RESEARCH ARTICLE

Theory on the Mechanism of DNA Renaturation: Stochastic Nucleation and Zipping

Gnanapragasam Niranjani, Rajamanickam Murugan*

Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India

* rmurugan@gmail.com

Abstract

Renaturation of the complementary single strands of DNA is one of the important processes that requires better understanding in the view of molecular biology and biological physics. Here we develop a stochastic dynamical model on the DNA renaturation. According to our model there are at least three steps in the renaturation process viz. nonspecific-contact formation, correct-contact formation and nucleation, and zipping. Most of the earlier two-state models combined nucleation with nonspecific-contact formation step. In our model we suggest that it is considerably meaningful when we combine the nucleation with the zipping since nucleation is the initial step of zipping and nucleated and zipping molecules are indistinguishable. Nonspecific contact formation step is a pure three-dimensional diffusion controlled collision process. Whereas nucleation involves several rounds of one-dimensional slithering and internal displacement dynamics of one single strand of DNA on the other complementary strand in the process of searching for the correct-contact and then initiate nucleation. Upon nucleation, the stochastic zipping follows to generate a fully renatured double stranded DNA. It seems that the square-root dependency of the overall renaturation rate constant on the length of reacting single strands originates mainly from the geometric constraints in the diffusion controlled nonspecific-contact formation step. Further the inverse scaling of the renaturation rate on the viscosity of reaction medium also originates from nonspecific contact formation step. On the other hand the inverse scaling of the renaturation rate with the sequence complexity originates from the stochastic zipping which involves several rounds of crossing over the free-energy barrier at microscopic levels. When the sequence of renaturing single strands of DNA is repetitive with less complexity then the cooperative effects will not be noticeable since the parallel zipping will be a dominant enhancing factor. However for DNA strands with high sequence complexity and length one needs to consider the underlying cooperative effects both at microscopic and macroscopic levels to explain various scaling behaviours of the overall renaturation rate.

Introduction

The biological function of DNA depends largely on its double stranded helical structure and its ability to unwind and rewind in a reversible manner. The double stranded structure of DNA (dsDNA) is mainly stabilized by weak hydrogen bonds between the nitrogen bases of complementary single strands (c-ssDNAs) and the hydrophobic forces arising from the base-stacking within the core of dsDNA polymer [1–2] These weak interactions melt down upon heating the solution containing dsDNA beyond the melting temperature which in turn yields the corresponding c-ssDNAs. These single strands exactly reunite (hybridize) back into their original double stranded helical form upon cooling the solution below the melting temperature. Melting temperature of dsDNA is defined as the temperature at which precisely half of the dsDNA melts into corresponding c-ssDNAs. Several molecular biological processes such as transcription, translation and replication and *in vitro* laboratory techniques are solely based on the reversible unwinding-rewinding property of dsDNA. Understanding the dynamics and mechanism of renaturation of c-ssDNAs in solution is important in recombination, design of primers for polymerase chain reaction, design of oligonucleotide probes for microarray chips, various membrane blotting techniques and other related DNA fingerprinting technologies [2–5]. In this context, several models describing the process of renaturation of c-ssDNAs in aqueous solution have been developed and experimentally verified [6–29]. Detailed understanding of the mechanism of renaturation of c-ssDNA at microscopic level is one of the important contemporary topics of interest in molecular biology and biological physics.

Renaturation of c-ssDNAs was initially thought [6] as one-step bimolecular second order chemical kinetic process as in **Scheme I** of Fig 1. Several experimental observations could not be explained by a simple one-step second order kinetics. One of such observations is that irrespective of the experimental conditions the overall bimolecular rate constant was directly proportional to the square-root of the average length of c-ssDNAs and inversely proportional to its sequence complexity. Whereas a one-step process predicted that the bimolecular collision rate was directly proportional to the average length of c-ssDNA. To comply with various experimental observations, Wetmur and Davidson [6] suggested a detailed two-step renaturation model that comprised of nucleation and zipping steps as in **Scheme II** of Fig 1. They had shown that the overall second order rate constant associated with the renaturation phenomenon could be expressed as a function of the average length of the sheared DNA and sequence complexity of the reacting c-ssDNAs. Here the sequence complexity is defined as the length of DNA with unique nucleotide sequence pattern [Fig 1].

