Evolutionary advantage of a dissociative search mechanism in DNA mismatch repair

Kyle Crocker[1], **James London**[2], **Andrés Medina**[1], **Richard Fishel**[2], **Ralf Bundschuh**[3]

[1]Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA

[2]Department of Cancer Biology and Genetics, The Ohio State University, Columbus, Ohio 43210, USA

[3]Department of Physics, Department of Chemistry and Biochemistry, Division of Hematology, Department of Internal Medicine, The Ohio State University, Columbus, Ohio 43210, USA

Abstract

Protein complexes involved in DNA mismatch repair diffuse along dsDNA as sliding clamps in order to locate a hemimethylated incision site. They have been observed to use a dissociative mechanism, in which two proteins, while continuously remaining attached to the DNA, sometimes associate into a single complex sliding on the DNA and sometimes dissociate into two independently sliding proteins. Here, we study the probability that these complexes locate a given target site via a semi-analytic, Monte Carlo calculation that tracks the association and dissociation of the sliding complexes. We compare such probabilities to those obtained using a nondissociative diffusive scan in the space of physically realistic diffusion constants, hemimethylated site distances, and total search times to determine the regions in which dissociative searching is more or less efficient than nondissociative searching. We conclude that the dissociative search mechanism is advantageous in the majority of the physically realistic parameter space, suggesting that the dissociative search mechanism confers an evolutionary advantage.

I. INTRODUCTION

DNA mismatch repair (MMR) is a molecular process by which errors in a DNA sequence indicated by mismatched base pairs are corrected. Failure of this process is the cause of many cancers [1], but a complete mechanistic description of the process does not yet exist [1-3]. The MMR process is evolutionarily conserved from prokaryotes to eukaryotes [4-6], so *Escherichia coli* MutS, MutL, and MutH proteins may be productively used to study MMR. In *E. coli*, MMR consists of the following steps: First, MutS recognizes a mismatched site on a DNA strand and associates with the DNA. This MutS then binds MutL from solution, which in turn can bind MutH. MutH then nicks the newly synthesized, erroneous DNA strand. Excision, followed by polymerization and ligation, complete the repair process [5,6].

Here, we describe a quantitative model of the process by which the MMR proteins determine which strand is newly synthesized. Since *E. coli* methylates its DNA strands whenever a GATC base sequence appears, a newly synthesized strand differs from existing strands in that it is not yet methylated. A MutL-activated MutH, therefore, nicks the new strand at

a hemimethylated site, and the strand containing the nick is excised. To create this nick, however, the hemimethylated site must first be recognized. The hemimethylated sites may be thousands of base pairs away from the mismatch (and therefore the place at which the MutS proteins bind to the DNA), so recognition of hemimethylated sites is not a trivial problem. Through single-molecule probing of the MMR process *in vitro*, Liu *et al.* recently found stable toroidal protein clamps (MutS and MutL) diffusing along the DNA strand while transiently associating and dissociating from each other (but, crucially, remaining on the DNA) in order to reach and recognize a hemimethylated site [3]. While these clamps eventually dissociate *from the DNA*, this occurs on a much longer timescale than the protein-protein association and dissociation *on the DNA*. It is therefore the protein-protein association-dissociation diffusion mechanism *on the DNA* that is the subject of our quantitative model presented here.

Several previous studies investigated sliding clamps on DNA. The toroidal, "sliding clamp" protein structure, which we are interested in here, was reported by O'Donnell *et al.* in the context of an *E. coli* polymerase, DNA polymerase III holoenzyme, which is stabilized on the DNA by the β-clamp that encircles the DNA [7]. More recently, Daitchmen *et al.* have used molecular-dynamics simulations to study the diffusion of these protein clamps and report on the way in which the physical properties of the clamps affect the diffusion dynamics [8]. However, all of the previous studies of sliding clamp proteins that we are aware of have focused on individual proteins rather than on interactions between clamps or the search process as a whole.

Protein search processes involving nontoroidal DNA binding proteins have also been studied extensively. Berg *et al.* [9] derived a complete mathematical model of the search process of a DNA binding protein in terms of association and dissociation rates, as well as geometrical considerations that account for sliding, microscopic jumps, intersegmental transfers, and three-dimensional diffusion. Similarly, Lomholt *et al.* [10] produced a mathematical model of a search that takes DNA coiling explicitly into account, which allows an extension of the Berg model to include consideration of intersegmental jumps. Furthermore, Benichou *et al.* [11] found that a combination of three- and one-dimensional diffusion is capable of performing a fast search in the cell nucleus using a fractal description of chromatin. Givaty *et al.* [12] developed a molecular simulation based on electrostatic forces of DNA binding proteins searching DNA and tracked their motion. They found that the most efficient DNA searches consist of ≈20% sliding and ≈80% three-dimensional diffusion. Reingruber *et al.* [13,14], derived exact expressions for the duration of transcription-factor searches for DNA promoter sites that involve fast and slow rates of one-dimensional diffusion in addition to three-dimensional diffusion capable of intersegmental transfer. They found that a coiled DNA conformation is necessary for a fast search, and that the addition of the fast one-dimensional diffusion state decreases search time. Works from Mirny *et al.* [15], Slutsky and Mirny [16], Bauer and Metzler [17], and Zhou [18] also examine the potential role of protein conformational changes in producing a fast search.

The focus of this publication is to quantitatively model the observed protein clamp association-dissociation mechanism present in MMR protein clamp diffusion. While this process is similar to the search mechanisms summarized in the previous paragraph in that

it is characterized by transitions between a slow searching state and a fast nonsearching state, the crucial difference lies in the lifetime distribution of the fast state. In particular, the transition from the dissociated fast state into the slow searching state is governed by three-dimensional diffusion in the nontoroidal proteins, whereas the toroidal structure of the proteins that we consider prevents quick release from the DNA and thus restricts their motion to a single dimension. This structure also prevents transfer between nearby DNA segments [7,8,19]. Additionally, although there have been many previous studies that consider searches that switch between fast and slow one-dimensional diffusion states, ours is the first of which we are aware that studies searches in which the switch between fast and slow states is mediated by the separate diffusion of two distinct clamps. As we demonstrate, this results in a qualitatively different lifetime distribution of the fast diffusing state compared with systems in which the fast diffusing states are caused by internal transitions of the diffusing protein. In particular, while the internal transitions generally follow a Poissonian distribution [13,14], in our system the transition out of the fast state follows a Lévy distribution.

After construction of the quantitative model, we investigate whether the one-dimensional association-dissociation mechanism serves to increase the efficiency with which a hemimethylated site is found, as compared with a more straightforward situation in which the proteins are unable to dissociate from one another (or, equivalently, there is only a single protein complex). If this were the case, it could provide an evolutionary pressure favoring the association-dissociation mechanism. We find that, although the association-dissociation mechanism makes little difference at the observed *E. coli* parameters, there is a much larger section of parameter space in which the association-dissociation mechanism is beneficial as opposed to detrimental when compared with a case in which the proteins do not dissociate.

This paper begins with a summary of the Liu *et al.* experiments underlying our model, including a tabulation of experimental parameters relevant to the model in Sec. II. In Sec. III, the model itself is described both physically and mathematically. Section IV presents our approach to calculating the probability of finding the hemimethylated site from the model. The main findings concerning the probability of finding the hemimethylated site are then presented in Sec. V, and finally the implications of those results are discussed in Sec. VI, along with potential future directions of research in this area. Several of the detailed derivations and validations are relegated to various Appendixes.

II.   EXPERIMENTAL OBSERVATION OF DISSOCIATIVE SEARCH MECHANISM

In this section, we briefly summarize the experimental observations by Liu *et al.* [3] that underlie the model developed here. Additionally, we compile in Table I experimentally measured quantities used to determine values of model parameters, since we refer to these quantities throughout the paper.

In the experiment by Liu *et al.*, interactions of *E. coli* DNA mismatch repair proteins MutS, MutL, and MutH with dsDNA were imaged via total internal reflection fluorescence microscopy. Of particular interest is what will be referred to as the dissociative search

mechanism, so-called because of the many cycles in which MutS and MutL associate into a single complex and then dissociate into two separate complexes before reforming a single complex as they diffuse along the DNA in order to locate the hemimethylated site. (We called this mechanism the "association-dissociation mechanism" in the introduction for clarity, but for the remainder of the paper we switch to the less cumbersome "dissociative mechanism." Note that these "dissociations" refer to the proteins separating from each other while *remaining* on the DNA, rather than dissociation of the proteins from the DNA itself.)

When MutS binds to a mismatch, it forms a stable clamp in the presence of ATP. It then diffuses along the DNA strand. MutL may then bind to MutS, forming a new clamp which diffuses more slowly along the DNA. This slower diffusion implies frequent interaction with the DNA backbone, thus indicating that the MutS-MutL clamp complex is capable of "searching" the DNA for a hemimethylated site [3]. Interestingly, MutL often dissociates from MutS, and the two proteins form two stable and independently diffusing clamps, each of which diffuses along the DNA much more quickly than the MutS-MutL complex and is therefore not interacting with the DNA frequently enough to perform a search [3]. If the dissociated clamps diffuse back into a state in which they are adjacent along the DNA, they are able to re-associate and continue to search the DNA together. This process is represented in cartoon form in Fig. 1(a). Finally, MutH associates with MutL in order to cleave the newly synthesized DNA strand at the hemimethylated site. Measured association lifetimes and diffusion constants for the dissociative search are compiled in Table I. Since some of the values depend on the presence or absence of MutH, the table provides values for both scenarios denoted "SLH" and "SL," respectively. Note that the diffusion of the MutS protein alone is ≈10 times faster than the diffusion of the MutS-MutL complex, and that the diffusion of the MutL protein in the absence of MutH is a factor of ≈20 faster than that of the MutS protein alone. In the presence of MutH, however, MutS and MutL diffuse at similar rates. Furthermore, the addition of MutH does not seem to have a significant effect on the MutS-MutL diffusion constant [3].

The objective of this work is to quantitatively study the effect of this dissociative mechanism on search efficiency. In particular, there are two competing effects of dissociative diffusion on search efficiency that make its overall effect unclear. Since it makes the overall diffusion faster (compared with a system that always remains in the slow, searching state), it increases the region of the DNA that the protein clamps are able to visit. However, since proteins in the dissociated state are unable to actually search the DNA, the amount of DNA actually searched may decrease if the proteins do not re-associate often enough.

III. MODEL

A. Model structure

To determine the efficiency of the dissociative DNA mismatch repair search mechanism, we propose the microscopic model illustrated in Fig. 1, explicitly in Fig. 1(a) and as a state diagram in Fig. 1(c). Fig. 1(b) shows the nondissociative mechanism to which the dissociative mechanism will be compared.

In the dissociative model, the search begins with an associated MutS-MutL protein complex. The initial associated complex then diffuses in one dimension along the DNA with diffusion constant $D_{SL,\mu}$. The MutS-MutL complex dissociates with some average lifetime $\tau_{A,\mu}$ into independent MutS and MutL clamps initially separated by a distance x_d with diffusion constants D_S and D_L, respectively. The individual MutS and MutL clamps diffuse along the DNA until they come into contact again. Once in contact, the proteins either re-associate with each other with an association probability p_A or continue independent diffusion starting from a distance x_d with probability $1 - p_A$. Since p_A is not well known, it will be tested over a broad range of values, but we find that the value of p_A has only a small effect on the overall search (see Sec. IV C 5). This process continues until the hemimethylated site is either found or the MutS clamp falls off the DNA, thereby setting the total search time t_S. The parameters characterizing the search process described by our model are summarized in Table II.

B.   The role of MutH

In the actual MMR scenario, MutH binds to MutL at some point during the search. It then remains complexed with MutL and results in a change of the diffusion constants, as shown in Table I. Since it is unclear at what point in the search MutH enters the process, we consider separately the case in which MutH is present from the beginning of the search ("SLH") and the case in which MutH enters only after the search is complete ("SL"). These can be seen as limiting cases of the actual process, in which MutH joins MutL at some point during the process. For convenience, we will in our descriptions only refer to MutL and MutS with the understanding that MutL may or may not be bound by MutH and consider the absence or presence of MutH simply by choosing parameter values from the third or fourth column of Table I, respectively.

C.   Figure of merit for search efficiency

We quantify the efficiency of the overall search process in terms of the probability that in at least one of n_s total individual searches, a MutS-MutL complex reaches the hemimethylated site at $x = x_{meth}$ within the maximal search time t_s. This probability corresponds to the probability that the HMS is located during the MMR process and is given by

$$P_{t_s, n_s}^{(*)} = 1 - [1 - P(x_{SL} = x_{meth}, t < t_s)]^{n_s}. \tag{1}$$

A single search starts when the MutS-MutL complex forms on the DNA for the first time. While the MutS and MutL clamps may dissociate many times from one another within a DNA search, the search ends when the HMS is found or when MutS or MutL dissociate *from* the DNA. The latter sets the total search time t_s, which will be varied around the experimental value of the MutS lifetime τ_s. We use the MutS lifetime to determine the total search time since it is shorter than the MutL DNA lifetime and thus provides the more stringent cutoff [3,6]. Once a search has ended unsuccessfully, the process moves on to the next search, up to n_s total searches. If the HMS has still not been found after n_s searches, then the overall process is unsuccessful (which occurs with probability $1 - P_{t_s, n_s}^{(*)}$).

While other studies of protein searches on DNA often quantify search efficiency in terms of the time it takes for a protein to locate the target site, we believe that the probabilistic approach described above is more relevant for our system due to the fact that the MutS protein eventually falling off the DNA ends the search attempt irreversibly. Once fallen off, every MutS protein must begin its search again at the mismatch site. Thus, while in other models of search processes, the protein can continue the search indefinitely until it eventually finds its target, in our case the protein must travel the distance from the mismatch to the HMS before it falls off the DNA or the search is in vain. We thus believe the probability $P_{t_s, n_s}^{(*)}$ to be the most physiologically relevant quantity. In Sec. IV, we describe how we calculate $P_{t_s, n_s}^{(*)}$.

