whose expression is given by eq. (3), is solution of

$$[\hat{L}_0 + \epsilon \hat{L}_1 + \epsilon^2 \hat{L}_2] f = 0, \quad (\text{A.1})$$

with the following expressions for the operators  $\hat{L}_0$ ,  $\hat{L}_1$  and  $\hat{L}_2$ :

$$\begin{aligned} \hat{L}_0 &= D_p \left[ \frac{\partial^2}{\partial u^2} + \frac{d\tilde{U}}{du} \frac{\partial}{\partial u} + \frac{d^2\tilde{U}}{du^2} \right] \\ &+ D_p \frac{a}{z_0} \left[ 2 \frac{\partial^2}{\partial u \partial \tilde{z}} + \frac{d\tilde{U}}{du} \frac{\partial}{\partial \tilde{z}} + \tilde{z} \frac{\partial}{\partial u} \right] \\ &+ (D_p + D_l) \left( \frac{a}{z_0} \right)^2 \left[ \frac{\partial^2}{\partial \tilde{z}^2} + \tilde{z} \frac{\partial}{\partial \tilde{z}} + 1 \right], \\ \hat{L}_1 &= D_p \left[ 2 \frac{\partial^2}{\partial \tau \partial u} + 2 \frac{a}{z_0} \frac{\partial^2}{\partial \tau \partial \tilde{z}} + \frac{d\tilde{U}}{du} \frac{\partial}{\partial \tau} + \frac{a}{z_0} \tilde{z} \frac{\partial}{\partial \tau} \right], \\ \hat{L}_2 &= D_p \frac{\partial^2}{\partial \tau^2} + D \left[ \frac{\tau}{2} \frac{\partial}{\partial \tau} + \frac{1}{2} \right]. \end{aligned}$$

At the lowest order in  $\epsilon$ , eq. (A.1) reads  $\hat{L}_0 f_0 = 0$ , with  $f_0 = g_0(\tau) h_0(u, \tilde{z})$ . This equation is rewritten as (5) by splitting the operator  $\hat{L}_0$  into three sub-operators, given by

$$\begin{aligned} \hat{L}_{0,u} &= \frac{\partial^2}{\partial u^2} + \pi \tilde{U}_0 \sin(2\pi u) \frac{\partial}{\partial u} + 2\pi^2 \tilde{U}_0 \cos(2\pi u), \\ \hat{L}_{0,uz} &= 2 \frac{\partial^2}{\partial u \partial \tilde{z}} + \pi \tilde{U}_0 \sin(2\pi u) \frac{\partial}{\partial \tilde{z}} + \tilde{z} \frac{\partial}{\partial u}, \\ \hat{L}_{0,z} &= \frac{\partial^2}{\partial \tilde{z}^2} + \tilde{z} \frac{\partial}{\partial \tilde{z}} + 1. \end{aligned}$$

Equation (5) has an analytical solution in the form  $h_0(u, \tilde{z}) = l_0(u) k_0(\tilde{z})$ , with

$$\begin{aligned} l_0(u) &= \frac{1}{I_0(\tilde{U}_0/2)} \exp\left(\tilde{U}_0 \cos(2\pi u)/2\right), \\ k_0(\tilde{z}) &= \frac{1}{\sqrt{2\pi}} \exp(-\tilde{z}^2/2), \end{aligned}$$

where  $I_0$  is the modified Bessel function of the first kind. One can easily check that: i)  $\hat{L}_{0,uz} h_0 = 0$ , ii)  $\hat{L}_{0,u} l_0 = 0$  and  $l_0(u)$  satisfies the expected periodic boundary conditions ( $l_0(u+1) = l_0(u)$ ), iii)  $k_0(\tilde{z})$  is solution of  $\hat{L}_{0,z} k_0 = 0$  with vanishing boundary conditions ( $k_0(\tilde{z} \rightarrow \pm\infty) = 0$  and  $d_z k_0(\tilde{z} \rightarrow \pm\infty) = 0$ ) [27]. The multiplicative constants for  $l_0$  and  $k_0$  have been determined to satisfy the probability normalization [27]

$$\int_0^1 du \int_{-\infty}^{\infty} d\tilde{z} h_0(u, \tilde{z}) = 1.$$

The estimation of  $h_1(u, \tilde{z})$  requires to solve eq. (6) with boundary conditions similar to those given for  $h_0(u, \tilde{z})$ .

