Crowding on DNA in Protein Search for Targets

Alexey A. Shvets and Anatoly B. Kolomeisky*

Department of Chemistry and Center for Theoretical Biological Physics,

Rice University, Houston, TX 77005, USA

(Dated: February 24, 2016)

Abstract

Proteins searching and recognizing specific sites on DNA is required for initiating all major biological processes. While the details of the protein search for targets on DNA in purified *in vitro* systems are reasonably well understood, the situation in real cells is much less clear. The presence of other types of molecules on DNA should prevent reaching the targets, but experiments show that, surprisingly, the molecular crowding on DNA influences the search dynamics much less than expected. We develop a theoretical method that allowed us to clarify the mechanisms of the protein search on DNA in the presence of crowding. It is found that the dimensionality of the search trajectories specifies if the crowding will affect the target finding. For 3D search pathways it is minimal, while the strongest effect is for 1D search pathways when the crowding particle can block the search. In addition, for 1D search we determined that the critical parameter is a mobility of crowding agents: highly-mobile molecules do not affect the search dynamics, while the slow particles can significantly slow down the process. Physical-chemical explanations of the observed phenomena are presented. Our theoretical predictions thus explain the experimental observations, and they are also supported by extensive numerical simulations.

Introduction. Protein-DNA interactions control all major biological processes involved in the transfer and maintenance of genetic information [1]. All these processes are initiated by protein molecules finding and recognizing specific sequences on DNA. Due to a large number of nonspecific sites and interactions with multiple molecules in cellular medium, the protein search is a very complex biochemical and biophysical problem. It has been extensively studied using a variety of experimental and theoretical techniques [2–24]. Although a significant progress in explaining protein search dynamics has been achieved, many aspects of these complex phenomena are still not clarified [4, 5, 15, 17].

Theoretical studies of the protein search phenomena identify three different regimes depending on the nature of the dominating motions [17]. When the protein molecule is strongly nonspecifically bound to DNA most of the time a 1D search regime is realized: the protein slides to the target through the DNA molecule. For the case of weak nonspecific attractions the protein finds the specific sequence by utilizing a 3D search, i.e., it comes to the target directly from the bulk solution. But the most interesting behavior is observed for intermediate range of nonspecific interactions when the search combines both 1D and 3D pathways. Here the fastest search times are typically found [4, 17]. This is known as a *facilitated diffusion* [4, 5]. Recent single-molecule experiments that can visualize the dynamics of individual molecules qualitatively support these views [9, 11, 12, 14, 23]. However, the majority of theoretical models usually consider an oversimplified picture of the unobstructed protein search for always open target sites on DNA [4, 5, 17]. This might be reasonable for some *in vitro* situations, but in live cells the medium is very crowded, and the DNA chains are usually covered by a large number of various biological molecules [1, 15, 26]. This should prevent the fast protein search for target sites on DNA [27, 28]. At the same time, experiments suggest that the crowding does not strongly influence the effectiveness of this process [14, 15]. Thus, the mechanisms of *in vivo* protein search in the presence of crowding particles on DNA remain not well understood [15].

In this Letter, we present a theoretical method that allows us to explicitly investigate the role of the crowding in the protein search for targets on DNA. Using a discrete-state stochastic framework, the protein search dynamics is analyzed in the presence of the crowding agent that can move along the DNA chain. This approach takes into account the most relevant physical-chemical processes in the system. We identify the mobility of the crowding particles as a key property determining the effect of the crowding in the protein search. It

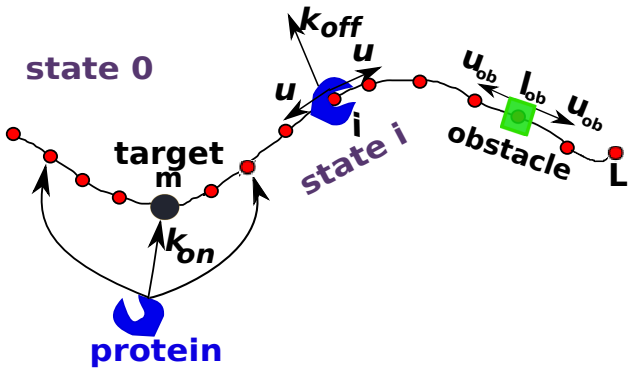

FIG. 1. A general scheme for the protein target search on DNA with a mobile obstacle. There are L nonspecific sites and 1 specific target site, which is located at the site m. A protein molecule can slide along DNA with the diffusion rate u, or it might dissociate into the solution with the rate k_{off}. From the solution the protein can associate to any site on DNA with the rate k_{on} per chain. In addition, the obstacle can diffuse along DNA with the rate u_{ob}.

is also argued that the crowding effects are stronger if the search is taking place via 1D diffusion along DNA in comparison with 3D search via the bulk solutions. Our theoretical predictions are tested by extensive computer Monte-Carlo simulations.