D. Simplifying assumptions

To create this model, we have made a number of simplifying assumptions. First, we only consider the case when there are two protein clamps (either MutS and MutL, or MutS and a MutL-MutH complex) on the DNA. This is not the case in the cell but serves as a first approximation. Similarly, there may be other molecules on the DNA aside from the MMR proteins involved in an individual search, which could also lead to complications. We also assume that a HMS is recognized each time a MutS-MutL or MutS-MutL-MutH clamp reaches it. While consideration of these complications is beyond the scope of the current work, we discuss qualitative expectations of the effect of these issues and discuss strategies for the expansion of our work to take these complicating factors into account in Sec. VI.

IV. CALCULATION OF SUCCESSFUL HEMIMETHYLATED SITE SEARCH PROBABILITY

In this section, we describe how we calculate our figure of merit of search efficiency, namely, the successful hemimethylated site search probability $P_{t_s, n_s}^{(*)}$ defined in Sec. IIIC. As a reminder, a successful search is defined as one in which the MutS-MutL complex visits the hemimethylated site before the cutoff time t_s passes at least once among n_s independent search attempts.

A. Successful hemimethylated search probability of nondissociative search

The baseline with which we compare $P_{t_s, n_s}^{(*)}$ is the equivalent probability for searches in which the clamps remain associated with each other for the entire search (pure one-dimensional diffusion), which will be referred to as "nondissociative" or "purely diffusive" searches and denoted $P_{t_s, n_s}^{(0)}$. These searches represent the limiting case in which the microscopic association lifetime $\tau_{A,\mu} \leftrightarrow \infty$, and they are illustrated in Fig. 1(b).

The probability of a successful nondissociative search can be derived analytically following Redner [20]. We start with the diffusion equation in the case of a MutS-MutL clamp:

$$\frac{\partial \, p(x,t \mid x_0)}{\partial t} = D_{SL} \frac{\partial^2 \, p(x,t \mid x_0)}{\partial x^2},$$

(2)

where D_{SL} is the diffusion constant associated with the MutS-MutL clamp and x_0 is the position along the DNA at which the nondissociative search begins. $p(x, t \mid x_0)dx$ is the probability that the clamp will be searching a position within dx around x at time t.

To consider the probability that the search reaches the hemimethylated site x_{meth}, we first solve this differential equation in the presence of an absorbing boundary condition at x_{meth}. Mathematically, this condition is expressed as $p(x_{meth}, t) = 0$, requiring that the MutS-MutL complex has not arrived at the hemimethylated site. Using the method of images, this solution is given by

$$p(x,t \mid x_0, x_{meth}) = \frac{1}{\sqrt{4\pi D_{SL}t}} \left[\exp\left(-\frac{(x-x_0)^2}{4D_{SL}t} \right) - \exp\left(-\frac{[x-(2x_{meth}-x_0)]^2}{4D_{SL}t} \right) \right],$$

(3)

which represents the spatial probability density of the search at some time t under the assumption that the search started at x_0 and has not yet reached x_{meth}. Note that this calculation treats the DNA as semi-infinite, since the search ends when the hemimethylated site is reached, and the clamp thus can never reach positions with $x > x_{meth}$. The total probability that at time t a clamp has only searched positions x such that $x < x_{meth}$ (i.e., the probability that the HMS has not been located) is obtained by integrating the probability density over all positions $x < x_{meth}$:

$$P(x < x_{meth}, t) = \int_{-\infty}^{x_{meth}} p(x,t \mid x_0, x_{meth})dx$$
$$= \mathrm{erf}\left(\frac{x_{meth} - x_0}{\sqrt{4D_{SL}t}} \right).$$

(4)

The probability that x_{meth} *has* been located by a single search is, therefore,

$$P_{t_s,1}^{(0)} = 1 - P(x < x_{meth}, t_s) = 1 - \mathrm{erf}\left(\frac{x_{meth} - x_0}{\sqrt{4D_{SL}t_s}} \right),$$

(5)

where the superscript (0) indicates that the probability is for a nondissociative search. The overall probability for at least one out of n_s searches being successful is given, in analogy to Eq. (1), by

$$P_{t_s,n_s}^{(0)} = 1 - \left[1 - P_{t_s,1}^{(0)} \right]^{n_s}.$$

(6)

B. Successful hemimethylated search probability of dissociative search

1. Association and dissociation event stepping simulation—To use the model described in Sec. III to calculate successful-search probabilities, we develop a Monte Carlo approach that samples from analytic one-dimensional diffusion probability distributions. This calculation breaks the problem of determining the successful-search probability into the cumulation of the probabilities that each individual microscopic association [the state shown in the top right of Fig. 1(c)] identifies the hemimethylated site.

The probability that an individual microscopic association reaches the HMS can be determined analytically [see Eq. (8) below] if the distance between the position at which the clamps associate x_0 and the hemimethylated site x_{meth} is known. Similarly, the probability distributions of the subsequent microscopic re-association position and time (i.e., the position and time at which MutS and MutL come back together following dissociation from each other) can be determined analytically and are given in Eqs. (11) and (12). In principle, this conceptual framework produces an analytic expression for the successful-search probability involving iterative convolution integrals. In practice, however, this expression is too complex to be used to compute values directly. In particular, we found that the most straightforward way to calculate the many integrals over diffusion position and association lifetime probability distributions was to randomly sample from these distributions many times. Each set of random samples produces a probability of either 1 or 0 that the hemimethylated site was successfully reached, and the average of many of these sets gives the successful-search probability $P(x_{\mathrm{SL}} = x_{\mathrm{meth}}, t < t_s)$ and therefore the overall successful-search probability $P_{t_s, n_s}^{(*)}$.

Another way to think of this iterative random sampling is to imagine that each set of random samples represents a path that the protein clamps can take along the DNA strand which results in either a successful or unsuccessful search. Each path occurs with a frequency proportional to its probability, and therefore setting the successful searches to 1 and the unsuccessful searches to zero and taking the average of many such searches produces the successful-search probability.

The following algorithm is used to carry out this experiment and will be called the association and dissociation event stepping simulation (ADESS):

(1) The clamps start immediately adjacent to each other. We set the starting position of the clamps to $x_0 = 0$, the step counting index to $i = 0$, and the elapsed time to $t_e = 0$. Input a position to search for on a one-dimensional axis (designated the "hemimethylated site" or simply "x_{meth}") representing its distance from the initial MutS-MutL association site on the dsDNA. Also choose a cutoff time t_s, representing dissociation of the MutS clamp from the dsDNA.

(2) Decide whether the adjacent clamps associate by sampling randomly from a uniform distribution between 0 and 1 and comparing the result to the probability p_A that adjacent clamps will associate. If the clamps do not associate, go to step 7.

(3) If the last association position is to the right of the HMS ($x_i > x_{\text{meth}}$), mirror the association position at the HMS by setting $x_i \equiv x_{\text{meth}} - (x_i - x_{\text{meth}})$. Due to the symmetry of the system, this operation does not impact the outcome of the search but avoids having to distinguish between cases with $x_i < x_{\text{meth}}$ and cases with $x_i > x_{\text{meth}}$ in what follows.

(4) Randomly select, using the method of inverse transform, an association lifetime from the probability distribution given by

$$p_{\text{assoc}}(t) = \tau_{\text{A},\mu}^{-1} \exp(-t / \tau_{\text{A},\mu}), \tag{7}$$

where $\tau_{\text{A},\mu}$ is the average microscopic association lifetime of the clamps. This represents the time for which the clamps are diffusing together during this association. Denote this time as t_{assoc}. If $t_e + t_{\text{assoc}} > t_s$ this is the last association period of the search and thus the last opportunity to find the HMS. In that case, since the search ends at t_s, shorten the length of the last association period to the remaining time $t_{\text{assoc}} \equiv t_s - t_e$. In either case, increment t_e by t_{assoc}.

(5) Decide whether the hemimethylated site x_{meth} has been reached given the previous association position and lifetime by sampling randomly from a uniform distribution between 0 and 1 and comparing the result r to the probability

$$P_{\text{find}}(t_{\text{assoc}}) = 1 - \text{erf}\left(\frac{x_{\text{meth}} - x_i}{\sqrt{4 D_{\text{SL},\mu} t_{\text{assoc}}}}\right) \tag{8}$$

that the site has been reached, where x_i is the previous association position and $D_{\text{SL},\mu}$ is the diffusion rate of the associated clamps. Note that this is simply Eq. (5) evaluated at $t = t_{\text{assoc}}$ with start position x_i. There are now three possibilities:

(1) If $r < P_{\text{find}}(t_{\text{assoc}})$, x_{meth} has been reached, so we proceed to step 9 and set the search value to 1.

(2) If $r < P_{\text{find}}(t_{\text{assoc}})$ and $t_e \geq t_s$, x_{meth} has not been reached and the search is over. We therefore proceed to step 9 and set the search value to 0.

(3) If $r > P_{\text{find}}(t_{\text{assoc}})$ and $t_e < t_s$, x_{meth} has not been reached, but the search continues. We therefore proceed to step 6.

(6) Use the previous association position x_i and lifetime to randomly select the next dissociation position x_{i+1} from the probability density function of Eq. (3) at $t = t_{\text{assoc}}$, with an additional normalization factor C that ensures that the probability that x_{meth} has not been reached is 1 at time $t = t_{\text{assoc}}$. This factor is necessary because we have already determined in the previous step that the hemimethylated site has not been reached:

$$p(x_{i+1}, t_{\text{assoc}} \mid x_i, x_{\text{meth}})$$
$$= \frac{C}{\sqrt{4\pi D_{\text{SL}} t_{\text{assoc}}}} \left[\exp\left(-\frac{(x_{i+1} - x_i)^2}{4D_{\text{SL}} t_{\text{assoc}}}\right) \right.$$
$$\left. - \exp\left(-\frac{[x_{i+1} - (2x_{\text{meth}} - x_i)]^2}{4D_{\text{SL}} t_{\text{assoc}}}\right) \right], \tag{9}$$

where

$$C = \frac{1}{\text{erf}\left[(x_{\text{meth}} - x_i) \,/\, \sqrt{4D_{\text{SL}, \mu} t_{\text{assoc}}}\right]} . \tag{10}$$

We increase i by one to indicate that the x_{i+1} determined here is the new position of the two newly dissociated clamps.

(7) Use the dissociation lifetime distribution

$$p_{\text{dissoc}}(t) = \frac{x_d}{\sqrt{4\pi D_{\text{rel}} t^3}} \exp\left[-\frac{x_d^2}{4D_{\text{rel}} t}\right] \tag{11}$$

to determine how long the clamps remain dissociated (see Appendix A). Here, x_d is as before the initial distance of the clamps following dissociation, and D_{rel} is the diffusion constant associated with the fluctuation of the distance between the clamps. Since each clamp is diffusing independently, the distance between them is also diffusing without bias in a particular direction. Denote this chosen lifetime t_{dissoc} and increment the total elapsed time t_e by t_{dissoc}. If the cutoff time has been reached ($t_e \geqslant t_s$) mark the search as unsuccessful and go to step 9.

(8) Using the lifetime chosen in the previous step t_{dissoc}, select the next possible association position x_{i+1} from the distribution of positions at which the relative position of the clamps returns to zero. This distribution is given by the solution to the unbounded diffusion equation with constant D_{CM} associated with the diffusion of the "center of mass" of the dissociated clamps (see Appendix A). In particular,

$$p_{\text{return}}(x_{i+1} \mid x_i, t_{\text{dissoc}})$$
$$= \frac{1}{\sqrt{4\pi D_{\text{CM}} t_{\text{dissoc}}}} \exp\left[-\frac{(x_{i+1} - x_i)^2}{4D_{\text{CM}} t_{\text{dissoc}}}\right] . \tag{12}$$

Increase i by one and return to step 2.

(9) Perform many such searches and assign a value of 1 to all those that are successful and 0 to those in which the cutoff time is reached without success. Take the average value of all of these searches to determine the successful-search probability. Divide the trials into 10 independent blocks of equal number of trials and calculate the search probability for each block to determine standard error.

This procedure gives $P_{t_S, 1}^{(*)}$, and the overall search probability for n_S searches is given by

$$P_{t_S, n_S}^{(*)} = 1 - \left(1 - P_{t_S, 1}^{(*)} \right)^{n_S}. \tag{13}$$

As noted in Sec. IV A, if this calculation is performed in the limit $\tau_{A,\mu} \to \infty$ in step 4, then $P_{t_S, n_S}^{(*)} \to P_{t_S, n_S}^{(0)}$.

2. Base pair stepping simulation confirms continuum approximation is appropriate—In our derivation above, we have treated the search process as continuous diffusion. This assumption needs to be justified, as Veksler and Kolomeisky argue that improper application of a continuum approximation can lead to misleading results [21]. Physically, the diffusion of the protein clamps is governed by the free-energy landscape of the DNA, which has base pair periodicity. This suggests that the most accurate description of the search process is somewhere between a discrete and continuous description. To verify that the continuum approximation is appropriate, we compare the continuous ADESS to a discrete simulation called the base pair stepping simulation (BPSS) that tracks the diffusion of the MMR proteins at a base pair level using the Gillespie algorithm [22]. This simulation is discussed in more detail in Appendix B, and it demonstrates excellent agreement with the ADESS, as shown in Appendix Fig. 6. This demonstrates that the continuum approximation employed in the ADESS is appropriate and also validates the two numerical codes against each other.

C. Determination of model parameters

The model described above is written in terms of several microscopic parameters. In this section we will determine the values of these parameters. Some of these parameters can be calculated directly from experimentally measured values and are summarized in Table III. For the remainder, summarized in Table IV, we need to make reasonable assumptions about their values.