Detailed renaturation studies on the sheared genomic DNA suggests [2] the presence of two different types of genomic sequences viz. nonrepetitive and repetitive. Here the degree of repetitiveness can vary from moderate to high. Moderately repetitive DNA contains short sequences that are repeated $10–10^3$ times whereas highly repetitive part of the genomic DNA contains several thousand repeats of short sequences with length of <100 bases (1 base = 3.4 x 10^{-10}m). The overall length of nonrepetitive sequences (the sequence complexity) increases along with the genome size and it seems to attain a plateau at ~2 x 10^9 bases [2]. Wetmur and Davidson further formulated a theoretical model on the renaturation phenomenon according to which the overall bimolecular rate was directly proportional to the nucleation rate apart from the ratio between the average length of c-ssDNA and its sequence complexity. They argued that the nucleation rate must be inversely proportional to the square-root of the average length of c-ssDNA. As a consequence of these facts, the experimentally observed overall bimolecular rate associated with renaturation is directly proportional to the square-root of the length of c-ssDNA. They further suggested that the inverse scaling of the nucleation rate on the length of c-ssDNA polymer must be owing to either the thermodynamic excluded volume effects

Scheme I

$$S + S' \xrightarrow{k_{HQ}} H$$

Scheme II

$$S + S' \underset{k_r}{\overset{k_N}{\rightleftharpoons}} Y_N \xrightarrow{k_Z} H$$
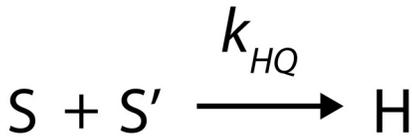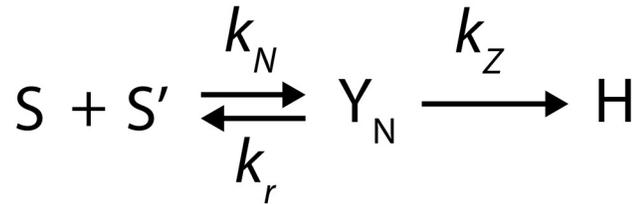
Sequence complexity	L	c
S1 5'-ATGATCTACGTAGTATGGCCTATGCCGGCT-3'	30	30
S2 5'-ATGATCTACGATGATCTACGATGATCTACG-3'	30	10
S3 5'-ATGATATGATATGATATGATATGATATGAT-3'	30	05

Fig 1. **Earlier models on DNA renaturation kinetics.** Renaturation of c-ssDNA strands was initially modelled as one-step bimolecular collision rate process as in **Scheme I** with an overall bimolecular association rate of k_{HQ}. S and S' are the concentrations of c-ssDNA and H is the concentration of completely renatured dsDNA. According to this model **Scheme I** the overall renaturation rate k_{HQ} should scale with the length of the reacting c-ssDNA strands in a linear manner. However experiments revealed a square-root dependency of the renaturation rate on the length of reacting c-ssDNA strands. To comply with the experimental observation a two-step mechanism was proposed as in **Scheme II** which comprised of nucleation and zipping. In this mechanism the nucleation rate (k_N) is inversely proportional to the square-root of the length of c-ssDNA strands. This scaling seems to emerge as a consequence of excluded volume effects of c-ssDNA polymer. Whereas the zipping rate (k_Z) is directly proportional to the length of c-ssDNA strands (L) and inversely proportional to the sequence complexity (c). Since the overall renaturation rate is directly proportional to both k_N and k_Z one observes a square-root dependency of the overall renaturation rate on the length of c-ssDNA strands. To generalize nucleation is modelled as a reversible process with a dissociation rate constant k_r. Here Y_N is the concentration of the nucleus. **Sequence complexity** of c-ssDNA is defined as the length of DNA with unique sequence pattern. For example consider sequences S1, S2 and S3 all with length of $L = 30$ bases. By definition the sequence complexity of S1 is $c = 30$ bases. Complexity of S2 is $c = 10$ bases since it has 3 repeats of ATGATCTACG with 10 bases length. In the same way, the complexity of S3 is $c = 5$ bases since it has 6 repeats of ATGAT with 5 bases length. The copy numbers $\rho = L/c$ of S1, S2 and S3 are 1, 3 and 6 respectively. The zipping rate in two-step renaturation models as in **Scheme II** is directly proportional to this copy number ρ. This means that the overall renaturation rate is inversely proportional to the sequence complexity of the reacting c-ssDNA strands.

doi:10.1371/journal.pone.0153172.g001

associated with the intra-strand dynamics or stearic hindrance associated with the diffusion controlled interpenetration of c-ssDNAs that is essential for the nucleation step.

Here one should note that the two-step model of Wetmur-Davidson will be inconsistent whenever the sequence complexity has same magnitude as that of the length of c-ssDNA. Subsequent experimental studies on the renaturation phenomenon were mainly focussed [16–28] on unravelling the molecular mechanisms and the underlying thermodynamics and kinetics aspects. In line with these experiments, several theoretical and computational models [6, 7, 14–17, 23–27] were also developed to explain the observed scaling behaviours of the overall second order renaturation rate on the size of reacting c-ssDNAs, temperature, ionic strength and viscosity of the reaction medium. Recent studies [23–27] considered either the excluded volume effects acting on the intrastrand dynamics or stearic hindrance associated with the diffusion controlled interpenetration of c-ssDNAs to explain the observed scaling behaviours of the overall renaturation rate on the length of c-ssDNA strands. Diffusion based models provide

correct viscosity dependence of overall renaturation rate constant compared to the models based on the framework of transition state theories (TST).