Model. We consider a single DNA molecule, a single searching protein and a single crowding particle in our system. The DNA molecule is represented as a chain consisting of $L+1$ binding sites as shown in Fig. 1. One of the sites serves as a target and it is located at the position m ($1 \leq m \leq L$). In addition, the chain contains a crowding particle, which also occupies 1 site. It can diffuse along the DNA chain with a rate u_{ob} (Fig. 1). The searching protein always starts from the solution that we label as a state 0. We assume that DNA is coiled in the solution, and the searching protein diffuses fast in the volume around DNA, so it can bind to any vacant site on DNA with equal probability with a binding rate k_{on} per each site (see Fig. 1). The attached protein can slide along the DNA chain with a diffusion rate u. The DNA-bound protein and the crowding particle interact with each other via a hard-core exclusion, i.e., they cannot pass each other. Finally, the protein can dissociate from DNA with a rate k_{off}, as shown in Fig. 1. Investigating the protein search dynamics, we average over all possible initial positions of the crowding particle on DNA. We can also extend the analysis to multiple crowders on DNA, but the main physics can be already understood from the case of the single crowding agent.

To analyze the search dynamics, we notice that if the diffusion rate of the crowding agent is relatively small in comparison with other rates in the system, then the protein will be able to find the target before the crowding agent can move from the original position. This suggests that the arrival times to the target can be well approximated as a linear combination of search times for the systems with static obstacles at different positions along the DNA chain. Such problems have been solved before [18], and this leads to the following approximate expression for the search time in our system,

$$\langle T_0^{ob} \rangle \simeq \frac{1}{L} \left(\sum_{l_{ob}=1}^{m-1} T_{ob}(l_{ob}) + \sum_{l_{ob}=1}^{L-m} T_{ob}(l_{ob}) \right), \tag{1}$$

where

$$T_i = \frac{k_{off} + k_{on}(L - S_i)}{k_{on} k_{off} S_i}, \tag{2}$$

which is valid for the static obstacle ($i = ob$) located at a distance l_{ob} from the target, as well as for a homogeneous chain without crowding molecules ($i = 0$). The auxiliary function S_i has a form [18]

$$S_{ob} = \frac{y(y^{-m} - y^m)}{(1-y)(y^m + y^{1-m})} + \frac{y(1 - y^{2l_{ob}-2})}{(1-y)(1 + y^{2l_{ob}-1})}, \tag{3}$$

with

$$y = \frac{k_{off} + 2u - \sqrt{k_{off}^2 + 4u k_{off}}}{2u}. \tag{4}$$

The results of our analytical calculations are presented in Fig. 2, where they are also compared with predictions from Monte Carlo simulations. Here the search times as a function of the scanning length $\lambda = \sqrt{u/k_{off}}$ (the average distance that the protein slides along the DNA before dissociating into the bulk solution) are computed for different mobilities of the crowding agent. As in the case of the protein search without obstacles [17], three search regimes are identified. For $\lambda < 1$ (weak nonspecific interactions between the searching proteins and DNA) the target can be found only by directly associating from the solution. This is the 3D search pathway. For intermediate values of the scanning length ($1 < \lambda < L$), the protein can also reach the target via sliding along the DNA chain. So this corresponds to a combination of the 3D and 1D search pathways. For strong nonspecific interactions the scanning length is large ($\lambda > L$), and 1D search pathways dominate.