The reason that the values of these parameters must be calculated or estimated rather than be measured directly is that the spatial resolution of the experiment is diffraction limited. Since the wavelength of visible light is on the order of hundreds of nm and the protein footprints are on the order of a few nm, the proteins interact on scales below the spatial sensitivity of the experiment. Importantly, this implies that the clamps can *appear* to be associated with each other in the experiment, when they are closer than the spatial resolution of the experiment, even though they may or may not be in actual physical contact. In contrast, in our model we define the associated state as the state in which the diffusion of the clamps is coupled, and the clamps have undergone some conformational change that allows them to interact more closely with the backbone and thus changes their diffusion rate. The dissociated state is the state in which the clamps are diffusing independently of each other. To avoid confusion, we will thus for the purposes of describing the calculation of model parameters from experimental observables denote the state in which the clamps are physically associated as "microscopically associated," the state in which the clamps are

physically dissociated but close enough that their positions are indistinguishable within the resolution of the experiment as "proximate," and the state in which the clamps are physically dissociated and far enough away that their positions are distinguishable as "macroscopically dissociated." In addition, we use "macroscopically associated" to describe clamps that could be either "microscopically associated" or "proximate" and "microscopically dissociated" for clamps that could be either "proximate" or "macroscopically dissociated."

1. Diffusion constants of individual clamps—Since diffusion is scale invariant, there is no reason to believe that the microscopic diffusion constants D_S and D_L of the individual clamps are different from their macroscopically measured values given in Table I. Rewriting the diffusion of two clamps of different diffusion constants in terms of relative and center-of-mass coordinate yields $D_{\text{rel}} = D_S + D_L$ for the diffusion of the relative coordinate and $D_{\text{CM}} = \frac{D_S D_L}{D_S + D_L}$ for the diffusion of the center-of-mass coordinate.

2. Association lifetime and complex diffusion constant—The experiment measures the lifetime $\tau_{\text{A,M}}$ and diffusion constant $D_{\text{SL,M}}$ of macroscopically associated clamps (see Table I). Since macroscopically associated clamps could be either microscopically associated or proximal, a macroscopic association event consists of a sequence of transitions between the microscopically associated state and the proximal state, where only after multiple excursions into the proximal state do the clamps finally reach a distance that can be resolved in the experiment and thus reach the macroscopically dissociated state. Thus, the macroscopically measured lifetime $\tau_{\text{A,M}}$ is an effective lifetime that integrates over many microscopic dissociation and re-association events, and the macroscopically measured diffusion constant $D_{\text{SL,M}}$ is a temporal average of the diffusion constant of microscopically associated clamps $D_{\text{SL},\mu}$ and the diffusion constant of the center of mass of individual clamps D_{CM} during their excursions in the proximal state.

In Appendix C A, we explicitly calculate how the macroscopically measured lifetime $\tau_{\text{A,M}}$ that integrates over multiple microscopic dissociation and re-association events depends on the microscopic parameters of the model. Solving this dependence for the microscopic association time yields

$$
\begin{aligned}
\tau_{\text{A},\mu} &= \frac{1}{[(\langle N_A \rangle - 1)p_A + 1]}\left[\tau_{\text{A,M}} - \frac{x_M(x_M - x_d)}{D_{\text{rel}}} \right] \\
&\approx \frac{\tau_{\text{A,M}}}{[(\langle N_A \rangle - 1)p_A + 1]},
\end{aligned}
\tag{14}
$$

where p_A is the probability that adjacent MutS and MutL clamps will associate, and $\langle N_A \rangle = x_M/x_d$ is the number of times the clamps are in a microscopically adjacent state (making microscopic association possible) in a single macroscopic association. x_d and x_M are the microscopic and macroscopic association distances, respectively, so $x_d \ll x_M$. The approximation in the second line of Eq. (14) holds for our specific values of the parameters as $\frac{x_M(x_M - x_d)}{D_{\text{rel}}} \approx 0.07$ s and $\tau_{\text{A,M}} \approx 30$ s. It implies that the time spent in the proximal state has a negligible contribution to the macroscopic association time due to the speed of the dissociated diffusion, even though the fact that a macroscopic association event consists of

multiple microscopic association events *is* relevant, as evinced by the prefactor $[(\langle N_A \rangle - 1)p_A + 1]^{-1}$. Accordingly (see Appendix C B), the excursions into the proximal state do not have a significant impact on the diffusion constant either due to their short durations. Thus,

$$D_{\text{SL}, \mu} \approx D_{\text{SL}, \text{M}}. \tag{15}$$

3.   Distance from the nearest hemimethylated site—In *E. coli*, hemimethylation occurs at GATC sites [23-25]. Thus, the distance from a random location in the genome to the nearest hemimethylated site is governed by the distance distribution of adjacent GATC sites, shown in Fig. 2 for the genome of *E. coli* K-12 MG1655, NCBI RefSeq assembly: GCF_000005845.2. While in 90% of the cases, the distance between neighboring GATC sites is 500 bp or less, the largest distances between adjacent GATC sites reach all the way to 4960 bp. Since the ability to repair mismatches in the genome should depend on being able to identify the closest hemimethylated site even in the worst case scenario of being right in the middle of the two furthest separated GATC sites, we will report search probabilities over a range of $x_{\text{meth}} = 500$–3000 bp.

4.   Total search time—The search continues until either MutS or MutL dissociates from the DNA. Since the experimentally determined MutS association lifetime $\tau_S = 185 \pm 35$ s is much shorter than the experimentally determined MutL association lifetime $\tau_L = 850 \pm 150$ s, the search time is limited by the MutS association lifetime and thus $t_s \approx \tau_S$.

5.   Dissociation distances, association probability, and number of searches
—Unlike the microscopic association lifetime, microscopic diffusion constants, and the distance from hemimethylated sites, the dissociation distances x_d and x_M, the association probability p_A, and the number of searches n_s are not determined by experimental observables, and thus cannot be calculated directly. Physical arguments, however, allow the estimation of x_d and x_M. In particular, the microscopic dissociation distance, i.e., the distance at which the clamps can be considered as independent, is on the order of $x_d \approx 1$ bp due to the base pair periodicity of the dsDNA free-energy landscape. The macroscopic dissociation distance, i.e., the distance at which two clamps can be resolved in the experiment as being independent, is determined by the diffraction limit and is expected to be about half the wavelength of the fluorescence. For one red and one green fluorophore, this distance is $x_M \approx 300$ nm ≈ 1000 bp.

Similar physical arguments are unable to provide an estimate for the association probability p_A, but arguments can be made to set limits on this parameter. As a probability, the upper limit on p_A is evidently 1. Approximation of a lower limit is made possible by the assumption that $p_A \geq P_{\text{assoc, soln}}$, where $P_{\text{assoc, soln}}$ is the probability that a MutL in solution colliding with a DNA-bound MutS will associate. This assumption is plausible since there is only one dimension (namely, rotation around the DNA) in which MutS and MutL clamps already associated with the DNA must align in order to associate with each other, rather than the three dimensions that must align when MutL is not already associated with the DNA. This assumption combined with published experimental results independent of the

experiments in Ref. [3] suggests that the association probability should be greater than 0.001 (see Appendix D), i.e.,

$$0.001 \leqslant p_A \leqslant 1 \,. \tag{16}$$

In Appendix E we show that the successful-search probability is not very sensitive to the choices of x_d, x_M, or p_A.

Finally, while experiments by Acharya *et al.* [26], Graham *et al.* [27], and Hombauer *et al.* [28] suggest that the DNA mismatch repair process involves multiple MutS-MutL searches for the hemimethylated site, the number of searches *in vivo* is not well known. We therefore perform this calculation with $n_s = 3$ and $n_s = 10$ to approximate this effect.

V. DISSOCIATIVE SEARCH PERFORMANCE

In this section we systematically compare the success probability of the dissociative search mechanism $P_{t_s, n_s}^{(*)}$ defined in Eq. (1) with that of a nondissociative search $P_{t_s, n_s}^{(0)}$, defined in Eq. (6). The goal is to determine if the dissociative search observed in the experiments by Liu *et al.* [3] confers an evolutionary advantage of increased success probability over the simpler nondissociative search. The successful-search probability of the dissociative search is calculated numerically using the ADESS approach presented in Sec. IV B 1, while the successful-search probability of the nondissociative search is given analytically by Eq. (5).

A. Dissociative and nondissociative searches result in similar single-search efficiency for experimental diffusion constants

To obtain an initial intuition about the behavior of the search probabilities, we first look at the single search ($n_s = 1$) probabilities at the experimentally determined values of the diffusion constants. Figure 3 shows the successful-search probability $P_{t_s, 1}^{(*)}$ of the dissociative search and $P_{t_s, 1}^{(0)}$ of the nondissociative search as a function of distance x_{meth} from the hemimethylated site. Here, the subscript t_s indicates the search cutoff time in seconds, and the subscript 1 indicates that the probability indicated is the success probability for only a single trial. Probabilities are shown for various search times t_s within roughly a factor of two from the experimental value of 185 s in both directions. The figure presents results for diffusion constants corresponding to the case where MutH is not associated with MutL and $p_A = 1$ in Fig. 3(a) and for diffusion constants corresponding to the case where MutH is associated with MutL and $p_A = 0.001$ in Fig. 3(b). These are chosen as the two extremes in terms of the differences between dissociative and nondissociative searches, as the results for MutH parameters at $p_A = 0.001$ and for non-MutH parameters at $p_A = 1$ are in between the two cases shown.

Surprisingly, the nondissociative search mechanism somewhat, but systematically, outperforms the dissociative mechanism for this choice of parameters, especially for the case of microscopic association probability $p_A = 0.001$. In spite of these differences somewhat favoring the nondissociative search mechanism, both search mechanisms result in sizable

successful-search probabilities of at least 0.4 for all parameter values explored here and thus both are likely to support successful DNA mismatch repair, in particular because multiple searches further increase this probability.

B.   Dissociative searches confer an advantage across a broad range of diffusion constants under physiological conditions

In the crowded *in vivo* environment, diffusion is likely significantly slower (10–100 fold) than *in vitro* [29]. Additionally the diffusion constants, hemimethylated site distances, and association lifetimes of mismatch repair proteins may vary across organisms. In light of these observations, we next characterize the comparative effect of the dissociative search mechanism across a wide range of possible diffusion rates D_S and D_{SL} at a number of HMS distances x_{meth}. Although we only explicitly vary the diffusion rate and hemimethylated site distance, this can be seen as variation of the dimensionless combination $\sqrt{Dt} \, / \, x$ on which the probability depends [see Eq. (5)]. Thus, we effectively study variations in association time t_s as well as hemimethylated site distance x_{meth} and diffusion rate.

To show the efficacy of the dissociative search mechanism in the scan described above, we calculate and plot the difference between the dissociative overall search probability and the nondissociative overall search probability, given by

$$\delta P_{t_s, n_S} \equiv P_{t_s, n_s}^{(*)} - P_{t_s, n_s}^{(0)}, \tag{17}$$

in Figs. 4 (for $n_s = 3$ searches) and 5 (for $n_s = 10$ searches). Positive values of $\delta P_{t_s, n_s}$ (which correspond to a dissociative mechanism advantage) are shown in solid green, whereas negative values of $\delta P_{t_s, n_s}$ (which correspond to a dissociative mechanism disadvantage) are shown in hatched brown. Regions in which $D_{SL} > D_S$ are blocked out in dotted gray, since these regions are physically unrealistic (the searching state is slower because it must interact more frequently with the DNA).

Overall, Figs. 4 and 5 demonstrate that there is a much broader range of diffusion constants, and therefore hemimethylated site distances and association times, for which the dissociative search mechanism is beneficial for mismatch repair hemimethylated site searches as compared with pure diffusion (indicated by the prevalence of solid green regions in these figures). For ten searches, the absolute difference in probability approaches $\delta P_{185s,10} = 1$ (dark green) for the cases in which dissociation is most favorable, whereas for three searches the maximum difference in probability is more modest, with $\delta P_{185s,3} \approx 0.5$. The case with three searches, however, exhibits a larger regime in which the dissociation mechanism is meaningfully beneficial. Furthermore, there are comparatively small regions of physically realistic parameter space in which the dissociative mechanism is significantly harmful compared with pure diffusion (hatched brown regions). The addition of MutH, the effect of which is shown in the rightmost columns, does not significantly impact these results, although it does slightly decrease the efficacy of the dissociative mechanism.

These results suggest that the dissociative mechanism may be evolutionarily conserved due to its beneficial effect on hemimethylated site searches, as evinced by largely positive values of $\delta P_{t_s, n_s}$ in these scans. We also note that the dissociative mechanism is particularly

favorable at low values of D_{SL} and large x_{meth}, where the purely diffusive search is less likely to succeed. We therefore speculate that this search mechanism may act as insurance against cases in which the hemimethylated site is difficult to locate.

VI. CONCLUSIONS

Experiments by Liu *et al.* [3] observed repeated protein-protein association and dissociation between MutS and MutL sliding clamps involved in identification of a hemimethylated site during DNA mismatch repair in *E. coli*. This naturally raises the question of whether locally searching the DNA in the MutS-MutL associated state and then quickly diffusing along the DNA to a different location when dissociated (i.e., independent MutS and MutL) actually provides an advantage to the search process. Here, we model the dissociative search process, calculate the probability that searching DNA mismatch repair proteins successfully locate the hemimethylated site, and compare the success rate of this dissociative search to the success rate of a simple diffusive search. We find that both search mechanisms are highly efficient for the majority of observed hemimethylated site distances at measured *in vitro* diffusion rates. Perhaps somewhat surprisingly, there is a slight disadvantage in terms of search probability conferred by the dissociative search mechanism for searches at these *in vitro* rates. We note, however, that there may be variation in diffusion rate, association lifetime, and hemimethylated site distance among different organisms and that it has been shown that *in vivo* diffusion can be slower than *in vitro* diffusion by one or two orders of magnitude [29]. Accordingly, we studied the effect of the dissociative search mechanism across a large range of the parameter space of diffusion rates, association lifetimes, and hemimethylated site distances and found that the dissociative mechanism is either neutral or favorable in most cases. Interestingly, we find the most significant advantages of the dissociative search in the parameter regime where the overall search probabilities in the absence of protein-protein dissociation are small. The dissociative search mechanism may therefore function as an insurance mechanism. This suggests that there is an evolutionary advantage conferred by the dissociative search mechanism.