Recently nucleation step in renaturation was modelled as an escape over free energy barrier [27] within the framework of Kramer's theory that deals with the dynamics of Brownian particle over a potential energy barrier. It was argued [27] that the square root dependency of nucleation rate on the length of c-ssDNA mainly originates from the entropic component of the free energy barrier associated with the Kramer's escape problem. Though this approach appeared to be a reasonable one, nature of the reaction coordinate and potential energy barrier associated with the renaturation process were not clearly defined. Moreover the exact connection between the entropic component of the free energy barrier and the observed scaling behaviour was not clearly established in detail. Here the entropic barrier associated with the nucleation step must originate from the freely moving single stranded overhangs of colliding c-ssDNA strands.

According to the current theoretical understandings over experimental and computational observations [23–29], the renaturation process should have at least three distinct steps namely (**a**) formation of nonspecific contact (**b**) nucleation or correct contact formation and (**c**) zipping [Fig 2A]. In the first step, the reacting c-ssDNAs collide with each other via three-dimensional (3D) diffusion controlled routes. This results in the formation of Watson-Crick (WC) base pairs at random nonspecific contacts between the reacting c-ssDNAs. Such nonspecific WC contacts randomly translocate along c-ssDNAs either via thermally driven one-dimensional (1D) slithering dynamics or internal displacement [29] mechanisms [Figs 2A and 3A–3C] until finding the correct-contact and initiate the nucleation process which is in turn followed by spontaneous zippering of c-ssDNAs.

Here one should note that steps (**a**) and (**b**) are purely stochastic dynamical processes similar to that of the site-specific DNA-protein interactions [17, 30–31]. Additionally zippering step (**c**) will be a stochastic dynamical process since there is a finite probability of dissociation of c-ssDNA strands at each step of the zipping reaction. These mean that one needs to apply stochastic dynamics based arguments rather than merely thermodynamics based ones to explain the observed scaling behaviours and underlying mechanisms. In this paper we will formulate such a stochastic dynamics based theoretical framework of renaturation phenomenon and explain various scaling properties associated with the renaturation rate.