Without the coupling term  $\hat{L}_{0,uz} h_1$ , one can find an analytical solution  $h_1^{\text{nc}}(u, \tilde{z})$  of (6), whose expression is given in eq. (7) and which involves the two functions  $l_1(u)$  and  $k_1(\tilde{z})$ , given by

$$\begin{aligned} l_1(u) &= \frac{l_0(u)}{I_0(\tilde{U}_0/2)} \left( \int_0^u dx e^{-\tilde{U}_0 \cos(2\pi x)/2} - u I_0(\tilde{U}_0/2) \right), \\ k_1(\tilde{z}) &= -\frac{z_0}{a} \frac{D_p}{D_p + D_l} \frac{1}{\sqrt{2\pi}} \exp(-\tilde{z}^2/2). \end{aligned}$$

Inserting the ansatz function  $h_1^*(u, \tilde{z})$ , defined in eq. (8), in eq. (6) one gets

$$\begin{aligned} & -\frac{B}{I_0^2(\tilde{U}_0/2)} \frac{a}{z_0} \tilde{z} k_0(\tilde{z}) + (A + B) \frac{a}{z_0} \tilde{z} l_0(u) k_0(\tilde{z}) \\ & + \left( B + A \frac{D_p}{D_p + D_l} \right) \pi \tilde{U}_0 \sin(2\pi u) l_0(u) k_0(\tilde{z}) = \\ & \left( \pi \tilde{U}_0 \sin(2\pi u) + \tilde{z} \frac{a}{z_0} \right) l_0(u) k_0(\tilde{z}). \end{aligned} \quad (\text{A.2})$$

Although no  $(A, B)$  pair is solution of (A.2), we suppose that these two constants satisfy

$$A \frac{D_p}{D_p + D_l} + B = 1 \quad \text{and} \quad A + B \left( 1 - \frac{1}{I_0^2(\tilde{U}_0/2)} \right) = 1.$$

Thus  $A$  and  $B$  only depend on  $D_p$ ,  $D_l$  and  $\tilde{U}_0$ , equality between terms with a sinusoidal dependence in eq. (A.2) is obtained, and, overall, eq. (A.2) is correctly fulfilled when  $a/z_0 \ll 1$ .  $A$  and  $B$  are easily derived from this set of linear equations:

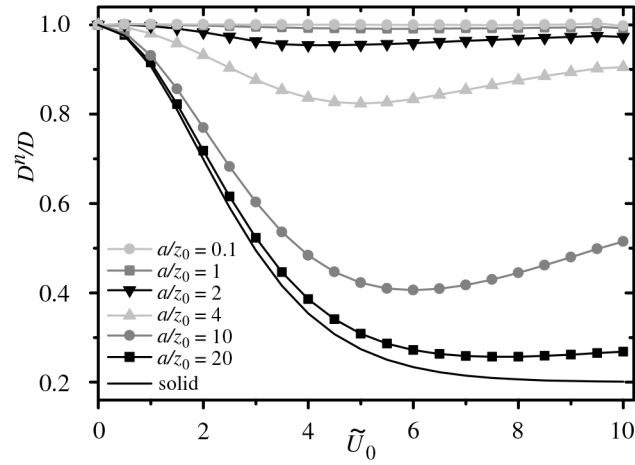
$$A = \frac{D_p + D_l}{D_p + D_l I_0^2(\tilde{U}_0/2)} \quad \text{and} \quad B = \frac{D_l I_0^2(\tilde{U}_0/2)}{D_p + D_l I_0^2(\tilde{U}_0/2)}.$$

The function  $\phi(u, \tilde{z})$ , whose expression is given in sect. 2, is then obtained by inserting  $h_1^*(u, \tilde{z})$ , as defined in eq. (8), in eq. (6).

## Appendix B. Dependence of $D^n/D$ on $a/z_0$

We investigated the dependence of the linear diffusion constant  $D^n$  (numerically estimated) on  $a/z_0$ . The ratio  $a/z_0$  is related to the amplitude of the label fluctuations: when  $a/z_0 \ll 1$ , the amplitude of the Brownian fluctuations of the label are sufficiently large to delocalize the label around the protein, while when  $a/z_0 \gg 1$ , *i.e.* when the spring is hard ( $k \rightarrow \infty$ ), the fluctuations of the label can be neglected, and we expect the protein-label complex to be similar to a rigid body.