We note, however, that, in addition to its role in the search for a hemimethylated site, MutL acts as a processivity factor for the DNA helicase UvrD, resulting in the excision that is necessary for the progression MMR process [30]. It therefore could be the case that the observed dissociative mechanism is evolutionarily preferred because the dissociation steps allow MutS to load multiple MutL proteins onto the strand, aiding in excision. This alternative hypothesis would be strengthened if further work determines that *in vivo* search efficiency is not increased by the dissociative mechanism, although it is also possible that the dissociative mechanism serves a dual purpose: both increasing search efficiency and loading multiple MutL proteins onto the DNA strand.

A. Mismatch repair is mathematically distinct from transcription factor binding site searches

Having described the MMR search process mathematically, it is instructive to compare it to other biological search processes that have been mathematically characterized. The closest such comparison is likely that of a transcription factor (TF) searching for its DNA

promoter site. While this search is different from the one we consider in that it has a three-dimensional mode, it also consists of a slow one-dimensional diffusion mode that is capable of searching the DNA as well as a fast one-dimensional diffusion mode that cannot search the DNA [13,14]. One could ask, therefore, whether the MMR search is simply a special case of the TF search. However, we find that the rate of switching between fast and slow diffusion modes is qualitatively different. Reingruber, Holcman, and Cartailler model TF searches, in which the switching dynamics is Poissonian [13,14], while we find that, due to the requirement that MutS and MutL proteins must find each other due to random diffusion to enter the slow state, the transition from fast to slow in MMR follows a Lévy distribution [i.e., $p(t) \sim t^{-2/3} \exp(-1/t)$, see Eq. (11)]. The most important difference between these distributions is that the Lévy distribution is fat-tailed, which means that there are long-lived slow states that never switch back to fast states on the relevant timescales. For instance, Cartailler and Reingruber find that, in the absence of three-dimensional diffusion, the optimal search is one such that the average time spent in fast and slow searching states is the same. Since the mean of the fast-state duration in MMR diverges, it is not even possible in this case to achieve such a state. One could imagine requiring that the modal fast time be the same as the average slow time, but in MMR this would require a "fast" diffusion rate that is slower than the "slow" searching rate and is therefore clearly not optimal.

B.   Simplifying assumptions suggest future directions

It is important to emphasize that our treatments of multiple searches and *in vivo* diffusion here are necessarily approximate. A more detailed treatment that accounts for the interactions between proteins that are initially involved in "separate" searches may be a fruitful avenue for future research: in principle the discrete base pair stepping simulation (discussed in Appendix B A) is capable of tracking more than two proteins, but the current computational cost is too high. Additionally, it is likely possible to expand the association and dissociation event stepping simulation to account for more than two proteins and the presence of other molecules on the DNA strand. In particular, the presence of other molecules on the DNA strand may provide a spatial constraint that prevents the occurrence of the of long-lived dissociation events that decrease the efficiency of the dissociative mechanism. Additionally, the association lifetimes of MutS and MutL suggest that there are likely about five MutL proteins on the DNA for every MutS. Since this search is a one-dimensional process, however, each individual protein only interacts with its nearest neighbors. We therefore do not foresee a significant effect specifically related to the *ratio* of proteins on the DNA, simply that the presence of (any) additional molecules may decrease the prevalence of long-lived dissociation events and render the search more efficient.

Another assumption that we make is that the first encounter of a MutS-MutL complex with a hemimethylated site results in its recognition followed by an incision. If recognition of the hemimethylated site is stochastic itself, this will also reduce the overall search probability. Incorporating this effect into our approach and quantifying its consequences on the search probabilities of the dissociative and nondissociative searches will be an interesting direction of future research.

A further potential avenue of study is the effect of a more physiological environment on the diffusion constants of the proteins. We note that the *in vivo* diffusion constants are likely to be smaller than the measured *in vitro* coefficients, but are not able to quantitatively predict the magnitude of this decrease. A study that determines the actual *in vivo* diffusion constants of mismatch repair proteins could therefore be very useful. Similarly, determination of diffusion constants in systems other than *E. coli* would be interesting.

C. Mathematical framework is likely applicable to other biological processes

Beyond describing the specifics of the MutS-MutL search process, our approach here is likely to be applicable to other diffusive processes along DNA in biology. For instance, Zessin *et al.* observe a fast and slow diffusion rate of proliferating cell nuclear antigen (PCNA), which is a eukaryotic protein similar to a β clamp that also forms a clamp structure during association with DNA [31]. Eukaryotes also exhibit three homologs to both MutS and MutL [6], combinations of which are likely to result in a variety of association, dissociation, and diffusion parameters. In this case, the broad parameter space characterized by our analysis may provide insight into MMR in many organisms.

Despite the work still necessary to fully understand the diffusive search process in DNA mismatch repair, we provide a broad characterization of the observed dissociative search mechanism along with a robust analytical and computational framework with which to study diffusion and interaction of protein clamps in DNA mismatch repair that can provide the basis for generalization to other sliding clamp systems in biology.

ACKNOWLEDGMENTS

Appendix

APPENDIX A: TIME AND LOCATION OF RE-ASSOCIATION

In this Appendix we derive the probability densities for the time to re-association and the re-association location of two clamps once they have disassociated from each other. These distributions are used in the ADESS approach to update the time and position after a microscopic excursion of the clamps.

A. Independent diffusion of two sliding clamps

While the two clamps are diffusing independently, the state of the system is given by positions x_S and x_L of the MutS and the MutL clamp along the DNA, respectively. The joint probability distribution for the two clamps follows the diffusion equation

$$\frac{\partial p(x_S, x_L \mid t)}{\partial t} = D_S \frac{\partial^2 p(x_S, x_L \mid t)}{\partial x_S^2} + D_L \frac{\partial^2 p(x_S, x_L \mid t)}{\partial x_L^2}. \tag{A1}$$

By analogy to the Schrödinger equation for a two-body quantum-mechanical problem, this equation can be rewritten in terms of relative and center-of-mass coordinates. In particular, substituting

$$x_{CM} \equiv \frac{\frac{1}{D_S}x_S + \frac{1}{D_L}x_L}{\frac{1}{D_S} + \frac{1}{D_L}}, \tag{A2}$$

$$x_{rel} \equiv x_S - x_L, \tag{A3}$$

$$D_{CM} \equiv \frac{D_S D_L}{D_S + D_L}, \text{ and} \tag{A4}$$

$$D_{rel} \equiv D_S + D_L \tag{A5}$$

yields

$$\frac{\partial\, p(x_{CM}, x_{rel} \mid t)}{\partial t} = D_{CM}\frac{\partial^2\, p(x_{CM}, x_{rel} \mid t)}{\partial x_{CM}^2} + D_{rel}\frac{\partial^2\, p(x_{CM}, x_{rel} \mid t)}{\partial x_{rel}^2}, \tag{A6}$$

which describes independent diffusion of the center of mass coordinate x_{CM} with diffusion constant D_{CM} and the relative coordinate x_{rel} with diffusion constant D_{rel}.

B. Time of re-association

In our model, the microscopic dissociation of the two clamps results in them being separated by the microscopic dissociation distance x_d. Since relative and center-of-mass position diffuse independently, the time to reassociation is the time the freely diffusing relative coordinate x_{rel} takes to reach $x_{rel} = 0$ when starting at $x_{rel} = x_d$. This problem is mathematically equivalent to the problem of the associated clamps reaching the hemimethylated site x_{meth} after starting at some position x_0. We can thus mirror image Eq. (4) [since $x_{rel} = 0$ provides a left boundary for this problem while x_{meth} provided a right boundary in the context of Eq. (4)] and replace x_0 with x_d, x_{meth} with 0, and D_{SL} with D_{rel} to obtain

$$P(t \mid x_{rel} > 0) = \text{erf}\left(\frac{x_d}{\sqrt{4D_{rel}t}}\right) \tag{A7}$$

for the probability that, at time t, the two clamps starting at an initial distance of x_d have not yet touched. The probability density associated with the return of the distance between the two clamps to 0 from a distance of x_d is therefore given by the negative derivative of this probability, i.e.,

$$p_{\text{dissoc}}(t) = -\frac{\partial P(t \mid x_{\text{rel}} < x_{\text{meth}})}{\partial t}$$
$$= \frac{x_d}{\sqrt{4\pi D_{\text{rel}} t^3}} \exp\left[-\frac{x_d^2}{4 D_{\text{rel}} t}\right]. \tag{A8}$$

C. Location of re-association

Since at the time of reassociation the two clamps are at the same location, all we have to do to find the location of this event is to follow the motion of the center-of-mass coordinate x_{CM} during the excursion. Since this is a free diffusion, the probability density for the location of the meeting point x of the two clamps after a time t given that they dissociated at some location x_0 is

$$p_{\text{return}}(x \mid x_0, t) = \frac{1}{\sqrt{4\pi D_{\text{CM}} t}} \exp\left[-\frac{(x - x_0)^2}{4 D_{\text{CM}} t}\right]. \tag{A9}$$

APPENDIX B: VALIDATION OF ADESS APPROACH AND CONTINUUM APPROXIMATION VIA DISCRETE BASE PAIR STEPPING SIMULATION

In this Appendix, we discuss the base pair stepping simulation (BPSS), which is used to validate the ADESS. In particular, it confirms that discrete and continuous calculations agree. We also show that results are not sensitive to factor-of-two changes in the estimated parameter.

A. Base pair stepping simulation

This simulation keeps track of the states of the system and uses Daniel Gillespie's "stochastic simulation" algorithm to transition between states [22]. Briefly, each simulation state consists of an either dissociated or associated MutS and MutL, as well as their position(s) along a DNA strand. Transitions between states occur at rates determined by the microscopic parameters, which allow us to track the timing of each state relative to the beginning of the simulation.

The allowed transitions are as follows:

i. For the dissociated state with MutS and MutL adjacent:

(a) MutS moves away from MutL with rate $k_S = D_S x_{\text{step}}^2$, where $x_{\text{step}} = 1$ bp is the simulation spatial step size;

(b) MutL moves away from MutS with rate $k_L = D_L x_{\text{step}}^2$;

(c) MutS and MutL form an associated complex with rate consistent with p_A, in particular $k_A = (k_S + k_L)\frac{p_A}{(1 - p_A)}$. with MutS and MutL spatially separated:

(a) MutS moves away from MutL with rate $k_S = D_S x_{step}^2$;

(b) MutL moves away from MutS with rate $k_L = D_L x_{step}^2$;

(c) MutS moves toward MutL with rate $k_S = D_S x_{step}^2$;

(d) MutL moves toward MutS with rate $k_L = D_L x_{step}^2$.

ii. For the associated state:

(a) Move left or right with rate $k_{SL} = D_{SL} x_{step}^2$ each;

(b) Dissociate with rate $k_D = 1/\tau_{A,\mu}$. After dissociation, the bases are placed 1 bp apart. This is achieved by moving one protein by 1 bp away from the last complex position and leaving the other protein at the last complex position. MutS is moved with probability $k_S/(k_S + k_L)$, and MutL is moved with probability $k_L/(k_S + k_L)$.

To calculate observables with this simulation, we start with the proteins in an associated state at position 0 and track their positions along the strand as a function of time. Assuming that the associated complex searches every position that it passes, the fraction of simulations in which the complex has passed a specific position in the given amount of time is the overall successful-search probability at that position, as in Fig. 6. Additionally, we can use the distance that separates dissociated MutS and MutL clamps at a given time to calculate the macroscopic association time. In particular, the time at which the distance between the clamps first reaches x_M is recorded for each simulation and then the distribution of these times is used to calculate a decay constant, as in Fig. 7 below.

B. Agreement with base pair stepping simulation

To validate the ADESS approach and the microscopic parameter calculation, we compare ADESS to the more time consuming BPSS. Since the BPSS approach follows every single diffusion step of the clamps, it becomes computationally unfeasible to obtain sufficient statistics for realistic values of the diffusion constants and we thus perform this validation for $D_S = 10^4$ bp^2/s, $D_{SL} = 10^3$ bp^2/s, and $D_L = 10^5$ bp^2/s, which are each about two orders of magnitude smaller than the actual experimentally determined diffusion constants. Figure 6 compares the search probability calculated using the BPSS approach and the search probability calculated using the ADESS approach and finds them to yield identical results within statistical error for the largest possible value $p_A = 1$ in Fig. 6(a) and the absolutely smallest possible value of $p_A = 10^{-4}$ considered in Appendix D in Fig. 6(b).
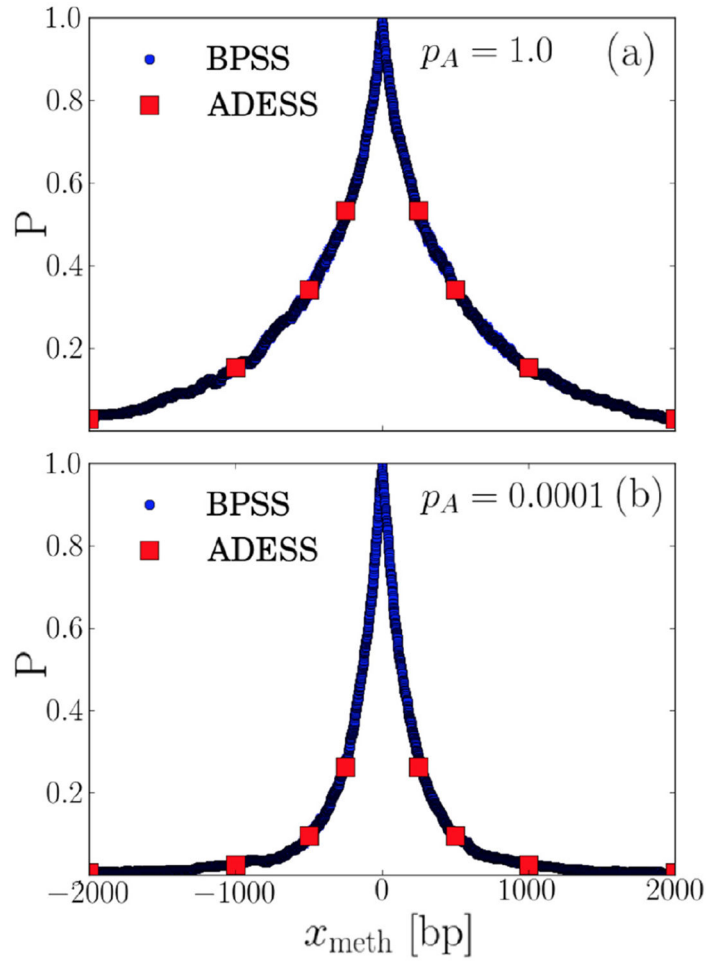
FIG. 6.