Results

Theoretical Formulation of DNA Renaturation Kinetics

At a coarse grained level one can consider the c-ssDNA polymer as a chain of nitrogen base beads with an average bond length of 1 base. The process of renaturation essentially involves collisions between the clusters of nitrogen base beads corresponding to c-ssDNA strands. The basic steps involved in the process of renaturation of c-ssDNA strands viz. (**a**) formation of nonspecific contact (**b**) nucleation or correct contact formation and (**c**) zippering can be well represented by **Scheme III** of Fig 2. Here 3D diffusion controlled collisions between the nitrogen base beads of c-ssDNA strands lead to the formation non-specific contacts between them that results in the formation of cn-ssDNA [Figs 2A and 3A]. Under non-specifically bound condition, the probe strand of cn-ssDNA searches over the template strand for the correct contact via a combination of 1D and 3D diffusion. Here the 1D diffusion comprises of facilitating processes such as slithering and internal displacement dynamics. Slithering dynamics involves [Fig 3B] a random search with unit base step size whereas internal displacement mechanism involves an inchworm type movement of the probe strand over the template strand with a step size of few hundreds to thousand bases [Fig 3C]. One should note that internal displacement mainly depends on the 3D conformational state of the template c-ssDNA strand.
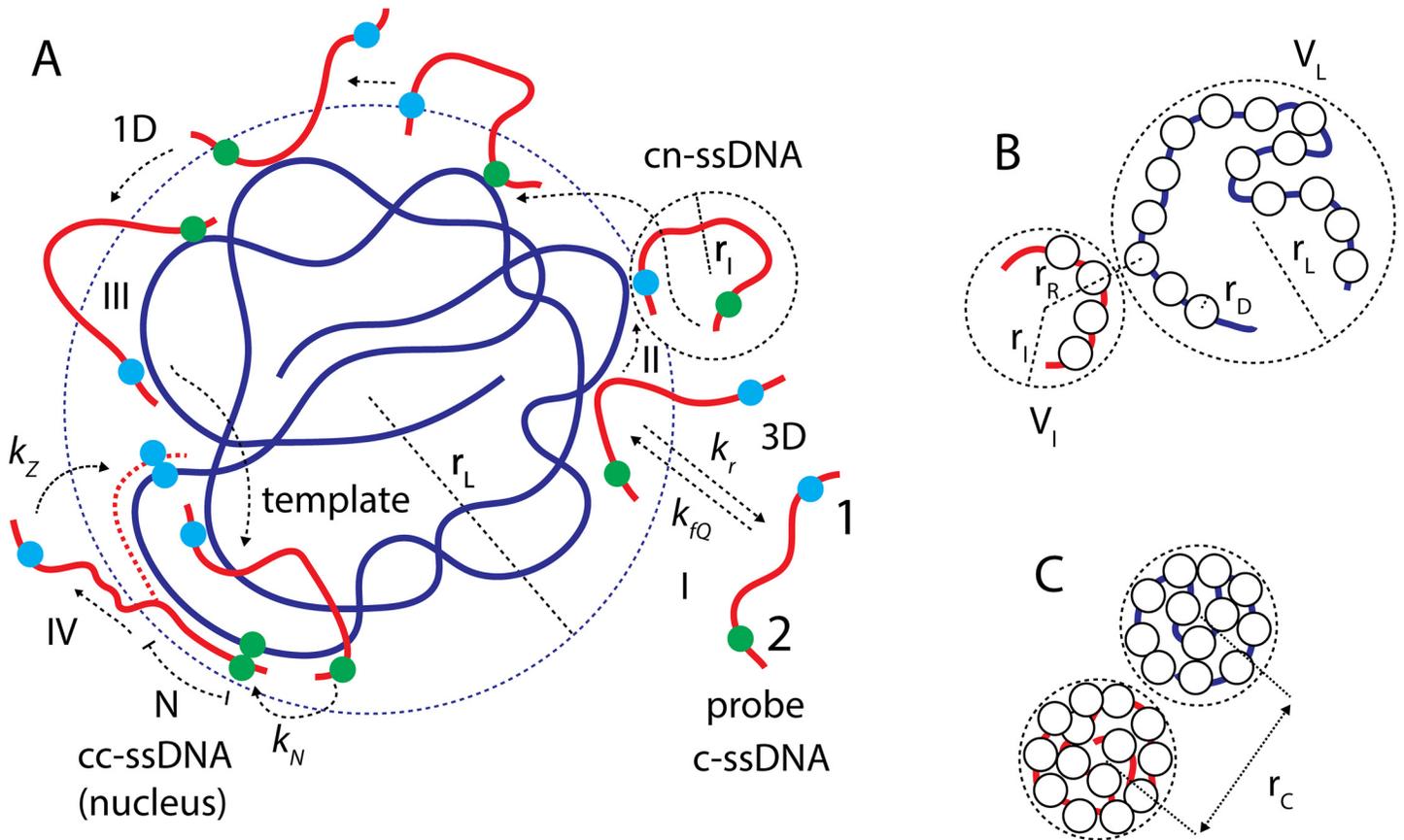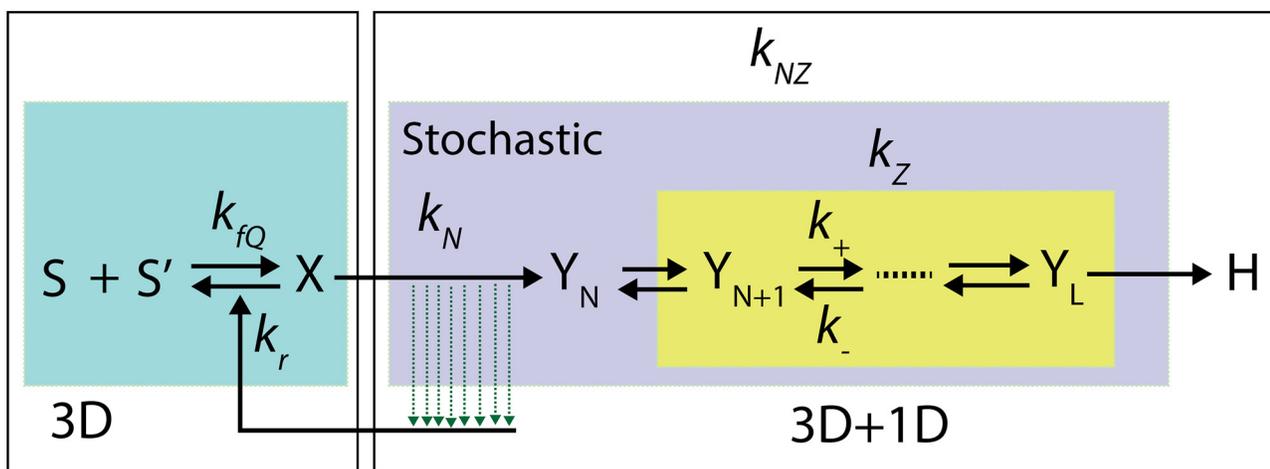
Fig 2. Basic steps of DNA renaturation phenomenon. A. Three basic steps in the renaturation of complementary single strands of DNA (c-ssDNA) are viz. nonspecific-contact formation, nucleation and zipping. Two arbitrary locations on the probe c-ssDNA are marked as 1 and 2 (blue and green dots respectively). Nonspecific-contact formation (cn-ssDNA) is purely a three dimensional (3D) diffusion controlled collision rate process (I) where the rate constant associated with the formation of nonspecific-contact scales with the length of colliding c-ssDNA molecules in a square root manner and it scales with the solvent viscosity in an inverse manner. Nucleation involves a one dimensional (1D) slithering dynamics (II) of one strand on the other strand of cn-ssDNA in the process of searching for correct-contact (cc-ssDNA). Internal displacement dynamics through inchworm movements (III) of one complementary strand on the other can facilitate the 1D diffusion dynamics. Upon finding the correct-contact and forming the nucleus, zipping of cc-ssDNA step (IV) follows.

