

On the origin of affinity–specificity negative correlation in DNA–probe interactions

R. Murugan*

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai, 400005, India

Received 6 January 2005; accepted 30 January 2005

Available online 21 April 2005

Abstract

Affinity (the fastness) and specificity (the correctness) are the two important factors that decide the efficiency of a nucleic acid probe to target its specific site on a DNA lattice. DNA–probe interactions differ from protein–ligand interactions in a way that here the specificity and the affinity of the interactions correlate negatively with each other. We present a simple phenomenological theory to explain the negative correlation between the specificity and the affinity of the probe towards its target site on the template DNA under solution conditions. We show that a simple random jump model can explain this fact and we also predict that the negative correlation between the affinity and specificity diminishes as the temperature increases or the viscosity of the medium decreases. Moreover, the length of target DNA and the distance between the initial position of the probe on the template DNA lattice and the target site increases the magnitude of affinity–specificity negative correlation. These results are consistent with experimental observations. Finally we propose practical strategies to coherently improve the specificity and the affinity with respect to important molecular biological techniques such as PCR and Southern blotting.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Specificity; Affinity; DNA–probe; PCR

1. Introduction

Recognition of a particular site on a large stretch of DNA by a small stretch of another DNA or RNA (these are called as probes that are mostly radio labeled and used to detect the presence of a specific site on a large set of genomic DNA) is an important phenomenon not only in molecular biology but also in many biotechnological and forensic applications such as Southern blotting, Northern blotting, DNA finger printing and nucleic acid arrays [1]. Among the characteristics of a good probe, the fastness of recognition (denoted as affinity) and the correctness of recognition (denoted as specificity) are very important for a better efficiency. Generally, sequence specificity and binding affinity of DNA–DNA, DNA–RNA and/or RNA–RNA interactions negatively correlate (also known as anticorrelation) with

each other [2], except for some special cases, such as oligonucleotide aptamers [3,4]. Here the negative correlation means that as the specificity for the target sequence increases the affinity will decrease and vice versa which is contradictory to most of the biomolecular processes such as enzyme–substrate and drug–protein interactions.

The main feature that differentiates nucleic acid interactions from other interactions is that in the former case the recognition process is actually a one-dimensional nucleation-zipping type process [5] whereas in the latter case it is generally a three dimensional lock-and-key type [6]. Both affinity and specificity in nucleic acids interactions play crucial role in important biological process such as replication of DNA, transcription, translation and genetic-recombination. Therefore there is always an urge to develop optimizing tools to design nucleic acid probes with enhanced specificity and affinity. In this context new type of probes are actively being designed and synthesized such as PNAs (Protein Nucleic Acids) [7] which has a protein like backbone with ordinary DNA bases. Binding of PNAs

* Tel.: +91 22 2280 4545x2737; fax: +91 22 2280 4610, 2280-4611.

E-mail address: RMurugan@gmail.com.

to DNA showed a remarkable enhancement of specificity and affinity. On the other hand, it is necessary to understand the origin of affinity–specificity negative correlation by theoretical or experimental means so that one can develop simple selection rules for better designing of nucleic acid probes. Recent results show that a two step model [5] can enhance the correlation between the affinity and the specificity [8] in case of PNA–DNA interaction. Nevertheless, the abovementioned model is a deterministic one and it does not include the underlying real microscopic processes which is actually stochastic in nature. In this article, we present a simple phenomenological random jump model which can predict a negative correlation between the specificity and affinity in probes–DNA interactions.

2. Theory and mathematical derivations

Let us denote a stretch of DNA by a one dimensional lattice of x_m base pairs in length. The probe is considered as a one dimensional unbiased random walker with a random step size of k base pairs (or a random jump) on the DNA lattice. Here the target DNA is larger than the probe DNA (in size or molecular weight) and mostly present in a supercoiled conformation which is true in most of the in vivo and in vitro conditions. The target site is located at the x_n th position such that $0 \leq x_n \leq x_m$ (see Fig. 1). The basic idea is as follows. Following the two-step model, the probe first non-specifically binds to the DNA lattice (in this case at $x=x_0$) and then performs a one dimensional search via performing random jumps with a jump size of k base pairs to locate the target site under nonspecifically bound-condition. Here a jump size k means that starting from a position x , in the next jump the probe can be found anywhere in the interval $x-k \leq x \leq x+k$ with equal probabilities. Here the helical ends of the target DNA that are