Comparison between successful-search probabilities $P^{(*)}_{185,1}$ for $t_s = 185$ s and $n_s = 1$ calculated using the BPSS (small blue circles) and ADESS (large red squares) validates the ADESS approach. Panel (a) is calculated with $p_A = 1$ and panel (b) with $p_A = 10^{-4}$. The statistical uncertainty is smaller than the size of the symbols.

Additionally, the BPSS allows us to validate (Eq. 14) for the microscopic association lifetime $\tau_{A,\mu}$ empirically. In particular, the BPSS approach lets us keep track of the distance between separate clamps and the times at which these distances occur. Using this feature, we calculate the time $t_{A,M}$ for which the clamps remain within the macroscopic association distance x_M of each other, i.e., the time until they first reach the macroscopically dissociated state. Figure 7 shows histograms of this time to reach the macroscopically dissociated state calculated from simulations that use the microscopic association lifetime calculated via (Eq. 14). We find that these simulated distributions accurately reproduce the experimentally measured macroscopic association lifetime $\tau_{A,M} \approx 30 \pm 3$ s (see Table I), indicating that (Eq. 14) correctly matches the microscopic association lifetime governing the multiple transitions between the microscopically associated and the proximal state to the macroscopic association lifetime observed in experiments.
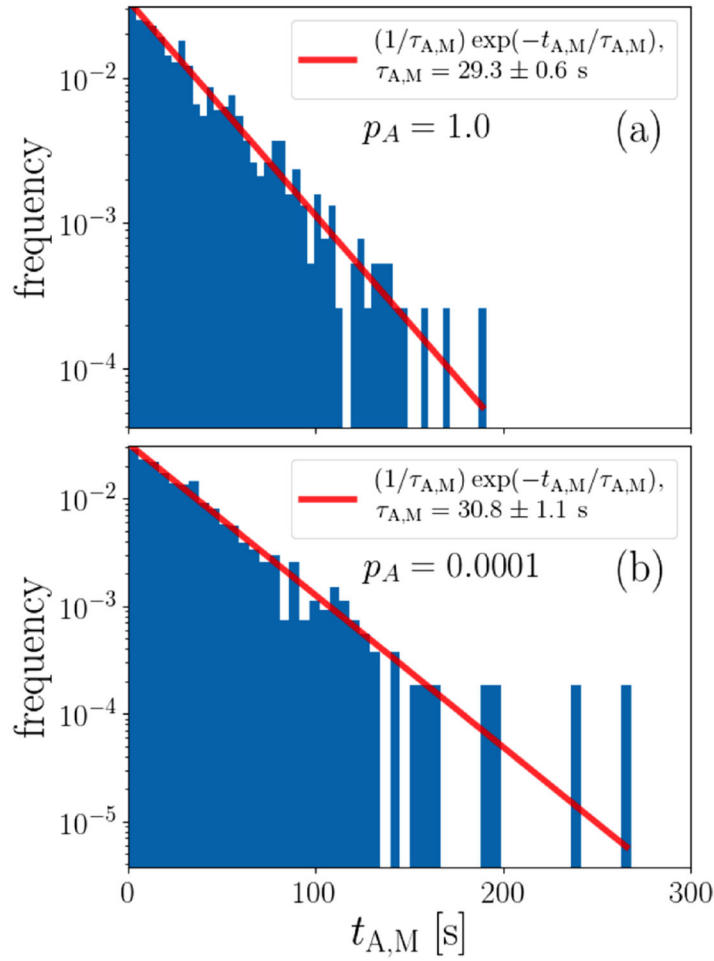
FIG. 7.

Histogram of simulated macroscopic association times for (a) $p_A = 1$ and (b) $p_A = 10^{-4}$ validates Eq. (14). The line is given by $\tau_{A,M}^{-1} \exp(-t_{A,M} / \tau_{A,M})$, where $\tau_{A,M}$ is the average of the simulated macroscopic association times. This line therefore demonstrates that the association time probability decays exponentially with a decay constant consistent with the macroscopic association lifetime of $\tau_{A,M} \approx 30 \pm 3$ s measured experimentally in Ref. [3]. The reported standard error of the decay constant is calculated by dividing the simulated data into ten independent blocks and calculating the mean of each of them independently.

APPENDIX C: MICROSCOPIC PARAMETER CALCULATION

The following are the full calculations used to determine the microscopic protein dynamics from experimental observables. In particular, we calculate the microscopic diffusion constant $D_{\mathrm{SL},\mu}$ and the microscopic association lifetime $\tau_{A,\mu}$. The calculations of P_M and $\tau(x)$ closely follow Ref. [32], a web published early draft of Ref. [33].

A. MutS-MutL association lifetime

First, we calculate the microscopic association lifetime. Consider first the macroscopic association lifetime, which can be written as

$$\tau_{A,M} = \tau_{A,\mu}[(\langle N_A \rangle - 1)p_A + 1] + \tau_R(\langle N_A \rangle - 1) + \tau_M, \tag{C1}$$

where N_A is the number of times the clamps are microscopically adjacent during a single macroscopic association, p_A is the probability of microscopic association given that the clamps are adjacent, τ_R is the average time to return to the adjacent state, and τ_M is the average time to reach a distance x_M without returning to the adjacent state (i.e., the average time to macroscopic dissociation). Note that removing a single adjacent state from the factor multiplied by p_A and multiplying it directly by $\tau_{A,\mu}$ ensures that there is at least one microscopic association in every macroscopic association. This must be true physically, since different diffusion rates are observed during macroscopic association.

Consider N_A for a complex starting in the aggregate state:

$$\begin{aligned} P(N_A = 1) &= P_M, \\ P(N_A = 2) &= (1 - P_M)P_M, \\ P(N_A = 3) &= (1 - P_M)^2 P_M, \\ P(N_A) &= (1 - P_M)^{N_A - 1} P_M, \end{aligned} \tag{C2}$$

where P_M is the probability for a newly microscopically dissociated complex to go to x_M. Thus,

$$\langle N_A \rangle = P_M \sum_{N_A = 1}^{\infty} N_A (1 - P_M)^{N_A - 1} = \frac{1}{P_M}. \tag{C3}$$

In order determine P_M we first consider P_M as a function of the distance between the clamps, which we will denote as x for the remainder of this section to avoid the more cumbersome notation of x_{rel} used in the rest of the paper. Evaluation of this function at $x = x_d$ will give P_M. [$P_M(x)$ will refer to the probability to go to x_M from some position x without visiting 0, while $P_M \equiv P_M(x_d)$ refers to the probability to go to x_M from x_d.] Additionally, since the clamps diffuse with intermittent DNA contact, $P_M(x)$ will be calculated under the assumption that the distance between clamps diffuses continuously. This allows us to write

$$\begin{aligned} P_M(x) &= \frac{1}{2}P_M(x + \delta x) + \frac{1}{2}P_M(x - \delta x) \\ 0 &= \frac{P_M(x + \delta x) - 2P_M(x) + P_M(x - \delta x)}{\delta x^2}, \end{aligned} \tag{C4}$$

and therefore

$$\frac{\partial^2 P_M(x)}{\partial x^2} = 0, \tag{C5}$$

with the boundary conditions

$$
\begin{aligned}
P_M(0) &= 0, \\
P_M(x_M) &= 1 \,.
\end{aligned}
\tag{C6}
$$

The unique solution of this differential equation is

$$
P_M(x) = \frac{x}{x_M},
\tag{C7}
$$

and thus

$$
P_M \equiv P_M(x_d) = \frac{x_d}{x_M},
\tag{C8}
$$

where x_d is the separation of the clamps immediately following dissociation. Therefore, we conclude that

$$
\langle N_A \rangle = x_M \,/\, x_d \,.
\tag{C9}
$$

To compute the microscopic association lifetime $\tau_{A,\mu}$ from Eq. (C1), it is also necessary to compute the average return time τ_R and the average time τ_M to reach x_M. To this end, consider the average time $\tau(x)$ for the distance between the clamps to reach either 0 or x_M given that the starting distance is x:

$$
\tau(x) = \sum_{\text{paths } p} t_p(x) P_p(x),
\tag{C10}
$$

where $t_p(x)$ is the time for a path of length x and $P_p(x)$ is the probability of such a path. Consideration of the effect of a single infinitesimal time step δt allows us to write

$$
\begin{aligned}
\tau(x) &= \sum_{\text{paths } p} t_p(x) P_p(x) \\
&= \sum_{\text{paths } p} \left[\frac{1}{2} t_p(x + \delta x) P_p(x + \delta x) \right. \\
&\quad \left. + \ \frac{1}{2} t_p(x - \delta x) P_p(x - \delta x) \right] + \delta t \\
&= \frac{1}{2} \tau(x + \delta x) + \frac{1}{2} \tau(x - \delta x) + \delta t \,.
\end{aligned}
\tag{C11}
$$

Thus, division by the square of some small spatial step δx^2 yields

$$
-\frac{2\delta t}{\delta x^2} = \frac{\tau(x + \delta x) + \tau(x - \delta x) - 2\tau(x)}{\delta x^2} \,.
\tag{C12}
$$

Therefore,

$$\frac{\partial^2 \tau(x)}{\partial x^2} = -\frac{2\delta t}{\delta x^2} = -\frac{2}{D_{\text{rel}}}, \tag{C13}$$

where we write the right-hand side in terms of the diffusion constant $D_{\text{rel}} = D_S + D_L$. The boundary conditions

$$\begin{aligned} \tau(0) &= 0, \\ \tau(x_M) &= 0 \end{aligned} \tag{C14}$$

allow us to conclude

$$\tau(x) = \frac{x}{D_{\text{rel}}}(x_M - x). \tag{C15}$$

We now write this quantity in terms of τ_R and τ_M as follows:

$$\langle N_A \rangle \tau(x_d) = \tau_R(\langle N_A \rangle - 1) + \tau_M. \tag{C16}$$

Thus, substitution into Eq. (C1) yields

$$\tau_{A,M} = \tau_{A,\mu}[(\langle N_A \rangle - 1)p_A + 1] + \langle N_A \rangle \tau(x_d). \tag{C17}$$

Finally, we conclude

$$\tau_{A,\mu} = \frac{\tau_{A,M} - \langle N_A \rangle \tau(x_d)}{[(\langle N_A \rangle - 1)p_A + 1]}, \tag{C18}$$

where $\langle N_A \rangle = x_M/x_d$.

B. Microscopic diffusion constant

Having computed the microscopic association lifetime, we turn our attention to the microscopic diffusion constant. During microscopic association, the observable quantity, that is, the diffusion of the center of mass of the oscillating dissociative complex, is given by

$$D_{M,SL} = P_A D_{SL,\mu} + P_D D_{CM}, \tag{C19}$$

where $D_{SL,\mu}$ and D_{CM} are the microscopically associated and dissociated complex diffusion rates, respectively, and $D_{M,SL}$ is the measured, macroscopic diffusion rate of the complex. P_A and P_D are the probabilities that the clamps are associated and dissociated, respectively. As argued in Sec. A 1, $D_{CM} = \frac{D_S D_L}{D_S + D_L}$. It follows that the quantity needed for the microscopic model, the microscopic diffusion constant, is given by

$$D_{SL,\mu} = \frac{1}{P_A}\left(D_{M,SL} - P_D \frac{D_S D_L}{D_S + D_L}\right). \tag{C20}$$

Since $D_{M,SL}$, D_S, and D_L are measured experimentally, we only need to write P_A and P_D in terms of observable quantities to obtain a value for $D_{SL,\mu}$. To do this, we observe that the probabilities that the proteins are microscopically associated and dissociated are given by the ratios of average time spent in an associated and dissociated state, respectively, divided by the sum of these times:

$$P_A = \frac{p_A \tau_{A,\mu}}{p_A \tau_{A,\mu} + \tau_R}, \tag{C21}$$

$$P_D = \frac{\tau_R}{p_A \tau_{A,\mu} + \tau_R}, \tag{C22}$$

where $\tau_{A,\mu}$ is the microscopic association time, and τ_R is the average time to return to the adjacent state. $\tau_{A,\mu}$ is multiplied by the association probability p_A because there are $1/p_A$ returns with time τ_R for every microscopic association. Note that τ_M does not enter these equations. This is because the final walk from $x_{rel} = 0$ to x_M has only a minor influence on the experimentally measured diffusion rate as τ_M represents only the last $\approx x_M^2 / D_{rel} \approx 0.1$ s of the ≈ 30 s macroscopic association.

Equation (C16) gives an expression for τ_R in terms of τ_M, so in order to determine τ_R we must first compute τ_M. Fortunately, we can calculate τ_M in a way that is analogous to the calculation of $\tau(x)$ in the previous section. Going back to a discrete picture, during a random walk that results in a separation distance $x = x_M$ before reaching $x = 0$, the first step after dissociation is from $x = x_d$ to $x = 2x_d$. Thus,

$$\tau_M = \tau_{step} + \langle N_{x_d} \rangle \tau_{x_d, M}(2x_d), \tag{C23}$$

where $\tau_{x_d, M}(x)$ is the average time for the distance between the clamps to reach either x_d or x_M, and N_{x_d} is the number of times the distance reaches x_d before going to x_M. Modifying the calculation of $\tau(x)$ with the appropriate boundary conditions

$$\begin{aligned} \tau_{x_d, M}(x_d) &= 0, \\ \tau_{x_d, M}(x_M) &= 0, \end{aligned} \tag{C24}$$

we find

$$\tau_{x_d, M}(x) = \frac{x - x_d}{D_{rel}}(x_M - x), \tag{C25}$$

which yields

$$\tau_M = \frac{x_d^2}{D_{rel}} + \frac{\langle N_{x_d} \rangle x_d}{D_{rel}}(x_M - 2x_d). \tag{C26}$$

Similarly, $\langle N_{x_d} \rangle$ can be computed in the same way that $\langle N_A \rangle$ was found earlier. In particular,

$$\langle N_{x_d} \rangle = \frac{1}{P_{x_d, M}}, \tag{C27}$$

where $P_{x_d,M}$ is the probability that the distance goes to x_M before x_d from distance $2x_d$.