Scheme III. According to this scheme both the nucleation and zipping are coupled stochastic dynamical processes. In this scheme there are three distinct steps in the process of renaturation viz. 3D diffusion mediated nonspecific contact formation with an on rate of k_{fQ} and off-rate of k_r, 1D and 3D diffusion mediated nucleation step with rate k_N and zipping which is a 1D diffusion process with a rate of k_Z. Before forming a successful nucleus with a critical size of N bases the colliding c-ssDNA strands undergo several rounds of nonspecific contact formation to form cn-ssDNA, 1D diffusion of one of the cn-ssDNA strands over the other and then dissociation. Upon formation of the nucleus (cc-ssDNA with N numbers of correct contacts) zipping process commences. Since nucleated cc-ssDNA is indistinguishable from zipping one, it is more appropriate to combine the nucleation with the zipping with an overall rate of $k_{NZ} = 1/(1/k_N + 1/k_Z)$ rather than with the nonspecific-contact formation step as in **Scheme II**. Conformational state of the reacting c-ssDNA molecules seems to significantly affect the reaction mechanism and scaling relationships associated with the overall renaturation rate on the size of the system. **B**, **C**. We can model the c-ssDNA chains as clusters of nitrogen bases so that the overall bimolecular rate associated with the formation of nonspecific contacts between spatially distributed base-clusters of c-ssDNAs is proportional of the product of concentrations of the total nitrogen bases in c-ssDNA molecules. The cylindrical surface area $C_M \sim 2\pi r_D M$ of a c-ssDNA molecule with a radius of r_D bases will be confined within the spherical solvent shell with surface area $V_M \simeq 4\pi r_M^2$ ($M = L$ for template and $M = l$ for probe c-ssDNA strands) where r_M is the radius of gyration of the respective c-ssDNA molecule. Under strongly condensed state of c-ssDNA one finds that $V_M < C_M$ (**C**) and when the DNA polymer is in a relaxed state then one find that $V_M > C_M$. At a coarse grained level one can model the bases of c-ssDNA as a chain of spherical beads with radius r_D. Under relaxed conformational state all these nitrogen base beads are distributed on the surface of the spherical solvent shell that covers a c-ssDNA molecule (**B**). Under condensed conformational state of c-ssDNA molecules significant fraction of nitrogen base beads will be inaccessible to the inflowing c-ssDNA molecules since they are buried inside the matrix of condensed c-ssDNA (**C**).

Condensed conformational state of template c-ssDNA always favours internal displacements while relaxed conformational state favours the slithering dynamics. Here slithering and internal displacement mechanism are similar to that of the sliding and intersegmental transfer dynamics as in case of site-specific DNA-protein interactions [Fig 3D]. Intersegmental transfer occurs whenever two distal segments of the same dsDNA chain come close by over 3D space via ring closure events. Under such conditions the non-specifically bound DNA binding protein molecules can move from one segment of DNA to another segment of the same DNA without dissociation. Analogous to intersegmental transfer, internal displacement occurs whenever the overhang parts of the probe strand of cn-ssDNA make contact with some other region of the template strand of cn-ssDNA without dissociation of the previous nonspecific contact. In this way the probe strand of cn-ssDNA can perform an inchworm type movement over the template strand of cn-ssDNA without physical dissociation [Fig 3C]. One should note that the probe strand of cn-ssDNA may dissociate after searching n numbers of nonspecific sites on the template strand where n is a random variable. These 1D random search processes in combination with several rounds of nonspecific association and dissociation will result in the finding of the correct-contact on the template strand of cn-ssDNA by the probe strand that in turn results in the formation of cc-ssDNA.

Upon formation of cc-ssDNA, the zipping reaction needs to progress against the entropic barrier associated with the single strand overhangs of cc-ssDNA. Nucleation occurs whenever the number of correct contacts in cc-ssDNA exceeds certain critical value (N) against the entropic barrier imposed by the freely moving single stranded overhangs [27]. Since the time that is required to locate the correct-contact is a random variable, nucleation rate also will be a random variable. Nucleation will be followed by zipping of cc-ssDNA into a completely renatured dsDNA. Here one should note that both the nucleation and zipping are parts of a continuous process after the formation of cc-ssDNA i.e. there is no clear cut timescale separation or boundary between them. Zipping step will also be a stochastic process since at each step of zipping there is a finite nonzero probability of dissociation of cc-ssDNA. When the c-ssDNA strands are repetitive then the zipping reacting can progress in parallel. As the zipping reaction progresses towards completion, in case of nonrepetitive c-ssDNA strands the stability of cc-ssDNA gradually increases and as a result the probability of dissociation of cc-ssDNA decreases.