Our approximate theory describes the protein search dynamics quite well for not very large scanning lengths ($\lambda < L$), when 3D pathways play important role in the search (see

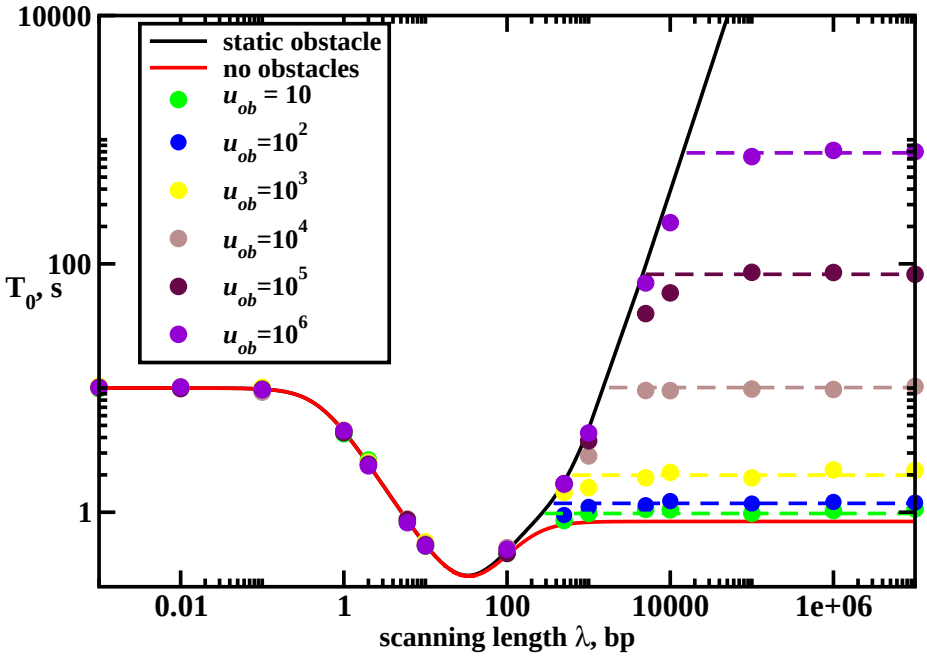

FIG. 2. A dynamic phase diagram for the protein search on DNA with a crowding particles. The DNA chain has the length $L = 10^3$ bp with the target at the position $m = L/2$. Parameters used for calculations are: $k_{on} = 0.1$ s^{-1}, $u = 10^5$ s^{-1} and different u_{ob} (in units of s^{-1}) as shown in the picture. Solid curves correspond to analytical results for homogeneous DNA without obstacles and for DNA with a static obstacle averaged over all initial positions of the crowding particle. Symbols describe Monte Carlo simulations whereas the dashed lines correspond to the approximate theory (see the text for the explanations).

Fig. 2). This is an expected result because the crowding particle cannot block the protein from finding the target all the time. But even for slow moving crowding objects there is always a regime in which the search times becomes independent of λ. Our theory that views the search dynamics as average over systems with static obstacles fails to predicts this. Another important observation from Fig. 2 is that increasing the mobility of the crowding agents leads to a situation when the protein effectively does not feel any crowding at all. The last observation is surprising since one expects that the protein and the crowding molecule interact many times by colliding into each other when the search is dominated by 1D motion.

To explain this peculiar dynamic behavior, we present the following arguments. Let us consider the 1D search regime when the protein is bound to DNA most of the time. We

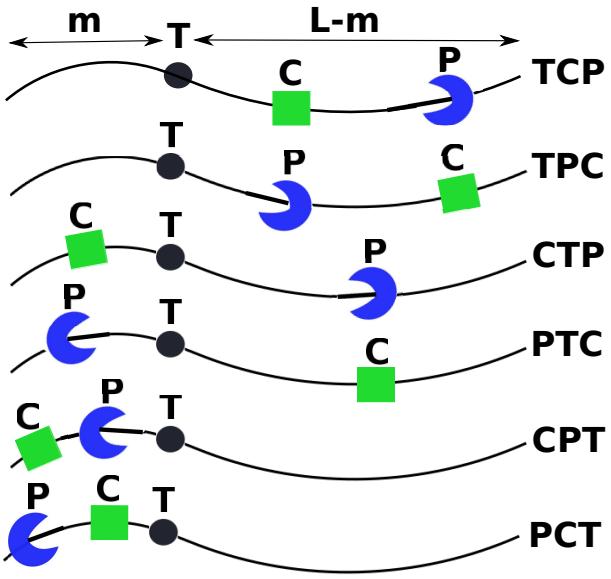

FIG. 3. A schematic view of different distributions of relevant particles during the 1D search on DNA. The capital letters T, P and C describe the target, the protein and the crowding agent, respectively. The labeling of configurations is explained in the text.