denoted as $x=0$ and $x=x_m$ are the reflecting boundaries and $x=x_n \pm \varepsilon$ is the absorbing boundary-layer i.e. once the probe hits this set of points, it binds there very tightly. Aforementioned reflecting boundary conditions are necessary because otherwise our two-state assumption will be violated. Let us assume that the jump size k obeys the inequality $k \geq 2\varepsilon$ which is valid since in the real situation the target DNA is in super coiled state that allows larger jump sizes. Now let us assume that there exists two points $\{a, b\}$ on the DNA lattice such that $a=x_n-\varepsilon$ and $b=x_n+\varepsilon$. This means that the range of recognition is $x_n \pm \varepsilon$ where $a \leq x_n \leq b$ i.e. when the probe hits any one of this set of points under random jump conditions it will be absorbed. Suppose, if we need highly specific recognition at $x=x_n$, we should insist the condition $\varepsilon \rightarrow 0$ due to fact that highly specific recognition can have only one possible way of binding which is present only at $x=x_n$ and no mismatch / looping is allowed. Therefore, the parameter ε is inversely proportional to the specificity of the probe. On the other hand, the rate of recognition is directly proportional to the affinity of the probe to its target site. To find the correlation between affinity and specificity of interaction of probe with its target, it is necessary to find the functional relationship between the rate of recognition and the specificity interval factor ε . Let us assume that at time $t=0$, the probe non-specifically binds to DNA at the position $x=x_0$ and performs a one dimensional unbiased random jumps with jump size of k base pairs along the DNA lattice to locate its target site. In the first nonspecifically-binding-step, the affinity and the specificity are not correlated due to fact that until the probe meets DNA, their dynamics are not correlated. With this background, we derive the expression for the correlation between the affinity and the specificity in the second i.e. searching step. The probability $P(x,t)$ of finding the probe at the x th base pair position of target DNA at any time t by unbiased random jumps obeys the following generalized master equation.

$$\partial_t P = w \sum_{i=1}^k [P_{i+1,t} + P_{i-1,t} - 2P_{i,t}]. \quad (1)$$

Here w is the unbiased transition rate. Since the target DNA is very large compared to the size of the probe DNA we can approximate the master Eq. (1) by its corresponding Fokker–Plank equation (FPE) [9].

$$\partial_t P(x,t) = (D/2) \partial_x^2 P(x,t). \quad (2)$$

Here $D = 2w \sum_{i=1}^k i^2$ is the phenomenological diffusion coefficient [9,11]. When the probe is added to a solution containing the template DNA with the corresponding target site, it first performs a three dimensional search to locate the template DNA before actually binding in a nonspecific manner. Since the two-step assumption has been insisted, once the probe non-specifically binds to DNA it exists in bound form until it locates the target site. As we stated in the earlier sections, this implicitly indicates that the helical ends of template DNA that are denoted by the base-pair positions

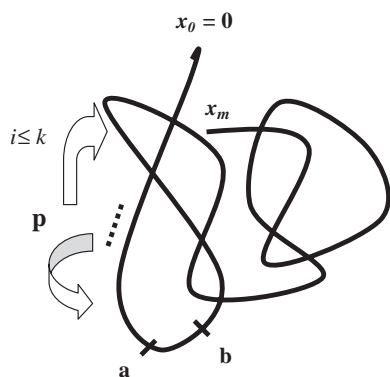


Fig. 1. Recognition of a specific site range (a, b) on a coiled state DNA by a small DNA probe p (dotted line) by a random jump with a jump size of k base pairs which means that starting from a position x , in the next step the probe can be found in the interval $x-k \leq x \leq x+k$ with equal probabilities. Here $x=0$ and $x=x_m$ are the reflecting boundaries and the absorbing range is (a, b) where $\varepsilon=(b-a)/2$ is the specificity determining factor i.e. specificity $\propto 1/\varepsilon$.