In **Scheme III** of Fig 2 [S] and [S'] are the concentrations (mol/lit, M) of the colliding c-ssDNAs whose lengths are L bps (template) and l bps (probe) respectively where by definition
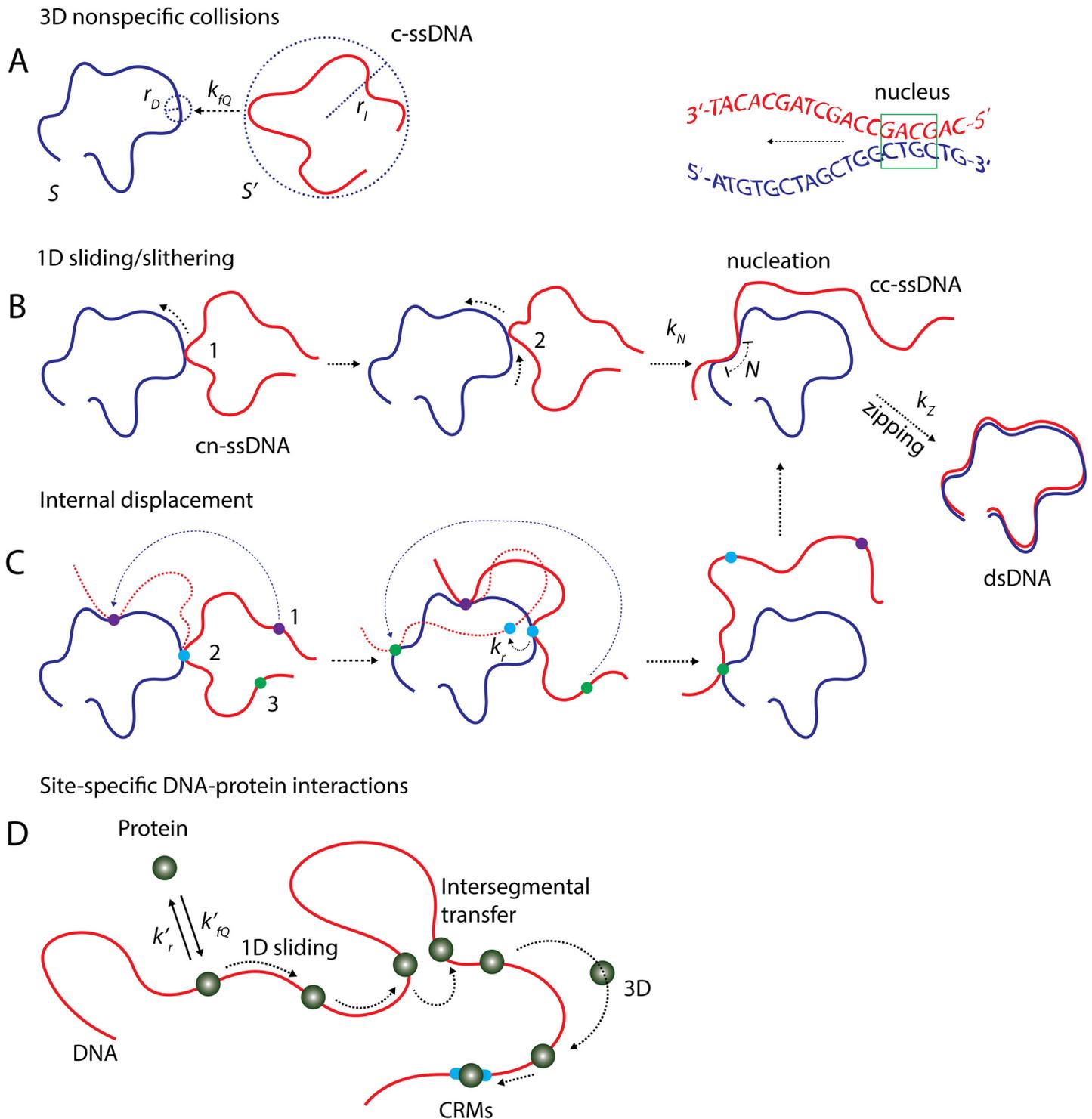
Fig 3. Mechanisms of DNA renaturation. A. Collision between c-ssDNA strands leads to the formation of nonspecific contacts (cn-ssDNA) at a diffusion controlled bimolecular collision rate of k_{fQ} (Q = C for condensed conformation and Q = R for relaxed conformational state of c-ssDNA). **B, C**. Slithering and internal displacement mechanism are involved in the nucleation step of renaturation of c-ssDNA strands. Here slithering is a 1D diffusion dynamics (with unit base step size) of one of the cn-ssDNA strands on the other in the process of searching for the correct-contact to initiate nucleation and zipping. Slithering involves local dynamics of individual bases of one strand of cn-ssDNA over the other. Internal displacement mechanism involves inchworm type movement of one of the cn-ssDNA strand over the other. Here two different segments of the same cn-ssDNA strand involved in the inchworm type 1D diffusion dynamics where second nonspecific contact is formed between cn-ssDNA strands before the dissociation the former one with a dissociation rate k_r. In the

