Faraday Discussions

Accepted Manuscript

Faraday Discussions

Royal Society of Chemistry

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

Single molecule study of non-specific binding kinetics of LacI in mammalian cells

Laura Caccianini,[a] Davide Normanno,[b,c,d,e] Ignacio Izeddin[a,f,]* and Maxime Dahan[a,]*

Many key cellular processes are controlled by the association of DNA-binding proteins (DBPs) to specific sites. The kinetics of the search process leading to the binding of DBPs to their target locus is largely determined by transient interactions with
10 non-cognate DNA. Using single-molecule microscopy we studied the dynamics and non-specific binding to DNA of the Lac repressor (LacI) in the environment of mammalian nuclei. We measured the distribution of LacI-DNA binding times at non-cognate sites and determined the mean residence time to be
15 τ_{1D} = 182 ms. This non-specific interaction time, measured in the context of an exogenous system such as that of human U2OS cells, is remarkably different than that reported for the LacI in its native environment in *E. coli* (< 5 ms). Such a striking difference (more than 30 fold) suggests that the
20 genome, its organization, and the nuclear environment of mammalian cells play an important role on the dynamics of DBPs and their non-specific DNA interactions. Furthermore, we found that the distribution of off-target binding times follows a power law, similar to what reported also for TetR in
25 U2OS cells. We argue that a possible molecular origin of such a power law distribution of residence times is the large variability of non-cognate sequences found in the mammalian nucleus by the diffusing DBPs.

1 Introduction

30 Biological processes such as transcription, replication or repair, and their regulation are controlled by the association of DBPs to specific sites. Understanding the mechanisms by which DBPs find their target sequence—typically of the order of ten base pairs (bp) among a genome composed by millions to billions of bp—is thus critical [1]. The current paradigm for the target search process involves a combination
35 of three-dimensional diffusion in the nucleoplasm and one-dimensional sliding on DNA (see review in Mirny *et al.*[2]). This facilitated diffusion (FD) model was first proposed[3] to explain the (apparently[4]) faster-than-diffusion DNA association rate of LacI proteins reported in the seminal work of Riggs and colleagues[5]. A key element in the FD model is that the non-specific interaction of DBPs with DNA largely
40 controls the search rate. Therefore, it is an important goal to properly determine the non-specific binding times (referred hereafter as τ_{1D}).

In past years, single molecule (SM) assays[6-8] have started complementing

conventional approaches such as recovery after photobleaching or fluorescence correlation spectroscopy (see review in Mueller *et al.*[9]) for the study of nuclear dynamics and organization. In particular, SM experiments provide direct ways of measuring non-specific binding kinetics in live cells. The first SM measurements of transcription factor dynamics in living cells were those reported by Elf and collaborators for LacI in *E. coli*[10] in which they could infer a rapid non-specific association to DNA, with $\tau_{1D} < 5$ ms, and later probe the FD model[11]. More recently, SM measurements have been extended to mammalian systems. Compared to *E. coli*, DBPs in mammalian cells must find their specific DNA sequence among a much larger pool of competing non-specific sites. Experiments in mammalian cells have now been performed on a few different systems, including p53 in human lung carcinoma H1299 cells[12], Sox2/Oct4 in mouse embryonic stem cells[13], glucorticoid receptors in human breast cancer MCF-7 cells[14], and TetR in human osteosarcoma U2OS cells[15]. Interestingly, in all these cases, the mean non-specific interaction times appear to be on the time of a few seconds, several orders of magnitude longer than the values reported for LacI in *E. coli*[10]. It thus raises the question of whether this markedly different non-specific binding kinetics is due to the intrinsic properties of the DBPs that were studied, to the difference in DNA organization between bacteria and mammalian cells, or to the experimental techniques that were used. In the present study, we address this question by measuring the diffusion dynamics and the non-specific binding properties of LacI in U2OS cells, in order to directly compare our results with those obtained on the same protein but in bacteria.