Using Eqs. (C4) and (C5) with boundary conditions

$$\begin{aligned} P_M(x_d) &= 0, \\ P_M(x_M) &= 1, \end{aligned} \tag{C28}$$

we get

$$P_{x_d, M} = \frac{x - x_d}{x_M - x_d}. \tag{C29}$$

Finally, since we assume that the walk starts at $x = 2x_d$,

$$\langle N_{x_d} \rangle = \frac{x_M - x_d}{x_d}. \tag{C30}$$

Appropriate substitutions and algebraic manipulations yield

$$D_{\text{SL}, \mu} = D_{\text{SL,M}} - \delta(D_{\text{CM}} - D_{\text{SL,M}}), \tag{C31}$$

with

$$\delta = R_x R_\tau \frac{(2 - \frac{R_x}{1 - R_x})(1 + \frac{R_x}{p_A(1 - R_x)})}{\frac{1}{(1 - R_x)} - R_\tau} \tag{C32}$$

$$\approx 2 R_x R_\tau \left(1 + \frac{R_x}{p_A}\right), \tag{C33}$$

where $R_x \equiv \frac{x_d}{x_M} \approx 10^{-3}$, $R_\tau \equiv \frac{x_M^2}{\tau_{A, M} D_{\text{rel}}} \approx 10^{-2}$ for the specific values of the parameters and the approximation in the second line holds since $R_x \ll 1$ and $R_\tau \ll 1$. For $p_A > 0.001$ the correction $\delta(D_{\text{CM}} - D_{\text{SL,M}})$ is ≈ 3 bp^2/s $\approx 0.01\%$ of $D_{\text{SL,M}}$ and even for the most extreme worst case value of $p_A = 10^{-4}$ considered in Appendix D we still find $\delta(D_{\text{CM}} - D_{\text{SL,M}})$ is ≈ 50 bp^2/s $\approx 0.1\%$ of $D_{\text{SL,M}}$.

Thus,

$$D_{\text{SL}, \mu} \approx D_{\text{SL,M}}. \tag{C34}$$

APPENDIX D: APPROXIMATION OF LOWER LIMIT OF ASSOCIATION PROBABILITY

The lower limit of the association probability can be calculated under the assumption that $p_A \geq P_{\text{assoc, soln}}$, where $P_{\text{assoc, soln}}$ is the probability that a MutL in solution colliding with a DNA-bound MutS will associate. As discussed in the main text, it should be easier for MutL and MutS to bind when they are both already somewhat aligned by their formation of clamp structures on the DNA.

The association probability $P_{\text{assoc, soln}}$ is given by the ratio

$$P_{\text{assoc, soln}} = k_{\text{on, expt}} / k_{\text{on, max}}, \tag{D1}$$

where $k_{\text{on, expt}}$ is the experimental rate at which MutL associates with MutS on DNA from solution, and $k_{\text{on, max}}$ is the rate at which MutS and MutL collide (e.g., the diffusion-limited rate).

We first focus on the diffusion-limited rate. The Smoluchowski equation yields an expression for the diffusion-limited rate constant for two uniform spheres [34]:

$$k_{\text{on, max}} = 4\pi D R, \tag{D2}$$

where D is the relative diffusion constant and R is the reaction radius.

Manelyte *et al.* give the MutS Stokes radius as $R_{\text{S,S}} \approx 3$ nm [35], and Grilley *et al.* give the MutL Stoke radius as $R_{\text{S,L}} \approx 6$ nm [36]. Therefore $R \approx R_{\text{S,S}} + R_{\text{S,L}} \approx 10$ nm.

To determine the relative diffusion constant D, we use the measured MutS diffusion along the DNA strand, $D_S = 0.043 \pm 0.016$ $\mu m^2/s$, and the Stokes-Einstein diffusion of MutL in water at room temperature, $D_{\text{L, soln}} = \frac{k_B T}{6\pi\eta R_{\text{S,L}}} \approx 4 \times 10^{-11} \, m^2/s \gg D_S$. Thus $D \approx 4 \times 10^{-11}$ m^2/s and the diffusion-limited on rate is

$$k_{\text{on, max}} \approx 10^9 \, M^{-1} s^{-1}. \tag{D3}$$

We can now turn to the experimental on rate. Liu *et al.* do not measure this rate directly, but they do find the fraction F_{SL} of an ensemble of DNAs on which MutS-MutL complexes associate in equilibrium to be high enough to perform the experiment, i.e., a significant fraction of their constructs shows association of a MutL at their experimental concentration of MutL [3]. We thus choose $F_{\text{SL}} = 0.1$ as a conservative "worst case" estimate with $F_{\text{SL}} \approx 1$ more likely. This, along with the known MutS dissociation constant with DNA, $K_{d,S} = 0.6$ μM [26] and the measured MutL off rate $k_{\text{off,L}} \sim 1/\tau_{\text{on,L}} \approx 1/850$ s can be used to estimate the desired on rate. The fraction of DNAs with MutS-MutL associated is given by

$$F_{\text{SL}} = [\text{SLDNA}] / [\text{DNA}] = \frac{k_{\text{on,L}}[\text{L}][\text{SDNA}]}{k_{\text{off,L}}[\text{DNA}]}, \tag{D4}$$

and thus

$$k_{\text{on},L} = \frac{k_{\text{off,L}} F_{SL} K_{d,S}}{[L][S]}. \tag{D5}$$

For the reported $[L] \approx 20$ nM and $[S] \approx 10$ nM,

$$k_{\text{on,L}} \approx 10^5 \, \text{M}^{-1} \, \text{s}^{-1} \tag{D6}$$

for the worst case estimate $F_{SL} = 0.1$ and $k_{\text{on,L}} \approx 10^6 \, \text{M}^{-1} \, \text{s}^{-1}$ for $F_{SL} = 1$. Thus we conclude that

$$P_{\text{assoc, soln}} \approx 0.001 \, \text{M}^{-1} \, \text{s}^{-1}, \tag{D7}$$

and therefore

$$0.001 \leqslant p_A \leqslant 1, \tag{D8}$$

which gets widened to $10^{-4} \leqslant p_A \leqslant 1$ in the worst case of $F_{SL} = 0.1$.

APPENDIX E: ROBUSTNESS OF RESULTS AGAINST VARIATION IN ESTIMATED PARAMETERS

Since several model parameters can only be estimated (see Table IV), we next determine how sensitive our model is to variations in these parameters. The parameter with the largest uncertainty is the microscopic association probability p_A. To gauge the sensitivity of the model to this parameter, we hold all other parameters constant at their values given in Tables III and IV (both in the presence of and the absence of MutH) while varying the microscopic association probability over its entire potential range given in Appendix D, all the way down to the "worst case" lower limit of 10^{-4}. Then, we numerically calculate the main observable of our model, namely the probability of a successful search, using the ADESS approach described in Sec. IV B 1.

Figure 8 shows the resulting search probabilities as a function of search distance x_{meth} for different values of the association probability p_A. We note that the successful-search probability is largely independent of the microscopic association probability p_A as long as $p_A \geqslant 0.001$ and then drops significantly for $p_A = 10^{-4}$. Since a significantly reduced search probability would be evolutionarily disadvantageous, this provides further evidence that the range $0.001 \leqslant p_A \leqslant 1$ is more realistic compared with the extreme case of $10^{-4} \leqslant p_A \leqslant 1$. In the realistic range, the search probability is largely insensitive to the value of p_A.
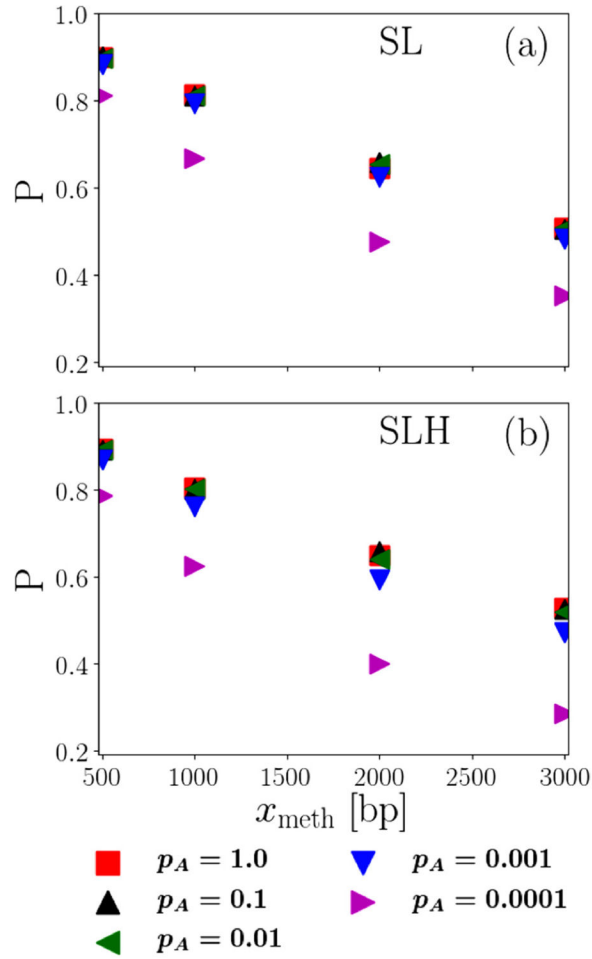
FIG. 8.

Search probability $P_{185,1}^{(*)}$ for $t_s = 185$ s and $n_s = 1$ searches as a function of search distance for different values of the association probability p_A indicates that the search probability is not sensitive to variation of p_A from 0.001 to 1 but that the search probability drops precipitously below 0.001. Panel (a) was calculated with non-MutH parameters, while panel (b) was calculated with MutH parameters. The statistical uncertainty is smaller than the size of the symbols.

We note that, naïvely, it appears unintuitive for the overall search probability to be so insensitive to three orders of magnitude of variation in the probability that two adjacent clamps successfully form a complex. However, we would like to point out that the microscopic association probability p_A appears in Eq. (14) for the microscopic association lifetime. Thus, different values for the microscopic association probability p_A yield different values for the microscopic association lifetime $\tau_{A,\mu}$ to keep the macroscopic association lifetime $\tau_{A,\mu}$ consistent with its measured value. The relative insensitivity of the search probability to the value of the microscopic association probability thus indicates that changes to the microscopic association lifetime compensate for the significant variation in microscopic association probabilities over three orders of magnitude. This also explains

the change in behavior at $p_A = 0.001$. Since the number of returns of the two clamps before final dissociation is $\langle N_A \rangle = 1000$ for our parameters, the denominator $(\langle N_A \rangle - 1)p_A + 1$ in Eq. (14) is larger than one for $p_A \geqslant 0.001$ and asymptotes to one for $p_A < 0.001$. Thus, for $p_A \geqslant 0.001$ the clamps go through multiple reassociation events before final dissociation, the lifetime of which compensates for the change in the microscopic association probability p_A. For $p_A < 0.001$, the probability for even a single reassociation is becoming small and the microscopic association lifetime $\tau_{A,\mu}$ is locked to the macroscopic association lifetime $\tau_{A,M}$ and is no longer able to compensate for changes in the association probability p_A.
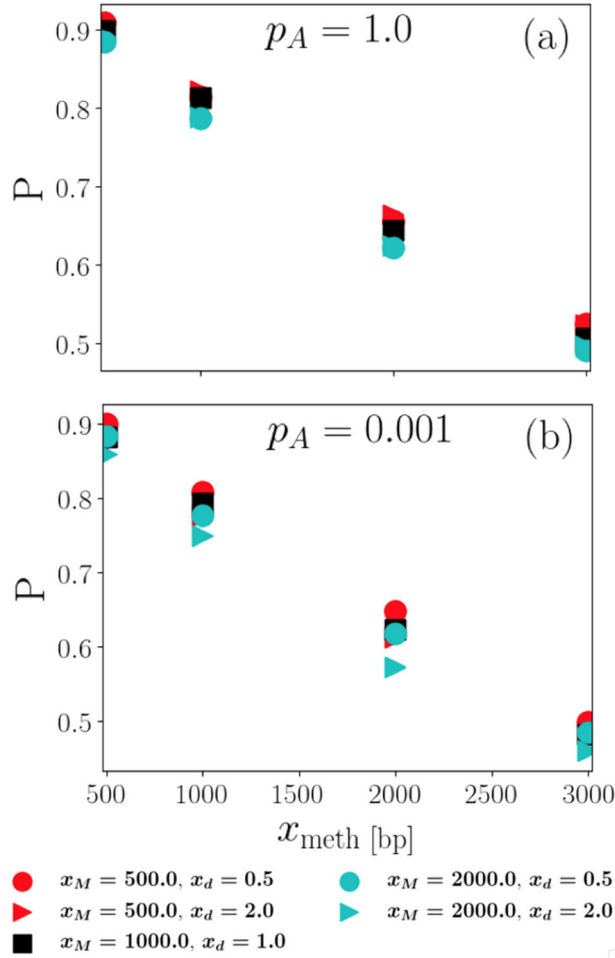
FIG. 9.

Comparison of successful-search probability for $n_s = 1$ and $t_s = 185$ s (i.e., $P^{(*)}_{185,1}$) for factor of two variations in the macroscopic and microscopic dissociation distances. This comparison reveals that these probabilities are not sensitive to the particular value chosen in these estimations. Panel (a) is calculated with $p_A = 1$ and panel (b) with $p_A = 0.001$. All results are shown for experimental parameters in the absence of MutH. The statistical uncertainty is smaller than the size of the symbols.