$x=0$ and $x=x_m$ act as the reflecting boundaries i.e. when the probe hit these helical ends then they eventually turned back to the template DNA rather than escaping into the solution. The reflecting boundary conditions are satisfied when $|\partial_x P(x,t)|_{x=0} = |\partial_x P(x,t)|_{x=x_m} = 0$. Therefore with the initial condition $P(x,t|x_0,0) = \delta(x-x_0)$ and in absence of a probability sink, the formal solution of Eq. (2) reads as,

$$P(x,t|x_0,0) = \frac{1}{x_m} + \frac{2}{x_m} \times \sum_{k=1}^{\infty} \cos\left(\frac{k\pi}{x_m}x_0\right) \cos\left(\frac{k\pi}{x_m}x\right) e^{-\frac{k^2\pi^2 Dt}{2x_m^2}}. \quad (3)$$

Here $\delta(x-x_0)$ is the Dirac-delta function. In the absence of a probability sink, the steady state probability P_s of finding the probe anywhere on the DNA lattice can be given as $\lim_{t \rightarrow \infty} P(x,t|x_0,0) = P_s = 1/x_m$. We know that the range of target sites spans the interval (a, b) and since $k \geq 2\varepsilon$ one should note that unlike in case of a random walk with unit step size under random jump conditions, each and every point in the interval (a, b) can be an absorbing point. Moreover there is a certain probability for a probe to jump across the interval (a, b) from the interval $0 \leq x \leq a$ to the interval $b \leq x \leq x_m$ and vice versa without actually getting absorbed. Therefore, the probability flux P_{flux}^{a-b} per unit time in this interval is given by the line integral of the probability current (i.e. $S(x,t) = 2^{-1} D \partial_x P(x,t)$) in (a, b) that is equal to $P_{\text{flux}}^{a-b} = \int_a^b S(x,t) dx$. In case of discrete step random jump one has to replace this integral with a sum of probability densities on lattice points in the interval (a, b) as $P_{\text{flux}}^{a-b} = \sum_{x=a}^b S(x,t)$. Since we approximated the master equation by the Fokker–Plank equation here, we use the integral representation. However, one should note that in case of a random walk with unit step size aforementioned summation formula or the line-integral formula are not valid since the probability current (and thus the probability flux) can be defined only in one point on the DNA lattice at a time. Now the total probability flux (denoted as Π) in the target interval (a, b) can be given as $\Pi = \int_0^{\infty} P_{\text{flux}}^{a-b} dt$. The average time taken by the probe (starting from the position $x=x_0$) to fall into the interval $x=x_n \pm \varepsilon$ is given by the conditional mean first passage time (denoted in a short form as MFPT) [9] that is denoted as $\tau(x_0) = \frac{1}{\Pi} \int_0^{\infty} t P_{\text{flux}}^{a-b} dt$. Finally, it is easy to verify that the conditional mean first passage time $\tau(x_0)$ with $x_0=0$ is given by,

$$\tau(0) = \frac{2x_m^2}{\pi^2 D} \left\{ \frac{\sum_{k=1}^{\infty} \frac{1}{k^4} \left\{ \cos\left(\frac{k\pi}{x_m}b\right) - \cos\left(\frac{k\pi}{x_m}a\right) \right\}}{\sum_{k=1}^{\infty} \frac{1}{k^2} \left\{ \cos\left(\frac{k\pi}{x_m}b\right) - \cos\left(\frac{k\pi}{x_m}a\right) \right\}} \right\}. \quad (4)$$

Here we use the condition $x_0=0$ to simplify our analysis. Now using the well-known relations (Eqs. (5) and (6)),

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \cos(kx) = \frac{\pi^2}{6} - \frac{\pi x}{2} + \frac{x^2}{4} + O(x^6), \quad (5)$$

$$\sum_{k=1}^{\infty} \frac{1}{k^4} \cos(kx) = \frac{\pi^4}{90} - \frac{\pi^2 x^2}{12} + \frac{\pi x^3}{12} - \frac{x^4}{48} + O(x^7). \quad (6)$$

We finally arrive at $\tau(0)$ as a function of b and a .

$$\tau(0) = \frac{x_m}{3D} = \left\{ \frac{-(b+a) + \frac{1}{x_m}(b^2 + ab + a^2) - \frac{1}{4x_m^2}(b^3 + a^2b + ab^2 + a^3)}{\frac{1}{2x_m}(b+a) - 1} \right\}. \quad (7)$$