2 Results

To address the non-specific LacI-DNA interactions in mammalian cells, we chose an engineered cell line containing repeated insertions of the *lac*O binding sequence. More precisely, we used the human osteosarcoma cell line U2OS 2-6-3, which contains 200 insertions of an artificial gene each including also 256 *lac*O repeats[16]. Since the human genome does not contain any *lac*O sequence, the sites of specific and non-specific interactions between LacI molecules and DNA could be unambiguously separated with this cellular system.

We transfected U2OS 2-6-3 cells with two different plasmids: NLS-LacI-GFP and NLS-LacI-HaloTag. After overnight transfection, we stained the cells with HaloTag TMR ligand (at high concentration). The NLS-LacI-GFP plasmid served as a visual aid to unambiguously locate the target locus. Fluorescence images of transfected nuclei in both GFP and TMR channels displayed a homogenous and diffuse population of LacI molecules and bright fluorescent loci corresponding to an accumulation of LacI recruited at the target sequence (Fig. 1).
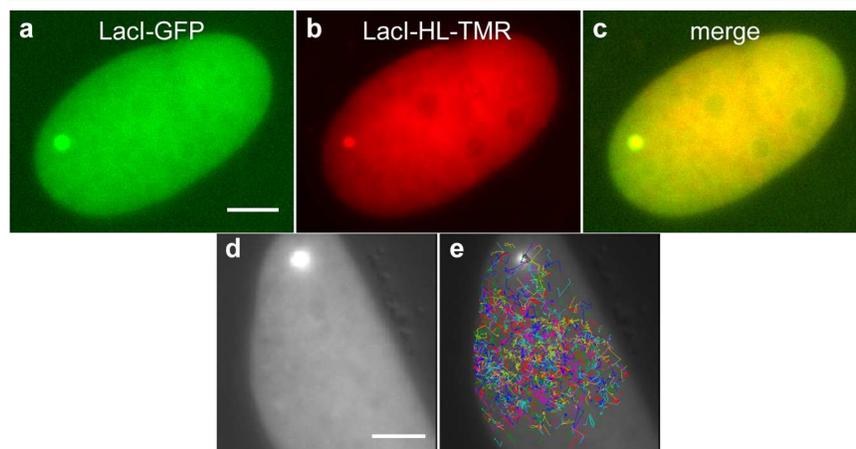
*Figure 1: **(a-c)** Fluorescence images of a U2OS 2-6-3 cell nucleus co-transfected with NLS-LacI-GFP (green) and NLS-LacI-HaloTag labeled with HaloTag Ligand-TMR at high concentration (1 nM) (red). The accumulation of proteins on the bright spot indicates the position of the binding locus (target site). **(d-e)** Fluorescence*
5 *image of a cell transfected with NLS-LacI-GFP (d) and NLS-LacI-HaloTag stained with low concentration (25 pM) of HaloTag Ligand-TMR for SM experiments. In panel (e), overlaid trajectories resulting from our single-particle detection and raw-tracking analysis; note both, traces at the specific binding site and elsewhere in the nucleoplasm, only the latter considered for our non-specific binding analysis. Scale*
10 *bar, 5 μm.*

2.1 Single-molecule dynamics of LacI in mammalian cells

To investigate LacI dynamics at the single molecule level, we used a low concentration of HaloTag TMR ligand (25 pM) to label the cells, which allowed us
15 to work in a diluted regime, suited for single-particle-tracking (SPT) experiments. We acquired SM fluorescence images of LacI-HaloTag Ligand-TMR proteins, under continuous illumination with a 561 nm laser (at ~1 kW/cm^2), with an EMCCD camera at a frame rate of 197 Hz. In these imaging conditions, the signal from a single dye was sufficient to overcome the background coming from out-of-focus
20 dyes and from endogeneous autofluorescence. We next localized the proteins position in each frame (with an accuracy ~20-40 nm) and computed a set of individual trajectories away from the specific target locus, which was identified in the GFP channel (Fig. 1). The length of the trajectories is limited either by axial diffusion or by photobleaching of the dye. We performed experiments in 6 cells
25 accumulating a total of 1,525 individual traces longer than 6 frames, with a median trace length of 29 frames (145 ms). We then computed the time-averaged mean square displacement (MSD) of each individual track. The MSD curves of a sample of 10 individual traces are plotted in Fig. 2a and reveal a wide range of diffusion behavior of LacI in the nucleus of mammalian cells. To quantify the diffusion
30 dynamics, we calculated the instantaneous diffusion coefficient D_{Inst} from the linear fit of the initial points (2nd to 5th) of each MSD curve. We found that the distribution of instantaneous diffusion coefficients spanned over three orders of magnitude, from 0.01 μm^2/s to more than 10 μm^2/s (Fig. 2b). The distribution was