consider the search processes with all possible initial positions of relevant particles. Both the protein molecule and the crowding agent with equal probability can be found anywhere on DNA. If we abbreviate P, T and C as the protein, the target and the crowding agent, respectively, then there are six (= 3!) possible arrangements of these species relative to each other on the DNA chain: see Fig. 3. We label them as PTC, TPC, PCT, CPT, TCP and CTP. Due to the symmetry, some of them equally probable, i.e.,

$$P_{PTC} = P_{CTP}, \quad P_{CPT} = P_{PCT}, \quad P_{TPC} = P_{TCP}. \tag{5}$$

Because the target is fixed at the site m, the probabilities for different configurations are proportional to the product of segment lengths where the protein and the crowding molecule can be found. It can be written as (see Fig. 3)

$$P_{PTC} = P_{CTP} = Am(L-m), \quad P_{CPT} = P_{PCT} = Am^2, \quad P_{TPC} = P_{TCP} = A(L-m)^2, \tag{6}$$

where the normalization coefficient A can be found from $\sum_i P_i = 1$ for all six configurations. Then one can easily calculate

$$P_{CTP} = \frac{m(L-m)}{2(L^2+m^2-mL)}, \quad P_{PCT} = \frac{m^2}{2(L^2+m^2-mL)}, \quad P_{TCP} = \frac{(L-m)^2}{2(L^2+m^2-mL)}. \tag{7}$$

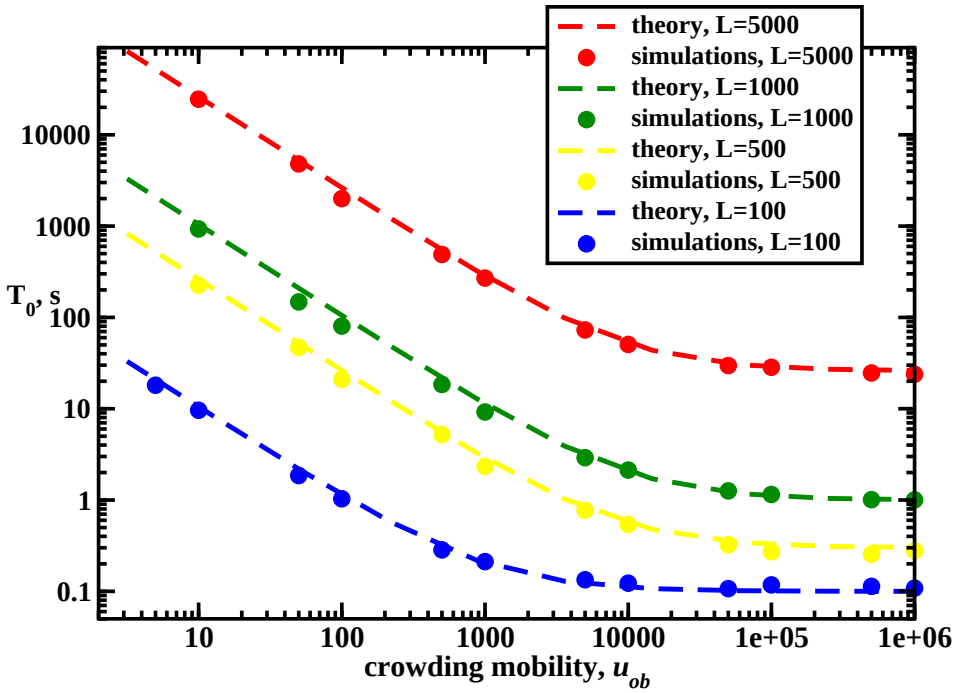

FIG. 4. Average times to reach the target as a function of the obstacle mobility u_{ob} (in units of s^{-1}). Parameters used for calculations are: $k_{on} = 0.1$ s^{-1}, $u = 10^5$ s^{-1}, $k_{off} = 10^{-7}$ s^{-1}, the position of the target is $m = L/2$, and variable DNA chain lengths L (in units of bp) as indicated in the plot. Symbols correspond to Monte Carlo computer simulations, and the dashed curves are theoretical predictions.