Using the relations $a=x_n-\varepsilon$ and $b=x_n+\varepsilon$ and defining a new parameter $\delta=2\varepsilon$, Eq. (7) can be rewritten as,

$$\tau(0) = \frac{x_m x_n}{3D} \left\{ \frac{2 - \frac{1}{x_m} \left(3x_n + \frac{\delta^2}{4x_n} \right) + \frac{1}{x_m^2} \left(x_n^2 + \frac{\delta^2}{4} \right)}{1 - \frac{x_n}{x_m}} \right\} = \alpha - \beta \delta^2. \quad (8)$$

where

$$\alpha = \frac{x_m x_n}{3D} \left(\frac{x_n^2}{x_m^2} - \frac{3x_n}{x_m} + 2 \right) \times \left(1 - \frac{x_n}{x_m} \right)^{-1} \text{ and } \beta = [12D]^{-1}.$$

It is clear from Eq. (4) that $\lim_{\delta \rightarrow \delta_m} \tau(0) = 0$ where δ_m is defined as $\delta_m = \sqrt{\alpha\beta^{-1}}$ and $\lim_{x_n \rightarrow x_m} \lim_{\delta \rightarrow 0} \tau(0) = x_m^2/3D$. The targeting rate k_A of the probe is simply the inverse of the mean first passage time. Therefore we obtain the following relationship,

$$k_A = \tau(0)^{-1} = [\alpha - \beta \delta^2]^{-1}. \quad (9)$$

Since δ is the specificity factor i.e. specificity $\propto \delta^{-1}$ and k_A is the affinity parameter, Eq. (9) clearly demonstrates the negative correlation between the affinity and specificity of probe–target interactions. Here one should note that the range of affinity rate is $0 \leq k_A \leq \infty$ and the range of the specificity factor is $0 \leq \delta \leq \sqrt{\alpha\beta^{-1}}$ which is difficult to handle and visualize. In order to circumvent this problem we rescale the variable by defining,

$$s = 1 - \sqrt{\frac{\beta}{\alpha}} \delta = 1 - \left[1 - \frac{x_n}{x_m} \right]^{1/2} \times \left[x_m x_n \left(\frac{x_n^2}{x_m^2} - \frac{3x_n}{x_m} + 2 \right) \right]^{-1/2} \delta. \quad (10)$$

$$k_A = \frac{1}{\alpha(1 - (1-s)^2)} = \frac{3D}{x_m x_n} \times \frac{1 - \frac{x_n}{x_m}}{\frac{x_n^2}{x_m^2} - \frac{3x_n}{x_m} + 2} \times \frac{1}{1 - (1-s)^2}. \quad (11)$$

And by redefining $A = 1 - e^{-k_A}$ one obtains the relation,

$$A = 1 - e^{-[x(1-(1-s)^2)]^{-1}}. \quad (12)$$

Here the ranges of new variables A and s are $0 \leq s \leq 1$ and $1 - e^{-x} \leq A \leq 1$ which is actually very easy to visualize

and handle. We should note that when the specificity factor $s=1$ (i.e. maximum specificity and specificity $\propto s$), $A_{\min}=1-e^{-\alpha^{-1}}$ which is not equal to zero indicating that the probe would still exhibits a minimum affinity. Therefore, A_{\min} is the magnitude of the affinity with highest specificity.

Since the recognition of the specific site by the probe happens in solution condition (both in vivo and in vitro), both the jump size k and the transition rate w are strongly influenced by the viscosity as well as by the temperature of the solution. The major free energy source of nonspecific interaction between the template DNA and the probe is hydrophobic in nature and the specific interaction is mainly due to the hydrogen bonding between the complementary base pairs of the probe and the template DNA. The hydrophobic interactions will therefore introduce additional impediment to the dynamics of the probe on the target DNA. In other words, the microscopic viscosity in the vicinity of the template DNA is expected to be slightly higher than the viscosity of the bulk-solution. However since the magnitude of the nonspecific interactions between the probe and the template DNA is in turn weakened by the electrostatic repulsion, for all the practical purposes we can assume that the microscopic viscosity is almost equal to the viscosity of bulk-solution. Due to this fact the one dimensional (phenomenological) microscopic-diffusion coefficient D associated with probe can also be expressed as $D=kT\gamma^{-1}$ (Stokes–Einstein relationship) where γ is the internal friction coefficient, and therefore A_{\min} can also be rewritten as follows.

$$A_{\min} = 1 - \exp \left[- \frac{3kT}{\gamma x_m x_n} \frac{1 - \frac{x_n}{x_m}}{\frac{x_n^2}{x_m^2} - \frac{3x_n}{x_m} + 2} \right] \quad (13)$$

When $x_n \ll x_m$, Eq. (13) further simplifies to,

$$A_{\min} = 1 - \exp \left[- \frac{3kT}{2\gamma x_m x_n} \right] \quad (14)$$

From Eqs. (13) and (14) we can conclude that when x_m and x_n are fixed (actually in biotechnological applications point of view these are all constants), under maximum specificity conditions ($s=1$) the affinity factor (A) can be enhanced by either increasing the temperature or reducing the viscosity of the medium. These facts have been earlier observed in the reassociation of complementary strands of DNA [10].