approximately bimodal with a first peak centered around 5 $\mu m^2/s$, which we attributed to proteins diffusing in the nucleoplasm, and a second one centered around 0.3 $\mu m^2/s$ and corresponding to proteins interacting (at least transiently) with DNA. The fraction of the traces displaying an instantaneous diffusion coefficient D_{Inst} smaller than 0.1 $\mu m^2/s$, which we considered the benchmark for an immobile protein—being the diffusivity of chromatin-bound markers as H2B histones of the same order[7,14]—was equal to $f \sim 8\%$. This fraction f can be expressed as $\tau_{1D}/(\tau_{1D}+\tau_{3D})$ where τ_{3D} is the average time spent in 3D diffusion before binding to
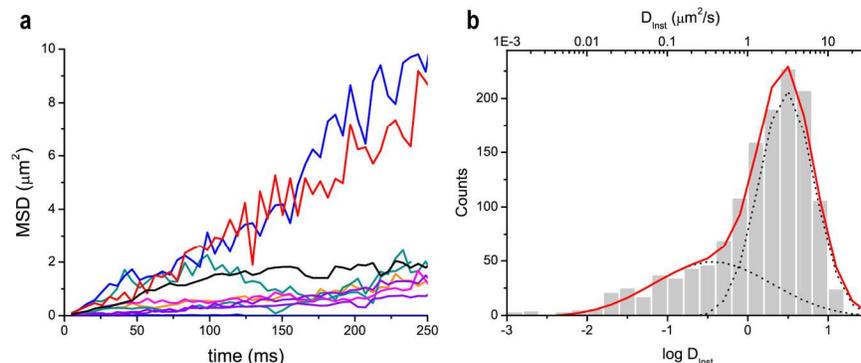
DNA. Thus, the value of f yields an estimate of the ratio $\tau_{3D} / \tau_{1D} \sim 11.5$. Note that $1/\tau_{3D}$ and $1/\tau_{1D}$ corresponds to the association and dissociation rates, respectively.

*Figure 2: **(a)** Mean square displacement (MSD) curves of 10 individual traces of LacI-HL-TMR molecules, exemplifying the broad distribution of LacI diffusivity. **(b)** Histogram of instantaneous diffusion coefficients D_{Inst} from individual MSD curves, extending over three orders of magnitude (1,525 trajectories, 6 cells).*

2.2 Non-specific binding time of LacI to DNA

We next aimed to more carefully examine the role of non-specific DNA interactions on the dynamics of LacI molecules and estimate τ_{1D}. To this end, we followed an approach recently developed for the analysis of the non-specific interactions of the Tet repressor in mammalian cells[15]. In brief, we analyzed the changes of instantaneous diffusion coefficient along each individual trajectory by computing D_{Inst} in a running window of 40 ms. We identified events in which the molecule reversely switched from fast to much slower diffusion (Fig. 3a), likely corresponding to a transition from 3D diffusion to non-specific binding to DNA. The distribution of the interaction times was approximately exponential with a decay time of 37 ms (Fig. 3b), a value which is ~8 times larger than the non-specific binding times of LacI en *E. coli* ($\tau < 5$ ms)[10] but smaller than other transcription factors reported in eukaryotes, like p53 (~ 1.7 s)[12] and Sox2 (~ 0.8 s)[8]. Nevertheless, we observed also long binding events, exceeding several seconds, which were difficult to capture due to the photobleaching of the fluorophore. In order to examine them and circumvent photobleaching, we performed time-lapse experiments in which 5 ms exposure images were interlapsed with periods of 50 and 500 ms. We then considered the proteins that remained localized within a region of half a pixel