illustration (**C**) three different locations of red colored strand of cn-ssDNA are marked as 1, 2 and 3. Initially position 2 of the probe c-ssDNA strand has a nonspecific contact with the template strand. In this condition the position 3 located on the freely moving overhang of probe strand makes contact with the template strand after which dissociation of the nonspecific contact at position 2 occurs. In this way the probe strand performs an inchworm type movement over the template strand. Occurrence of internal displacements in turn speeds up the 1D diffusion dynamics up to certain extent as in case of the intersegmental transfers via ring closure events associated with the site-specific DNA-protein interactions [31]. Both slithering and internal displacement mechanism are thermally driven stochastic processes which independently contribute to the 1D diffusion coefficient D_o. Correct contact formation leads to nucleation with rate k_N beyond the critical nucleus size of $N \sim$ 4–7 bases which in turn results in the zipping of cc-ssDNA strands into dsDNA with a rate of k_Z. **D**. Slithering seems to be analogous to the sliding mode of searching in the site-specific DNA-protein interactions whereas internal displacement is similar to that of the intersegmental transfer dynamics via ring closure events. Here two distal segments of the same DNA polymer come nearby in 3D space through thermally driven looping dynamics so that the nonspecifically bound protein molecule moves between them. As in DNA renaturation k'_{fQ} is the rate constant associated with the forward 3D diffusion mediated nonspecific binding of proteins with DNA and k'_r is the rate constant associated with the reverse dissociation step. Before reaching the CRMs (specific binding site) the protein molecules perform several rounds of 3D diffusion mediated association with DNA at random locations, 1D diffusion (which includes various modes of facilitating processes such as sliding, hopping and intersegmental transfers) along the DNA polymer and dissociations.

$L \geq l$, $[X]$ is the concentration of c-ssDNAs with nonspecific contacts (cn-ssDNA), $[Y_N]$ is the concentration of c-ssDNAs with correct-contact (cc-ssDNA) and nucleation, and $[H]$ is the concentration of completely renatured dsDNA form. Here one should note that $[Y_N]$ is a special form of $[H]$ since the complementary strands are already nucleated and aligned in $[Y_N]$. Since it is very difficult to identify and quantity $[Y_N]$ in experiments we consider $[H]$ as the main product of renaturation in this paper. Further k_{fQ} (M^{-1}s^{-1}) and k_r (s^{-1}) are the forward second-order and reverse first-order rate constants associated with the formation of nonspecific contacts between colliding c-ssDNAs, k_N (s^{-1}) is the nucleation rate constant and k_Z (s^{-1}) is the zippering rate constant. Since $[Y_N]$ is a hidden intermediate we consider the overall rate constant associated with both nucleation and subsequent spontaneous zipping as k_{NZ} which is the inverse of total time required for nucleation and zipping processes together i.e. $k_{NZ} = 1/(1/k_Z + 1/k_N)$. Typical values of the size of nuclei [6] seems to be $N \sim$ 4–7 bases. Here we set the subscript $Q = C$ for condensed conformational state of c-ssDNA and $Q = R$ for relaxed conformational state of c-ssDNA throughout this paper. The set of differential rate equations associated with **Scheme III** of Fig 2 can be written as follows.

$$d[X]/dt = k_{fQ}[S][S'] - (k_r + k_{NZ})[X]; \ d[H]/dt = k_{NZ}[X]; \ d[H]/dt \simeq \bar{k}_{HQ}[S][S'] \qquad (1)$$

Since nucleation step involves several rounds of 1D slithering and internal displacement dynamics of one of the cn-ssDNA strand over the other in combination with nonspecific association and dissociation events, k_{NZ} will be a function of n. Here n is the number of nonspecific sites scanned by cn-ssDNA strands on each other before dissociation. With this background one can define the overall second order rate constant as $\bar{k}_{HQ} = \int_0^L k_{HQ}p(n)dn$ where we have defined $k_{HQ} = k_{fQ}k_{NZ}/(k_{NZ} + k_r)$ and $p(n)$ is the probability density function associated with the random variable n. We will derive the explicit expressions for k_{NZ} and $p(n)$ in subsequent sections. Various parameters and variables defined throughout this paper are summarized in Table 1. While deriving the expression for $\bar{k}_{HQ}$ we have assumed that the nucleation and zipping are the rate limiting ones so that $d[x]/dt \simeq 0$ in the timescales of nucleation and zipping.

Essentially the first reaction in **Scheme III** of Fig 2 can be thought as collisions between the spatially distributed clusters of nitrogen bases corresponding to the two reacting c-ssDNAs as in case of a mean field approach [27]. As a result, the overall bimolecular rate associated with the formation of nonspecific contacts between spatially distributed base-clusters of c-ssDNAs is proportional of the product of concentrations of the total nitrogen bases in c-ssDNA molecules. The cylindrical surface area $C_M \sim 2\pi r_D M$ of a c-ssDNA molecule will be confined within