3. Results and discussions

Specificity and affinity are the two central factors that decide the efficiency of any molecular probe. In case of protein–ligand system, enhancement of one factor automatically enhances the other in a coherent manner which is due to fact that the dynamics of protein and ligand are positively correlated. Unfortunately, in case of nucleic acid probes the specificity and the affinity are negatively correlated. Never-

theless, nucleic acid probes are important constituents in many molecular biological techniques. In this context our theory provides many important strategies to reduce the affinity–specificity negative correlation. Correlation diagram for the affinity parameter A and the specificity parameter s at different values of α^{-1} is shown in Fig. 2. Here the affinity and specificity are displayed in percentage. This clearly shows that as α^{-1} increases, the negative correlation between the specificity and affinity diminishes. At very high α^{-1} , it is easy to conclude that the correlation between specificity and affinity will become zero or uncorrelated. This is due to the fact that the two-step assumption will break down at very high α^{-1} and the problem becomes eventually three dimensional in nature.

The dynamics of the probe on the template–DNA lattice can be modeled as Brownian motion of the probe in potential well with infinite potential walls at the helical ends of the template DNA lattice. However, aforementioned the potential well is not ideally flat but may contain many local minima apart from the global minima that is present at the specific site (denoted as G_g). Here a local minimum (denoted as G_l) can be understood as the binding site of the probe with a significant number of mismatches. The global minimum can be understood as the binding site of the probe with minimum or no mismatches and the non-specific binding is simply a binding with the number of mismatches that is equal to the size of the probe itself (denoted as G_n). The specificity is directly proportional to the difference in the potentials between the specific site and the local minima ($G_g - G_l$) or the nonspecific site ($G_g - G_n$). It is obvious to note that $|G_g - G_l| < |G_g - G_n|$ and therefore the presence of local minima will actually reduce the specificity of the probe towards the target site on the template DNA. Similarly, it is easy to conclude that the affinity of the probe to the target site on the template DNA is also reduced in presence of local minima. This is due to the fact that the probe is temporarily trapped in the local minima which in turn retard the dynamics of the probe towards its target site on the

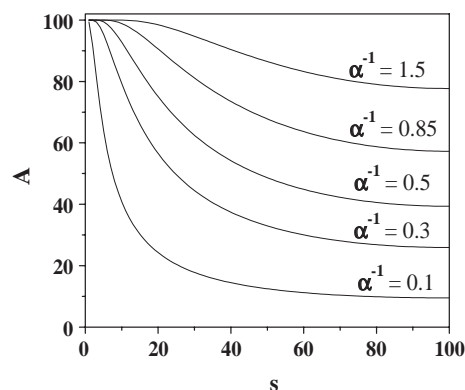


Fig. 2. Correlation diagram for the affinity parameter A and the specificity parameter s at different values of α^{-1} which linearly scales with temperature. Here the affinity and specificity are given in terms of percentage. This clearly indicates that as α^{-1} increases, the negative correlation between the specificity and affinity diminishes.

template DNA. Suppose let us assume that all the aforementioned arguments are valid at the temperature T , then by raising the temperature of the system from T to another temperature T_{opt} in such a way that the inequality $|G_1 - G_n|/RT_{\text{opt}} = 1$ is satisfied, one can coherently improve the specificity as well as the affinity of the probe towards its target site on the template DNA under consideration. Because by raising the temperature we mainly avoid the temporary trapping of probe in the local minima which in turn increase the affinity towards the specific site whereas the specificity remains almost constant. In other words, we can minimize the affinity–specificity anticorrelation by raising the temperature. However, beyond certain temperature T_m one may obtain the inequality $|G_g - G_n|/RT_m = 1$ which means that the specific interaction is similar to the nonspecific interaction between the probe and the template DNA i.e. the affinity–specificity anticorrelation disappears.