(80 nm) and calculated their survival probability $SP(\tau)$, defined as the probability to stay bound for a time longer than τ (Fig. 3c). The same analysis was also performed for the case of continuous imaging experiments and yielded to a power law distribution (t^{γ}) of the non-specific binding survival probability, with $\gamma = -1.12$, all the more evident in the normalized curves over all the time regimes under
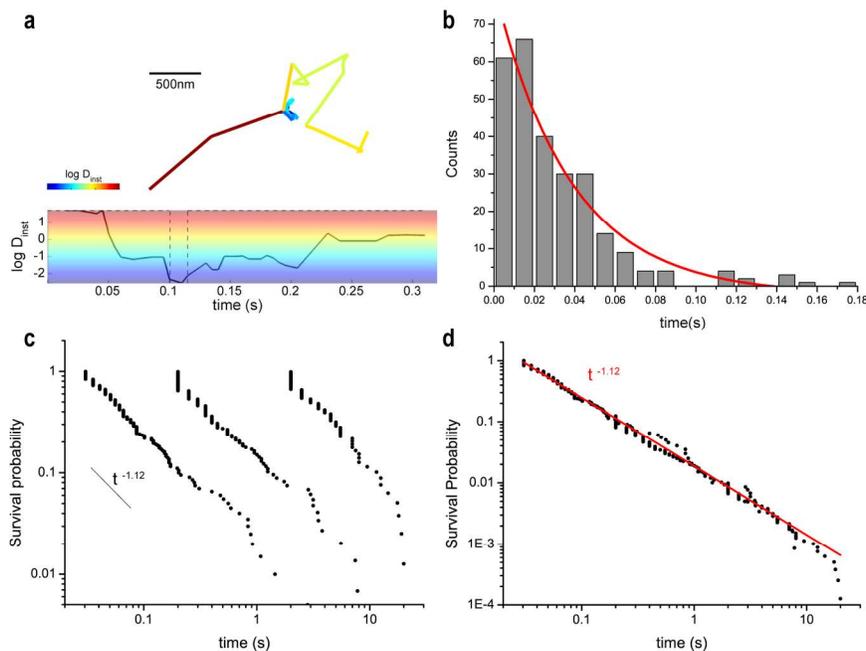
investigation (Fig. 3d). The numerical integration of the global survival probability curve provided us with an estimation of the mean non-specific binding time $\tau_{1D} = 182$ ms, from which we could deduce $\tau_{3D} = 2.1$ s.

*Figure 3: **(a)** Example of a molecule undergoing transition from the diffusive state to non-specific DNA binding and its instantaneous diffusion coefficient D_{Inst} calculated over a running window of 40 ms. **(b)** Distribution of binding times retrieved with the running window analysis, measured with a 95% confidence that the events represent real deviation from diffusive behavior. In red, exponential fit with a characteristic decay time of $\tau = 37$ ms. **(c)** Survival probability of the non-specific binding events of LacI molecules for (from left to right) continuous imaging (197 Hz) and time-lapse (20 and 2 Hz) experiments. The black line is a visual guide corresponding to t^{γ} with $\gamma = -1.12$. (d) Global survival probability curve resulting from rescaling of the three temporal series. The red line represents the best fit to data with a power law t^{γ} function, yelding to $\gamma = -1.12$.*

3 Discussion

Our SM experiments provide quantitative information on the diffusion dynamics of LacI proteins in the nucleus of human cells. In particular, with this exogenous

system we gained insight into the role of non-specific interactions with non-cognate DNA sequences and their impact on nuclear protein's diffusion. Compared to the results obtained in *E. coli*, we note several striking differences. First, the mean non-specific binding times (182 ms) in U2OS is about 30 times larger than the upper
5 bound (5 ms) estimated in *E. coli*[10]. While we do not think it explains the discrepancy between the results, it is worth mentioning a technical difference between our experiments and those in bacteria. In *E. coli*, the non-specific binding time was not obtained, as in our case, by analyzing individual trajectoriess, due to the limited brightness and photostability of the fluorescent proteins used to tag LacI
10 in those experiments. Instead, an estimate of τ_{1D} was indirectly determined by measuring the broadening of the point-spread function (PSF) as a function of the illumination time. A width of the PSF exceeding the diffraction limit was interpreted as evidence that the protein had detached during the recording time.