The overall search time is an average over all initial configurations presented in Fig. 3. However, it is clear that the largest contributions to the search time will come from the configurations where the crowding particle is found between the protein and the target, such as TCP and PCT configurations (see Fig. 3). In these cases, the crowding agent can block the protein from reaching the target. Therefore, for $u_{ob} < u$, we can approximate the contributions to the search time due to blocking as

$$\langle T_{bl} \rangle \simeq P_{PCT} T_{PCT} + P_{TCP} T_{TCP}. \tag{8}$$

The average blocking times from the initial configurations PCT and TCP can be estimated as

$$T_{PCT} \simeq \frac{(m/2)^2}{2u_{ob}}, \quad T_{TCP} \simeq \frac{((L-m)/2)^2}{2u_{ob}}, \tag{9}$$

which are the average times required for the crowding agent to pass the target and make the path open for the protein. For the crowding agent to do that they have to diffuse the

distances $m/2$ or $(L-m)/2$ for the PCT or TCP configurations, respectively (see Fig. 3). Substituting these expressions into Eqs. (7) and (8), leads to the total blocking time

$$\langle T_{bl}\rangle = \frac{m^4 + (L-m)^4}{16u_{ob}(L^2+m^2-mL)}. \tag{10}$$

When the target is in the middle, $m = L/2$, this equation gives $\langle T_{bl}\rangle = L^2/96u_{ob}$, while the blocking time is larger for the targets near the end of the DNA chain, $m/L \ll 1$, where we obtain

$$\langle T_{bl}\rangle \simeq \frac{L^2}{16u_{ob}}\left(1-\frac{3m}{L}\right). \tag{11}$$

Finally, the total search time can be found as a combination of the blocking time and the average search time for the system *without* the crowding particle,

$$\langle T_0^{ob}\rangle \simeq T_0^{(0)} + \langle T_{bl}\rangle, \tag{12}$$

where the explicit expressions for $T_0^{(0)}$ are known [17]. One should also note here that our theoretical arguments are valid for all target positions as long as they are not at the end of the DNA chain ($m \neq 1$ or $m \neq L$). In this case, the crowding molecule can never create paths for the searching protein to slide directly into the target. The details of the search dynamics in this case are discussed in the Supplementary materials.

The results of our theoretical calculations for the 1D search regime are presented in Fig. 2 and Fig. 4, and excellent agreement is found in comparison with Monte Carlo computer simulations. But the most important result from our theoretical arguments is the understanding of the role of the crowding agent mobility. Slow crowding molecules (small u_{ob}) significantly decelerate the search dynamics by blocking the sliding of proteins to the target (see Fig. 4). Increasing the mobility (large u_{ob}) lowers the blocking ability, and the search dynamics is quite fast (Fig. 4). Faster crowding molecules can move quickly beyond the target, freeing the path for the protein to reach the specific site. If the mobility of the crowding agent is low, then it will take a very long time before such path can be created. The last result is counter-intuitive since one would expect many collisions between the protein and the crowding agent that could slow down the search process. To understand this we notice that if the protein and the crowding agent are sitting on the neighboring sites, because of the high mobility of the crowding molecule it will move faster away, clearing the previously occupied site. This process will eventually lead to the protein reaching for the target.

It is important to discuss our theoretical predictions for realistic situations using transcription factors binding to their specific sites. It is known that DNA are heavily covered by many DNA-binding proteins [1, 27]. It should block the sliding of transcription factors to their targets, but because the mobility of many DNA-binding proteins is similar ($u \sim u_{ob}$) our theory suggests that this should not be a big problem. In addition, experiments indicate that the search for transcription factors is taking place at the conditions where both 3D and 1D search modes coexist [13, 14], and this should lower the effect of the crowding. Furthermore, the search could be even faster because of the lower length of DNA that should be scanned by the protein.

Conclusion. We present a theoretical approach that allowed us to explicitly investigate the role of the crowding on dynamics of protein search for specific target sites on DNA. We found that there are two important features of the search process that help proteins to avoid the expected negative effects of the crowding. One of them is a mobility of the crowding molecules on DNA, which increases the probability for direct sliding into the target. Fast crowding molecules move away and clear the path for the protein motion to the specific sites. Another one is a dimensionality of the search pathways. Increasing the contribution of 3D binding to the target via the bulk solution decreases the influence of the crowding agents. Our theoretical predictions are fully supported by extensive Monte Carlo computer simulations. The proposed theoretical method provides a simple and convenient way of explaining the dynamics of protein-DNA interactions using fundamental physical-chemical ideas. This should lead to better understanding the mechanisms of complex biological processes.