Since α^{-1} in Eq. (12) is directly proportional to the temperature, the affinity–specificity negative correlation behavior can be easily tuned by adjusting the temperature. This is important in many molecular biological techniques such as PCR and Southern blotting. The PCR based techniques like DNA-fingerprinting and RAPD (Random Amplified Polymorphic DNA) requires specific amplification of the target site using a random primer i.e. affinity to the target site is low. In order to avoid or reduce the non-specific amplification, one can increase the annealing temperature or reduce the viscosity of the medium. One more striking feature we observe from Eq. (14) is that the length of the target DNA actually increases the negative correlation behavior which means that in case of shorter DNA the affinity–specificity negative correlation will tend to zero. This is an important issue to be discussed. Under the evolutionary process the genome size of organisms is also increasing which in turn increases the negative correlation between the affinity and specificity in case of vital enzymes such as DNA polymerase (here the target site is the origin of replication) and RNA polymerase (here the target site is the promoter) with respect to their target site on the genome. The organism can trespass this problem either by condensing its DNA or developing some machinery to energetically drive the probe towards the target site. Under condensing condition, it has been shown earlier that [11] random jumps are facilitated leading to the enhancement of phenomenological diffusion coefficient ($D = 2w\sum_{i=1}^k i^2$, where k is the jump size and w is the transition rate) which in turn reduces the affinity–specificity anticorrelation. Similarly, the affinity–specificity anticorrelation of DNA–probe interactions increases with the distance of the target position (i.e. x_n) from the initial position. Due to this fact when a probe comes closer to its target site the affinity–specificity negative correlation tends to minimum. In biochemical terms, this phenomenon is known as positive-cooperative effect. On the other hand, when the probe moves away from its target site, affinity–specificity negative correlation increases which is a negative-cooperative effect.

4. Conclusions

We present a simple phenomenological theory to explain the negative correlation between the specificity and the affinity of the DNA–probe to target its target site on the template DNA in solution conditions. We show that a simple random jump model can explain this fact and we predict that the negative correlation between the affinity and specificity diminishes as the temperature increases or the viscosity of the medium decreases. Moreover, the length of target DNA and the distance between the initial position of the probe on the DNA lattice and the target site actually increases the affinity–specificity negative correlation. These results are consistent with experimental observations. Finally, we propose practical strategies to improve the specificity and the affinity with respect to important molecular biological techniques such as PCR and Southern blotting.

Acknowledgments

This work was supported by TIFR, Mumbai. The author thanks the Referee for constructive comments. The author is recipient of Kanwal-Rekhi fellowship for career development.

References

- [1] B. Lewin in *Genes VIII*, Prentice Hall, (2003); B. Alberts, et al. in *Molecular Biology of the Cell*, Garland Science (2002).
- [2] A. Lomakin, M.D. Frank-Kamenetskii, A theoretical analysis of specificity of nucleic acid interactions with oligonucleotides and peptide nucleic acids, *J. Mol. Biol.* 276 (1998) 57–70.
- [3] M. Famulok, et al., Nucleic acid aptamers—from selection in vitro to applications in vivo, *Acc. Chem. Res.* 33 (2000) 591–599.
- [4] E.N. Brody, L.J. Gold, Aptamers as therapeutic and diagnostic agents, *Biotechnology* 74 (2000) 5–13.
- [5] P.H. von Hippel, O.G. Berg, On the specificity of DNA–protein interactions, *Proc. Natl. Acad. Sci. U. S. A.* 83 (1986) 1608–1612.
- [6] D.E. Koshland, The key-lock theory and the induced fit theory, *Angew. Chem., Int. Ed. Engl.* 33 (1994) 2375–2378.
- [7] P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science* 254 (1991) 1497.
- [8] V.V. Demidov, M.D. Frank-Kamenetskii, Two sides of the coin: affinity and specificity of nucleic acid interactions, *TIBS* 29 (2004) 62.
- [9] (a) C.W. Gardiner, *Handbook of Stochastic Methods*, Springer-Verlag, Berlin, 1986, p. 142;
(b) N.G. VanKampen, *Stochastic Processes in Physics and Chemistry*, Elsevier Science, Amsterdam, 2004, p. 295.
- [10] J.G. Wetmur, N. Davidson, DNA renaturation kinetics, *J. Mol. Biol.* 31 (1968) 349–370.
- [11] R. Murugan, DNA–protein interactions under random jump conditions, *Phys. Rev., E* 69 (2004) (011911).