15 In our experiments, we were able to recover the full distribution of non-specific binding times, and not only the mean value. The observed distribution is broad and scales as a power law. The longest binding events exceeded a few seconds and were not found for the LacI in bacteria. Interestingly, a power law distribution has been found also for the *SP* of non-specific binding times of TetR in U2OS 2-6-3 cells[15],
20 but with a different exponent (-0.7 *vs.* -1.12 for the LacI). The molecular origin of these broad distributions is not well understood yet. A possible cause is the large variability of non-specific sequences encountered by the diffusing proteins in the nucleus of mammalian cells, going from random sequences to quasi-specific sites, leading to heterogeneity in the binding affinities. Notably, it can be shown that from
25 the cumulative sum of an exponential distribution of DNA-binding energies, a power law distribution of binding times, can emerge[17,18]. This scenario is in contrast to bacteria, where the genomic DNA is much shorter than in eukaryotes and the number of non-cognate sequences, which might act as (transient) decoy sites, much smaller.

30

The fraction *f* of time spent non-specifically bound to DNA is higher in bacteria (~87 % in *E. coli vs.* ~8% in U2OS 2-6-3 cells). This is equivalent to say that τ_{3D} is much shorter in bacteria, less than 1 ms compared to ~2 s in our measurements. This in turn means that the association rate $1/\tau_{3D}$ is much higher in bacteria. To evaluate
35 the theoretical diffusion-limited association rate *k*, it is necessary to estimate the concentration *c* of DNA (expressed in terms of base-pairs). Assuming a volume of 1 μm^3 for *E. coli* and of 500 μm^3 for a human cell nucleus and given the respective size of the genome (4.6 millions bp and 3 billions bp), *c* is in fact comparable in both cases and on the order of 10 mM. Thus, we expect $k = D\,c\,a$ (where $D \sim 5$
40 μm^2/s is the diffusion coefficient and *a* is taken equal to 1 nm) to be on the order of 10^5 s^{-1}. Even though neither of the measurements (in *E. coli* or in human cells) reached the diffusion limit, the discrepancy with the measurement of τ_{3D} in U2OS is especially striking. This might be due to the fact that genomic DNA is globally more accessible in prokaryotes than in higher eukaryotes, where DNA is compacted
45 within nucleosomes and can be covered by many other proteins, specifically or non-specifically bound and competing for the same accessible sites. Also, the bound and unbound times for different DBPs in mammalian cells are closer to what we observed for LacI proteins in the same cellular system than in bacteria (Table 1). This points to the potential role of the nuclear environment of mammalian cells in

dictating the diffusion dynamics of DBPs and their interactions with DNA.

Protein	Cell	τ_{1D}	τ_{3D}	Ref.
LacI	U2OS	182 ms	2.1 s	This study
LacI	*E. coli*	< 5 ms	< 1 ms	Elf *et al*[10]
TetR	U2OS	2 s	6 s	Normanno *et al.*[15]
p53	H1299	1.7 s	1.8 s	Mazza *et al.*[12]
GR	MCF / U2OS	1.5 s	–	Gebhardt *et al.*[14]
Sox2	Mouse ES	0.8 s	3.3 s	Chen *et al.*[8]

5 *Table 1: Comparison between the 3D diffusion and the non-specific binding times reported for different TFs in several cellular lines from different organisms.*

Conclusion

10 Our experiments illustrate how kinetic rates can now be retrieved using single molecule imaging of DBPs in live cells. They also point to open questions on the role of chromatin organization on the specific or non-specific association rates with DNA. Unquestionably, together with powerful genome-editing tools[19] and advanced imaging techniques[20,21], single molecule techniques are expected to play an 15 important role in our understanding of nuclear dynamics and organization.