Similar to our analysis of the sensitivity of the association probability p_A, we vary the values of the dissociation distances x_d and x_M by a factor of two in each direction to determine

the sensitivity of the search probability to changes in these parameters at both limits of p_A. Figure 9 demonstrates that, for $p_A = 1$ and $p_A = 10^{-3}$, variation of the dissociation distances x_d and x_M by a factor of two only introduces a relative difference of up to 13%. We thus conclude that the difference between the approximate and exact values of the dissociation distances x_d and x_M will not significantly affect our results.

References

[1]. Martin-Lopez JV and Fishel R, The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. Fam. Cancer 12, 159 (2013). [PubMed: 23572416]

[2]. Reyes GX, Schmidt TT, Kolodner RD, and Hombauer H, New insights into the mechanism of DNA mismatch repair, Chromosoma 124, 443 (2015). [PubMed: 25862369]

[3]. Liu J, Hanne J, Britton BM, Bennett J, Kim D, Lee Jong-Bong, and Fishel R, Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair, Nature (London) 539, 583 (2016). [PubMed: 27851738]

[4]. Wang Jean Y. J. and Edelmann W, Mismatch repair proteins as sensors of alkylation DNA damage, Cancer Cell 9, 417 (2006). [PubMed: 16766259]

[5]. Iyer RR, Pluciennik A, Burdett V, and Modrich PL, DNA mismatch repair: Functions and mechanisms, Chem. Rev. (Washington, DC, U. S.) 106, 302 (2006).

[6]. Fishel R, Mismatch repair, J. Biol. Chem 290, 26395 (2015). [PubMed: 26354434]

[7]. O'Donnell M, Kuriyan J, Kong XP, Stukenberg PT, and Onrust R, The sliding clamp of DNA polymerase III holoenzyme encircles DNA. Mol. Biol. Cell 3, 953 (1992). [PubMed: 1358275]

[8]. Daitchman D, Greenblatt HM, and Levy Y, Diffusion of ring-shaped proteins along DNA: case study of sliding clamps, Nucleic Acids Res. 46, 5935 (2018). [PubMed: 29860305]

[9]. Berg OG, Winter RB, and Von Hippel PH, Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory, Biochemistry 20, 6929 (1981). [PubMed: 7317363]

[10]. Lomholt MA, van den Broek B, Svenja-Marei Kalisch J, Wuite GJ, and Metzler R, Facilitated diffusion with DNA coiling, Proc. Natl. Acad. Sci. USA 106, 8204 (2009). [PubMed: 19420219]

[11]. Bénichou O, Chevalier C, Meyer B, and Raphaël Voituriez, Facilitated Diffusion of Proteins on Chromatin, Phys. Rev. Lett 106, 038102 (2011). [PubMed: 21405302]

[12]. Givaty O and Levy Y, Protein sliding along DNA: Dynamics and structural characterization, J. Mol. Biol 385, 1087 (2009). [PubMed: 19059266]

[13]. Reingruber Jürgen and Holcman D, Transcription factor search for a DNA promoter in a three-state model, Phys. Rev. E 84, 020901(R) (2011).

[14]. Cartailler J and Reingruber J, Facilitated diffusion framework for transcription factor search with conformational changes, Phys. Biol 12, 046012 (2015). [PubMed: 26200216]

[15]. Mirny L, Slutsky M, Wunderlich Z, Tafvizi A, Leith J, and Kosmrlj A, How a protein searches for its site on DNA: The mechanism of facilitated diffusion, J. Phys. A: Math. Theor 42, 434013 (2009).

[16]. Slutsky M and Mirny LA, Kinetics of protein-DNA interaction: Facilitated target location in sequence-dependent potential, Biophys. J 87, 4021 (2004). [PubMed: 15465864]

[17]. Bauer M and Metzler R, Generalized facilitated diffusion model for dna-binding proteins with search and recognition states, Biophys. J 102, 2321 (2012). [PubMed: 22677385]

[18]. Zhou Huan-Xiang, Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins, Proc. Natl. Acad. Sci. U. S. A 108, 8651 (2011). [PubMed: 21543711]

[19]. Kong X-P, Onrust R, O'Donnell M, and Kuriyan J, Three-dimensional structure of the β subunit of E. coli DNA polymerase III holoenzyme: A sliding DNA clamp, Cell (Cambridge, MA, U. S.) 69, 425 (1992).

[20]. Redner S, A Guide to First-Passage Processes (Cambridge University Press, Cambridge, 2001).

[21]. Veksler A and Kolomeisky AB, Speed-selectivity paradox in the protein search for targets on DNA: Is it real or not? J. Phys. Chem. B 117, 12695 (2013). [PubMed: 23316873]

[22]. Gillespie DT, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem 81, 2340 (1977).

[23]. Lacks S and Greenberg B, Complementary specificity of restriction endonucleases of Diplococcus pneumoniae with respect to DNA methylation, J. Mol. Biol 114, 153 (1977). [PubMed: 20509]

[24]. Hattman S, Brooks JE, and Masurekar M, Sequence specificity of the P1 modification methylase (MEco P1) and the DNA methylase (mEco dam) controlled by the Escherichia coli dam gene, J. Mol. Biol 126, 367 (1978). [PubMed: 370402]

[25]. Geier GE and Modrich P, Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. J. Biol. Chem 254, 1408 (1979). [PubMed: 368070]

[26]. Acharya S, Foster PL, Brooks P, and Fishel R, The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair, Mol. Cell 12, 233 (2003). [PubMed: 12887908]

[27]. Graham WJ, Putnam CD, Kolodner RD, et al. , The properties of Msh2–Msh6 ATP binding mutants suggest a signal amplification mechanism in DNA mismatch repair, J. Biol. Chem 293, 18055 (2018). [PubMed: 30237169]

[28]. Hombauer H, Campbell CS, Smith CE, Desai A, and Kolodner RD, Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates, Cell (Cambridge, MA, U. S.) 147, 1040 (2011).

[29]. Konopka MC, Shkel IA, Cayley S, Record MT, and Weisshaar JC, Crowding and confinement effects on protein diffusion in vivo, J. Bacteriol 188, 6115 (2006). [PubMed: 16923878]

[30]. Liu J, Lee R, Britton BM, London JA, Yang K, Hanne J, Lee J-B, and Fishel R, MutL sliding clamps coordinate exonuclease-independent Escherichia coli mismatch repair, Nat. Commun 10, 5294 (2019). [PubMed: 31757945]

[31]. Zessin PJM, Sporbert A, and Heilemann M, PCNA appears in two populations of slow and fast diffusion with a constant ratio throughout S-phase in replicating mammalian cells, Sci. Rep 6, 18779 (2016). [PubMed: 26758689]

[32]. Ben-Naim E, Krapivsky PL, and Redner S, "Random walk/diffusion," http://physics.bu.edu/~redner/542/book/rw.pdf (2008), accessed: 2020-06-22.

[33]. Krapivsky P, Redner S, and Ben-Naim E, A Kinetic View of Statistical Physics, edited by Cambridge University Press (Cambridge University Press, Cambridge, 2010).

[34]. Smoluchowski MV, Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. Z. Phys. Chem 92U, 129 (1917).

[35]. Manelyte L, Urbanke C, Giron-Monzon L, and Friedhoff P, Structural and functional analysis of the MutS C-terminal tetramerization domain, Nucleic Acids Res. 34, 5270 (2006). [PubMed: 17012287]

[36]. Grilley M, Welsh KM, Su SS, and Modrich P, Isolation and characterization of the Escherichia coli MutL gene product. J. Biol. Chem 264, 1000 (1989). [PubMed: 2536011]
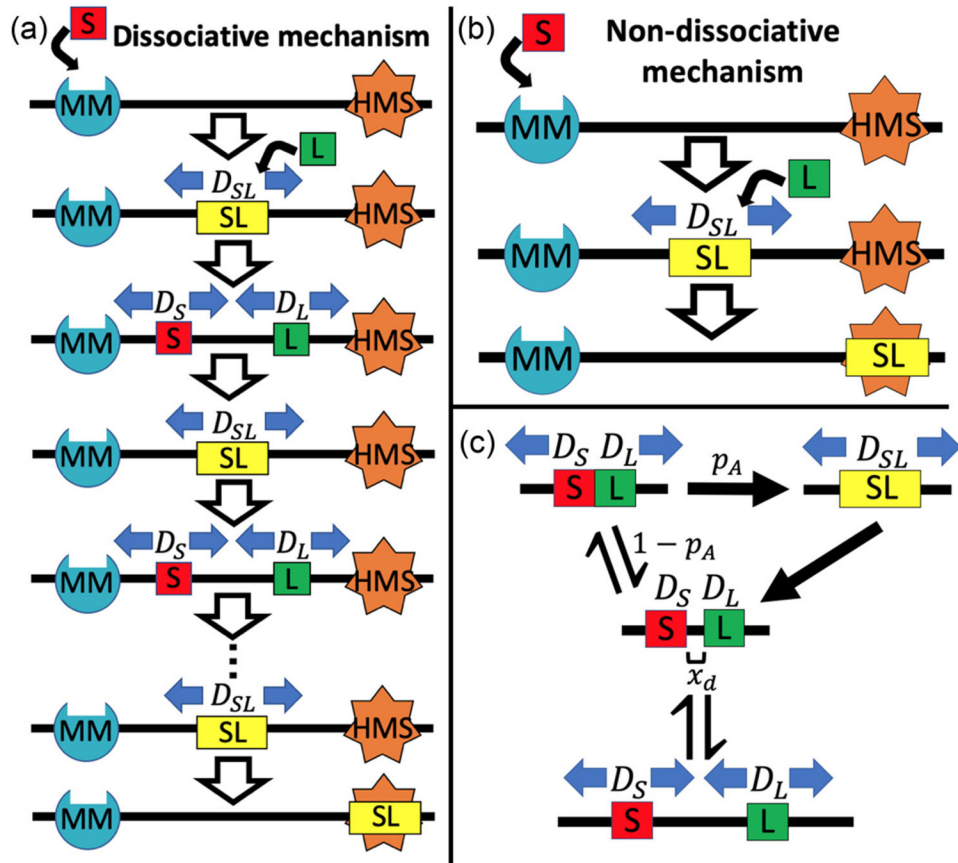
FIG. 1.

(a) Illustration of the dissociative search mechanism. First, MutS forms a sliding clamp at the mismatch site. Then MutS recruits MutL, and they form a MutS-MutL sliding clamp complex capable of recognizing the hemimethylated site. This complex dissociates into independently diffusing MutS and MutL sliding clamps, which in turn re-associate to form the MutS-MutL complex again. This cyclical process continues until either one or both complexes fall off of the DNA or the MutS-MutL complex finds the hemimethylated site. (b) Illustration of the nondissociative search mechanism. First, MutS forms a sliding clamp at the mismatch site. Then MutS recruits MutL, and they form a MutS-MutL sliding clamp complex capable of recognizing the hemimethylated site. In contrast with the dissociative search mechanism, in the nondissociative case MutS and MutL are unable to dissociate from each other. The search therefore continues in the MutS-MutL complex state until one or both of the proteins falls off of the DNA or the hemimethylated site is found. (c) State diagram of the model used to calculate successful hemimethylated site search probabilities in DNA mismatch repair. The DNA is modeled as a one-dimensional track along which the proteins travel. MutS and MutL protein clamps diffuse along the DNA with rates specified by D_S and D_L, respectively, until they come into contact with each other. They may then either diffuse away from each other without associating or associate into a combined, searching MutS-MutL. These happen with probabilities $1 - p_A$ and p_A, respectively. The combined MutS-MutL can in turn diffuse along the DNA with a new rate specified by D_{SL}, and is