We used the numerical approach presented in sect. 2 to calculate  $D^n$  for different values of  $a/z_0$  and we compared  $D^n$  to  $D$ , the linear diffusion constant obtained from an analytical solution of the Fokker-Planck equation solved under the assumption  $a/z_0 \ll 1$ . In fig. 4, we plot the ratio  $D^n/D$  as a function of the amplitude  $\tilde{U}_0$  of the DNA-protein potential for different values of  $a/z_0$ . When  $z_0 \geq a$ ,



**Fig. 4.** Ratio  $D^n/D$  for a savCy3-labeled protein ( $D_l^{\text{rot}} = D_p^{\text{rot}}/4$ ) as a function of the amplitude  $\tilde{U}_0$  of the DNA-protein interaction potential for different values of the parameter  $a/z_0$ . For  $a/z_0 \gg 1$  the diffusion coefficient  $D^n$  of the labeled protein is close to that of a rigid body diffusing along the DNA (solid curve).

$D^n \simeq D$ , as already shown in fig. 3. When  $a/z_0$  increases,  $D^n$  deviates from  $D$  and, for large values of  $a/z_0$ ,  $D^n/D$  converges towards  $D_s/D$  (fig. 4), where  $D_s$  is the diffusion coefficient of a solid composed of a protein stuck to the label:

$$D_s = \frac{D_p D_l}{D_p + D_l} \frac{1}{I_0^2(\tilde{U}_0/2)}.$$

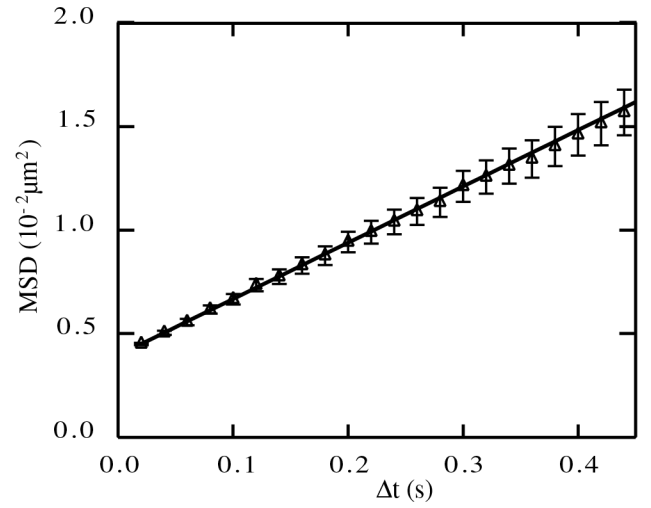
Altogether, these results make our numerical resolution credible.

## Appendix C. Experimental determination of $D_{\text{savCy3}}$

The linear diffusion constant  $D_{\text{savCy3}}$  of EcoRV labeled with savCy3 was determined as previously described in [21]. Briefly, DNA molecules were attached specifically by their ends to the wall of a flow cell, and Total Internal Reflection Fluorescence Microscopy (TIRFM) was used to visualize interactions of enzymes with the DNA. Tracking of proteins moving on DNA allowed us to reconstruct EcoRV trajectories and to compute the Mean Square Displacement (MSD) of the enzyme. We then plot the MSD averaged over all interaction events, and we deduced  $D_{\text{savCy3}}$  by least-square fitting of the first five points of this MSD (fig. 5).

## References

1. A.D. Riggs, S. Bourgeois, M. Cohn, J. Mol. Biol. **53**, 401 (1970).
2. O.G. Berg, R.B. Winter, P.H. von Hippel, Biochemistry **20**, 6929 (1981).
3. P.H. von Hippel, O.G. Berg, J. Biol. Chem. **264**, 675 (1989).



**Fig. 5.** Mean Square Displacement (MSD) for EcoRV attached to fluorescently labeled streptavidin. Since the enzyme slides along the DNA, the MSD (triangles) depends linearly on time. A linear fit (solid curve) gives for the diffusion constant  $D_{\text{savCy3}} = 1.2 \pm 0.1 \cdot 10^{-2} \mu\text{m}^2 \text{s}^{-1}$ .