The work was supported by the Welch Foundation (Grant C-1559), by the NSF (Grant CHE-1360979), and by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY-1427654).

* tolya@rice.edu

[1] Alberts, B., et al. *Molecular Biology of Cell* 6th ed. (Garland Science, New York, 2014).

[2] A.D. Riggs, S. Bourgeois, M. Cohn, J. Mol. Biol. **48**, 67 (1970).

[3] S. E. Halford and J.F. Marko, Nucl. Acid Res. **32**, 3040 (2004).

[4] L.A. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J.S. Leith, and A. Kosmrlj, J. Phys. A:

Math. Theor. **42**, 434013 (2009).

[5] A.B. Kolomeisky, Phys. Chem. Chem. Phys. **13** 2088 (2011).

[6] O.G. Berg, R.B. Winter, and P.H. von Hippel, Biochemistry **20**, 20, 6948 (1981).

[7] O.G. Berg and P.H. von Hippel, Ann. Rev. Biophys. Biophys. Chem. **14**, 131 (1985).

[8] R.B. Winter, O.G. Berg, and P.H. von Hippel, Biochemistry **20**, 6961 (1981).

[9] D.M. Gowers, G.G. Wilson, G.G. and S.E. Halford, Proc. Natl. Acad. Sci. USA **102**, 15883 (2005).

[10] G. Kolesov, Z. Wunderlich, O.N. Laikova, M.S. Gelfand, and L.A. Mirny, Proc. Natl. Acad. Sci. USA **104**, 13948 (2007).

[11] Y.M. Wang, R.H. Austin, and E.C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).

[12] J. Elf, G.-W Li, and X.S. Xie, Science **316**, 1191 (2007).

[13] A. Tafvizi, F. Huang, J.S. Leith, A.R. Fersht, and L.A. Mirny, Biophys. J. **95**, L1-L3 (2008).

[14] P. Hammar, P. Leroy, A. Mahmutovic, E.G. Marklund, O.G. Berg, and J. Elf, Science **336**, 1595 (2012).

[15] A. Mahmutovic, O. G. Berg and J. Elf, Nucl. Acid Res. **43**, 3454 (2015).

[16] L. Zandarashvili, D. Vuzman, A. Esadze, Y. Takayama, D. Sahu, Y. Levy, and J. Iwahara, Proc. Natl. Acad. Sci. USA **109**, E1724 (2012).

[17] A. Veksler and A.B. Kolomeisky, J. Phys. Chem. B **117**, 12695 (2013).

[18] A.A. Shvets, M. Kochugaeva and A.B. Kolomeisky, to appear in J. Phys. Chem. B (2016).

[19] A. Marcovitz and Y. Levy, Biophys. J. **104**, 2042 (2013).

[20] E.F. Koslover, M.A.D. de la Rosa, and A.J. Spakowitz, Biophys. J. **101**, 856 (2011).

[21] A. Afek, J.L. Schipper, J. Horton, R. Gordan, D.V. Lukatsky, Proc. Natl. Acad. Sci. USA **111**, 17140 (2014).

[22] M. Sheinman, O. Benichou, Y. Kafri, and R. Voituriez, Rep. Prog. Phys. **75**, 026601 (2012).

[23] A. Tafvizi, F. Huang, A.R. Fersht, L.A. Mirny, and A.M. van Oijen, Proc. Natl. Acad. Sci. USA **108**, 563 (2011).

[24] A.B. Kolomeisky and A. Veksler, J. Chem. Phys. **136**, 125101 (2012) .

[25] A. Esadze, C.A. Kemme, A.B. Kolomeisky, and J. Iwahara, Nucl. Acids Res. **42** 7039 (2014).

[26] C. A. Brackley, M. E. Cates, and D. Marenduzzo, Phys. Rev. Lett. **109**, 168103 (2012).

[27] H.-X. Zhou, FEBS Lett. **587**, 1053 (2013).

[28] H. Flyvbjerg, S.A. Keatch, and D.T.F. Dryden, Nucl. Acids Res. **34** 2550 (2006).