capable of searching the DNA over which it passes. After an average lifetime $\tau_{\mathrm{A},\mu}$, this complex dissociates into separate MutS and MutL DNA clamps.
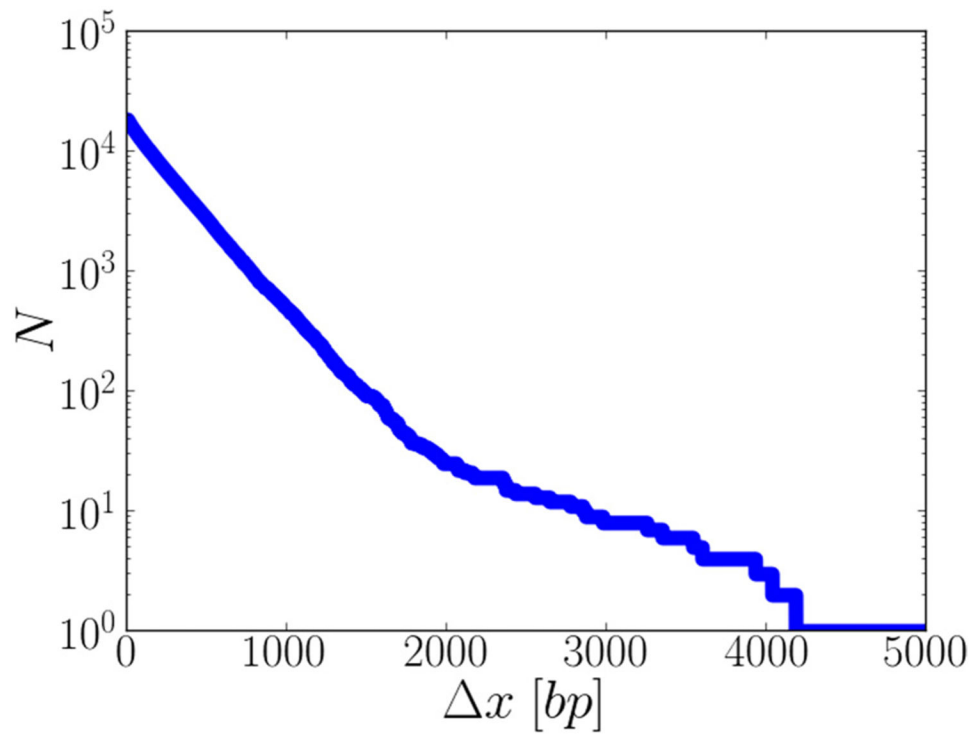
FIG. 2.
Cumulative distribution of hemimethylated site distances in the *E. coli* genome. For
each separation (in base pairs) on the horizontal axis, the vertical axis shows the
number of adjacent GATC sites in the *E. coli* K-12 MG1655, NCBI RefSeq assembly
GCF_000005845.2 genome with at least that separation. The largest separation between any
two GATC sites in the *E. coli* genome is 4960 bp.
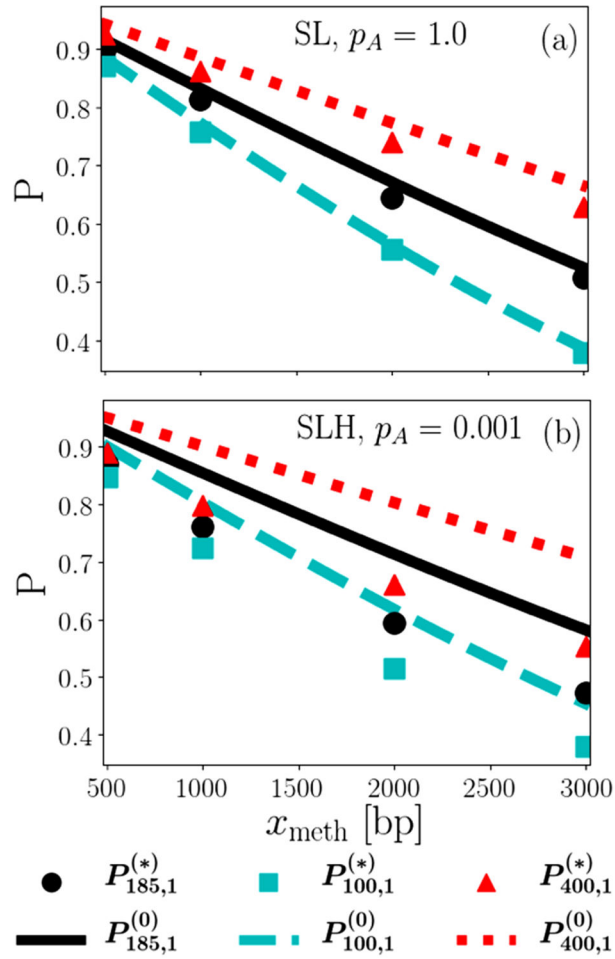
FIG. 3.

Preliminary results indicate neither advantage nor disadvantage conferred by dissociative search, but that both are effective in locating the HMS. In this figure, successful-search probabilities of dissociative ($P_{t_s,1}^{(*)}$) and nondissociative ($P_{t_s,1}^{(0)}$) single searches are plotted as a function of distance x_{meth} from the hemimethylated site for different search times, indicated by the first subscript. Nondissociative search results are plotted as continuous lines (superscript 0), whereas dissociative search results are plotted using discrete symbols (superscript *). Results in panel (a) use experimental parameters in the case where MutH is not associated with MutL (SL) and the upper limit $p_A = 1$ and (b) when MuH is associated with MutL (SLH) and the lower limit $p_A = 0.001$. The statistical uncertainty is smaller than the size of the symbols.
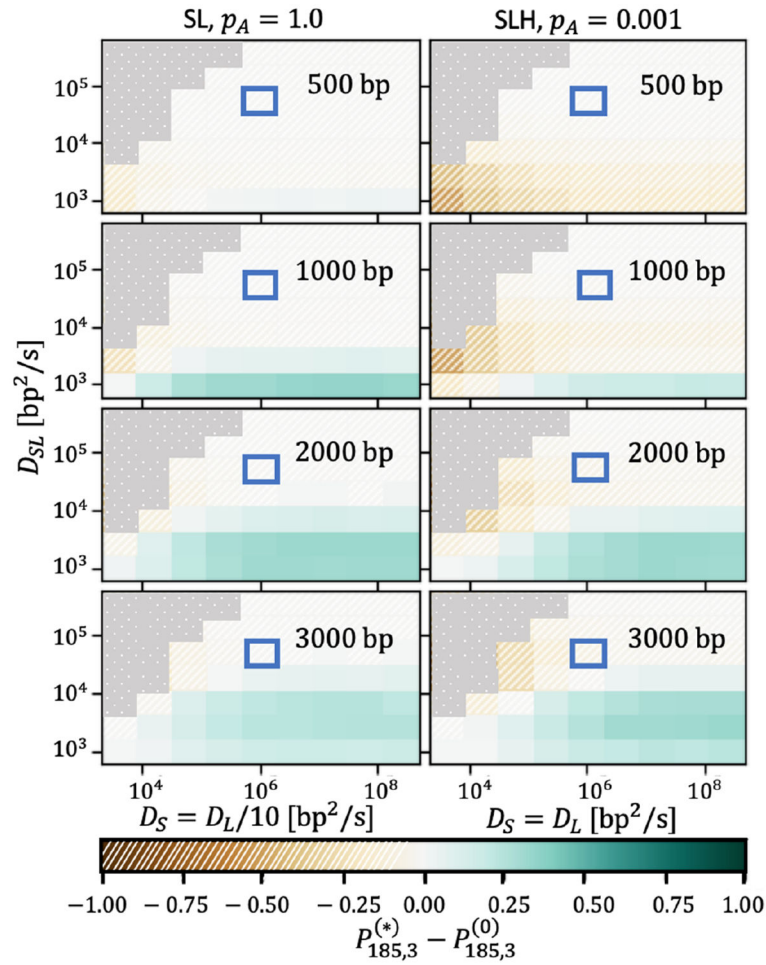
FIG. 4.

The dissociative search mechanism is advantageous over a broad range of diffusion constants (solid green regions). The MutS-MutL diffusion rate D_{SL} is varied on the vertical axis, and individual MutS diffusion rate D_S is varied on the horizontal axis. The color scale indicates the absolute difference between the ADESS dissociative and analytic nondissociative probabilities $\delta P_{185,3}$ for $n_s = 3$ searches with a $t_s = 185$ s cutoff time. Regions where $\delta P_{185,3} < 0$ are hatched and brown, while regions with $\delta P_{185,3} > 0$ are solid and green. In the left column, $D_L = 10 D_S$ (appropriate in the absence of MutH) with $p_A = 1.0$, while in the right column, $D_L = D_S$ (appropriate in the presence of MutH) with $p_A = 0.001$. These columns represent the two extremes in terms of $\delta P_{185,3}$, as the results for MutH parameters at $p_A = 0.001$ and for non-MutH parameters at $p_A = 1$ are in between the two cases shown. The square outlined in blue indicates the order of magnitude of experimental diffusion constants, and the regions of the coefficient space such that the diffusion of the "slow" searching state is faster than that of the "fast" state (that is, $D_{SL} > D_S$) are blocked out in dotted gray, since these regions are nonphysical. The hemimethylated site distances x_{meth} are increased as one moves down the figure vertically, and are indicated in the upper-right corner of each panel.
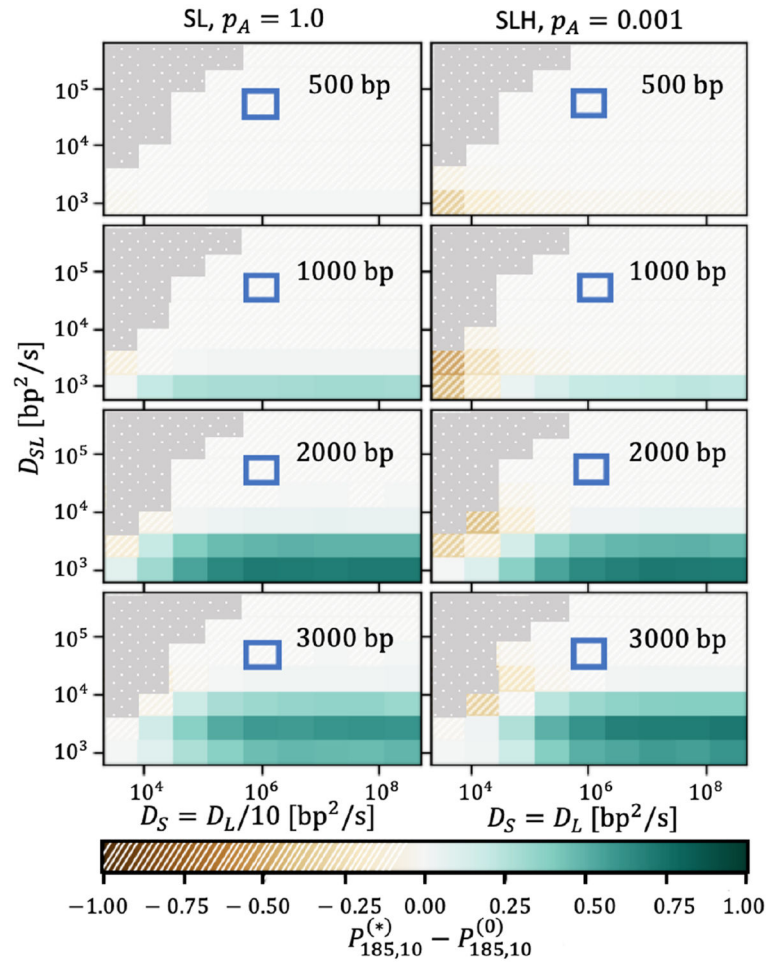
FIG. 5.

The dissociative search mechanism is advantageous over a broad range of diffusion constants (solid green regions). The MutS-MutL diffusion rate D_{SL} is varied on the vertical axis, and individual MutS diffusion rate D_S is varied on the horizontal axis. The color scale indicates the absolute difference between the ADESS dissociative and analytic nondissociative probabilities $\delta P_{185,10}$ for $n_s = 10$ searches with a $t_s = 185$ s cutoff time. Regions where $\delta P_{185,10} < 0$ are hatched and brown, while regions with $\delta P_{185,10} > 0$ are solid and green. In the left column $D_L = 10 D_S$ (appropriate in the absence of MutH) with $p_A = 1.0$, while in the right column $D_L = D_S$ (appropriate in the presence of MutH) with $p_A = 0.001$. These columns represent the two extremes in terms of $\delta P_{185,10}$, because the results for MutH parameters at $p_A = 0.001$ and for non-MutH parameters at $p_A = 1$ are in between the two cases shown. The square outlined in blue indicates order of magnitude of experimental diffusion constants, and the regions of the coefficient space such that the diffusion of the "slow" searching state is faster than that of the "fast" state (that is, $D_{SL} > D_S$) are blocked out in dotted gray, since these regions are nonphysical. The hemimethylated site distances x_{meth} are increased as one moves down the figure vertically and are indicated in the upper-right corner of each panel.

TABLE I.

Summary of relevant quantities measured by Liu *et al.* [3]. The column labeled "SL value" gives the value of each quantity in the absence of MutH, whereas the column labeled "SLH value" gives the value of each quantity in the presence of MutH. Note that some values were not measured separately in the presence of MutH (indicated by NA), so we will assume that MutH does not change these values. In Liu *et al.*'s experiment, 17.3 kb of dsDNA was stretched over 4.4 μm, so we use 4.4 μm = 17.3 kb to convert from reported values in μm by Liu *et al.* to units of bp, which are more convenient for our model.

Quantity	Symbol	SL value	SLH value
Search complex diffusion constant	$D_{SL,M}$	$(6 \pm 3) \times 10^4$ bp²/s	$(8 \pm 5) \times 10^4$ bp²/s
MutS diffusion constant	D_S	$(7 \pm 2) \times 10^5$ bp²/s	NA
MutL diffusion constant	D_L	$(1.4 \pm 0.6) \times 10^7$ bp²/s	$(6 \pm 5) \times 10^5$ bp²/s
MutS-MutL association lifetime	$\tau_{A,M}$	30 ± 3 s	NA
MutS-DNA association lifetime	τ_S	185 ± 35 s	NA
MutL-DNA association lifetime	τ_L	850 ± 150 s	NA

TABLE II.

Summary of stochastic model parameters. The left column describes the physical quantity, and the right column gives the associated symbol.

Parameter	Symbol
MutS-MutL diffusion constant	$D_{SL,\mu}$
MutS-MutL association lifetime	$\tau_{A,\mu}$
MutS diffusion constant	D_S
MutL diffusion constant	D_L
Initial separation after dissociation	x_d
MutS-MutL association probability	p_A
Total search time	t_S

TABLE III.

Model parameter values calculated from experimental observables. The column labeled "SL value" gives each value in the presence of MutH. The column labeled "SLH value" gives each value in the absence of MutH, whereas the

Parameter	Symbol	SL value	SLH value
Dissociated clamps relative position diffusion constant	D_{rel}	$(1.5 \pm 0.6) \times 10^7$ bp^2/s	$(1.3 \pm 0.5) \times 10^6$ bp^2/s
Dissociated clamps center-of-mass diffusion constant	D_{CM}	$(7 \pm 3) \times 10^5$ bp^2/s	$(3.2 \pm 2.8) \times 10^5$ bp^2/s
MutS-MutL diffusion constant	$D_{SL,\mu}$	$(6 \pm 3) \times 10^4$ bp^2/s	$(8 \pm 5) \times 10^4$ bp^2/s
MutS-MutL association lifetime	$\tau_{A,\mu}$	0.03 s $\leqslant \tau_{A,\mu} < 30$ s	0.03 s $\leqslant \tau_{A,\mu} < 30$ s
Distance from hemimethylated site	x_{meth}	500–3000 bp	500–3000 bp
Total search time	t_s	185 ± 35 s	185 ± 35 s

TABLE IV.

Estimated values for model parameters.

Parameter	Symbol	Value
Adjacent MutS-MutL association probability	p_A	$10^{-3} \leqslant p_A \leqslant 1$
MutS-MutL microscopic dissociation distance	x_d	1 bp
MutS-MutL macroscopic dissociation distance	x_M	1000 bp
Number of searches	n_s	3–10