4. S.E. Halford, J.F. Marko, Nucleic Acids Res. **32**, 3040 (2004).
5. M. Coppey, O. Bénichou, R. Voituriez, M. Moreau, Biophys. J. **87**, 1640 (2004).
6. H.-X. Zhou, Biophys. J. **88**, 1608 (2005).
7. T. Hu, A.Y. Grosberg, B.I. Shklovskii, Biophys. J. **90**, 2731 (2006).
8. B. van den Broek, M.A. Lomholt, S.M. Kalisch, R. Metzler, G.J. Wuite, Proc. Natl. Acad. Sci. U.S.A. **105**, 15738 (2008).
9. M. Sheinman, Y. Kafri, Phys. Biol. **6**, 016003 (2009).
10. C. Loverdo, O. Bénichou, R. Voituriez, A. Biebricher, I. Bonnet, P. Desbiolles, Phys. Rev. Lett. **102**, 188101 (2009).
11. M. Díaz de la Rosa, E.F. Koslover, P.J. Mulligan, A.J. Spakowitz, Biophys. J. **98**, 2943 (2010).
12. J. van Mameren, E.J. Peterman, G.J. Wuite, Nucleic Acids Res. **36**, 4381 (2008).
13. J. Hilario, S.C. Kowalczykowski, Curr. Opin. Chem. Biol. **14**, 15 (2010).
14. J. Gorman, E.C. Greene, Nat. Struct. Mol. Biol. **15**, 768 (2008).
15. H. Kabata, O. Kurosawa, I. Arai, M. Washizu, S.A. Margaron, R.E. Glass, N. Shimamoto, Science **262**, 1561 (1993).
16. A. Granéli, C.C. Yeykal, R.B. Robertson, E.C. Greene, Proc. Natl. Acad. Sci. U.S.A. **103**, 1221 (2006).
17. P.C. Blainey, A.M. van Oijen, A. Banerjee, G.L. Verdine, X.S. Xie, Proc. Natl. Acad. Sci. U.S.A. **103**, 5752 (2006).
18. Y.M. Wang, R.H. Austin, E.C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).
19. J.H. Kim, R.G. Larson, Nucleic Acids Res. **35**, 3848 (2007).
20. J. Gorman, A. Chowdhury, J.A. Surtees, J. Shimada, D.R. Reichman, E. Alani, E.C. Greene, Mol. Cell **28**, 359 (2007).
21. I. Bonnet, A. Biebricher, P.-L. Porté, C. Loverdo, O. Bénichou, R. Voituriez, C. Escudé, W. Wende, A. Pingoud, P. Desbiolles, Nucleic Acids Res. **36**, 4118 (2008).



22. A. Tafvizi, F. Huang, J.S. Leith, A.R. Fersht, L.A. Mirny, A.M. van Oijen, *Biophys. J.* **95**, L01 (2008).
23. N. Shimamoto, *J. Biol. Chem.* **274**, 15293 (1999).
24. P.C. Blainey, G. Luo, S.C. Kou, W.F. Mangel, G.L. Verdine, B. Bagchi, X.S. Xie, *Nat. Struct. Mol. Biol.* **16**, 1224 (2009).
25. N.M. Kad, H. Wang, G.G. Kennedy, D.M. Warshaw, B. van Houten, *Mol. Cell* **37**, 702 (2010).
26. A. Biebricher, W. Wende, C. Escudé, A. Pingoud, P. Desbiolles, *Biophys. J.* **96**, L50 (2009).
27. K.B. Zeldovich, J.-F. Joanny, J. Prost, *Eur. Phys. J. E* **17**, 155 (2005).
28. M. Slutsky, L.A. Mirny, *Biophys. J.* **87**, 4021 (2004).
29. N.S. Heyman, J.M. Burt, *Biophys. J.* **94**, 840 (2008).
30. H. Risken, *The Fokker-Planck Equation: Methods of Solutions and Applications*, 3rd edition, *Springer Series in Synergetics* (Springer, 1996).
31. M. Dahan, S. Lévi, C. Luccardini, P. Rostaing, B. Riveau, A. Triller, *Science* **302**, 442 (2003).
32. P. Fuchs, A. Parola, P.W. Robbins, E.R. Blout, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3351 (1975).
33. Wolfram Research, Inc., *Mathematica*, Version 6.0 for Linux x86, Champaign, Illinois (2007).
34. J.M. Schurr, *Biophys. Chem.* **9**, 413 (1979).
35. B. Bagchi, P.C. Blainey, X.S. Xie, *J. Phys. Chem. B* **112**, 6282 (2008).
36. F.K. Winkler, A.E. Protá, in *Restriction Endonucleases*, Vol. **114**, edited by A. Pingoud (Springer-Verlag, Berlin, Heidelberg, 2004).
37. A.B. Kochaniak, S. Habuchi, J.J. Loparo, D.J. Chang, K.A. Cimprich, J.C. Walter, A.M. van Oijen, *J. Biol. Chem.* **284**, 17700 (2009).
38. A.G. Cherstvy, A.B. Kolomeisky, A.A. Kornyshev, *J. Phys. Chem. B* **112**, 4741 (2008).