References

20 *a Laboratoire Physico-Chimie Curie, Institut Curie, PSL Research University, CNRS UMR168, Université Pierre et Marie Curie-Paris 6, 11 rue Pierre et Marie Curie, 75005 Paris, France*
b Centre de Recherche en Cancérologie de Marseille, CNRS UMR7258, 27 boulevard Leï Roure, 13273 Marseilles, France
c INSERM U1068, 13009 Marseilles, France
25 *d Institut Paoli-Calmettes, 13009 Marseilles, France*
e Aix-Marseille Université, 13284 Marseilles, France
f Currently at: Institut Langevin, ESPCI ParisTech, PSL Research University, 1 rue Jussieu, 75005 Paris, France
* Correspondence: maxime.dahan@curie.fr (MD), ignacio.izeddin@espci.fr (II)

30

1.   D. Normanno, M. Dahan, and X. Darzacq, *BBA - Gene Regulatory Mechanisms*, 2012, **1819**, 482–493.
2.   L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, *J. Phys. A:*
35   *Math. Theor.*, 2009, **42**, 434013.
3.   O. G. Berg, R. B. Winter, and P. H. von Hippel, *Biochemistry*, 1981, **20**, 6929–6948.
4.   S. E. Halford, *Biochem. Soc. Trans*, 2009, **37**, 343.
5.   A. Riggs, S. Bourgeois, and M. Cohn, *Journal of molecular biology*, 1970, **53**, 401:417.
6.   I. I. Cisse, I. Izeddin, S. Z. Causse, L. Boudarene, A. Senecal, L. Muresan, C. Dugast-
40   Darzacq, B. Hajj, M. Dahan, and X. Darzacq, *Science*, 2013, **341**, 664–667.
7.   I. Izeddin, V. Récamier, L. Bosanac, I. I. Cissé, L. Boudarene, C. Dugast-Darzacq, F. Proux, O. Bénichou, R. Voituriez, O. Bensaude, M. Dahan, and X. Darzacq, *eLife*, 2014,

3.

8.   J. Chen, Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, and Z. Liu, *Cell*, 2014, **156**, 1274–1285.

9.   F. Mueller, T. J. Stasevich, D. Mazza, and J. G. McNally, *Critical Reviews in Biochemistry and Molecular Biology*, 2013, **48**, 492–514.

10. J. Elf, G. W. Li, and X. S. Xie, *Science*, 2007, **316**, 1191–1194.

11. P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, *Science*, 2012, **336**, 1595–1598.

12. D. Mazza, A. Abernathy, N. Golob, T. Morisaki, and J. G. McNally, *Nucleic Acids Research*, 2012.

13. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang, *Cell*, 2013, **155**, 1479–1491.

14. J. C. M. Gebhardt, D. M. Suter, R. Roy, Z. W. Zhao, A. R. Chapman, S. Basu, T. Maniatis, and X. S. Xie, *Nature Methods*, 2013.

15. D. Normanno, L. Boudarene, C. Dugast-Darzacq, J. Chen, C. Richter, F. Proux, O. Bénichou, R. Voituriez, X. Darzacq, and M. Dahan, *Nature Communications*, in press, 2015.

16. S. M. Janicki, T. Tsukamoto, S. E. Salghetti, W. P. Tansey, R. Sachidanandam, K. V. Prasanth, T. Ried, Y. Shav-Tal, E. Bertrand, R. H. Singer, and D. L. SPECTOR, *Cell*, 2004, **116**, 683–698.

17. J. P. Bouchaud and A. Georges, *Physics Reports-Review Section of Physics Letters*, 1990, **195**, 127–293.

18. R. Metzler and J. Klafter, *The Random walk's guide to anomalous diffusion*, 2000.

19. J. D. Sander and J. K. Joung, *Nat Biotechnol*, 2014, 1–9.

20. S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, X. Darzacq, C. Wu, C. I. Bargmann, D. A. Agard, M. Dahan, and M. G. L. Gustafsson, *Nature Methods*, 2013, **10**, 60–63.

21. B. Hajj, M. El Beheiry, I. Izeddin, X. Darzacq, and M. Dahan, *Phys. Chem. Chem. Phys.*, 2014.