Table 1. Various symbols and their definitions

Symbol	Definition	Remarks
$[S]$, $[S']$	Concentration of complementary single strands of DNA (= [c-ssDNA])	mol/lit, M
$[X]$	Concentration of complementary strands with nonspecific contact between them (= [cn-ssDNA])	M
$[Y_N]$	Concentration of nucleated complementary single strands with nucleus size of N base pairs (= [cc-ssDNA])	M
$[H]$	Concentration of double stranded DNA (= [dsDNA]).	M
k_t	Smolochowski type 3D diffusion controlled collision rate limit[a].	$M^{-1}s^{-1}$
D_L, D_l	(= $k_BT/6\pi\eta r_M$) 3D diffusion coefficients associated with c-ssDNA polymers with length M bases ($M = L, l$).	$bases^2s^{-1}$
l_p, l_d	Step sizes associated with the 1D slithering (l_p = 1 base) and internal displacement dynamics ($l_d > l_p$)[b].	bases
D_{os}	(= $l_p^2 (w_f p_f + w_r p_r)$) 1D diffusion coefficient for the searching for the correct contact via slithering [c].	$bases^2s^{-1}$
D_{oi}	(= $l_d^2 (w_{f,d} p_{f,d} + w_{r,d} p_{r,d})$) 1D diffusion coefficient associated internal displacement[d].	$bases^2s^{-1}$
D_o	(= $D_{os} + D_{oi}$)	$bases^2s^{-1}$
r_D	Radius of DNA cylinder or radius of a nitrogen base bead. (1base = 3.4 x 10^{-10} m)	bases
L, l	Lengths of template and probe c-ssDNA strands, $L \geq l$.	bases
r_L, r_l	Radius of gyration of c-ssDNA polymer in aqueous solution whose lengths are L (template) and l (probe) [e].	bases
δ_C, δ_R	Multiplication factor corresponding to geometric constraints for the collision of c-ssDNA strands [f].	dimensionless
κ	Onsager radius*.	bases
ψ_Q	Multiplication factor for the electrostatic repulsion component associated with the collision of c-ssDNA.	dimensionless
χ_Q	Multiplication factor for the overall electrostatic component associated with the collision of c-ssDNA.	dimensionless
k_{fQ}	Bimolecular collision rate constant for the formation of nonspecific contacts (= $k_t \chi_Q \delta_Q$), ($Q = C, R$).	$M^{-1}s^{-1}$
k_r	Rate constant associated with the dissociation of cn-ssDNA.	s^{-1}
λ	(= L/n) minimum number of 1D slithering of cn-ssDNA of n bases required to find the correct contact.	dimensionless
N	Critical number of correct contacts in cc-ssDNA that can be considered as nucleus. ($N \sim$ 4–7 bases).	bases
τ_N	(= $\lambda n^2/12D_o$) overall time required for nucleation.	s
τ_Z	Zippering time to generate a completely renatured dsDNA from nucleation.	s
k_N	Rate constant associated with the formation of nucleus (= $1/\tau_N$, where τ_N is the nucleation time).	s^{-1}
k_Z	Macroscopic zippering rate constant (= $1/\tau_Z$, where τ_Z is the zippering time).	s^{-1}
k_{NZ}	(= $1/(1/k_Z + 1/k_N)$ = $1/(\tau_Z + \tau_N)$) coupled stochastic nucleation-zipping rate constant.	s^{-1}
C_L, C_l	Area of cylindrical surface of c-ssDNA ($C_L = 2\pi r_D^2 L$, $C_l = 2\pi r_D^2 l$) whose lengths are L and l bases.	$bases^2$
V_L, V_l	Area of spherical solvent shell that covers c-ssDNA polymer (= $4\pi r_M^2$).	$bases^2$
r_C, r_R	Reaction radius associated with the collision of one c-ssDNA polymer with another c-ssDNA[g].	bases
c	Sequence complexity of c-ssDNA which is the length of unique sequence found in a given stretch of DNA.	bases
ρ	(= L/c) number of repeats in c-ssDNA with length L bases and complexity of c bases.	dimensionless
k_+, k_-	Microscopic forward (k_+) and reverse (k_-) transition rate constants associated with the zipping process.	s^{-1}

(*Continued*)

Table 1. (*Continued*)

Symbol	Definition	Remarks
K_Z	($= k_-/k_+$) equilibrium constant associated with the microscopic zipping process.	dimensionless
$D_{\pm}$	($= l_p{}^2 k_+$) microscopic diffusion coefficient associated with the zipping process when $k_+ = k_-$ where $l_p = 1$ base.	bases^{2}s^{-1}
n	Number of possible nonspecific contacts scanned by cn-ssDNA before it dissociates into c-ssDNA.	dimensionless
u	Number of correct contacts at a given time point in the reaction associated with the zipping cc-ssDNA.	dimensionless
β	($= L/l_p$) maximum number of correct contacts in the completely renatured dsDNA.	dimensionless
k_p	($= k_+ l_p$) microscopic zipping rate constant.	bases s^{-1}
Y_A	Distance travelled by nonspecific contact via 1D slithering before cn-ssDNA dissociates.	bases
Y_E	($= k_p/k_r$) distance travelled in the zipping reaction before cn-ssDNA dissociates.	bases
k_{HQ}	Bimolecular rate constant associated with the complete renaturation [h].	M^{-1}s^{-1}
L_{opt}	optimum value of the length of c-ssDNA at which k_{HR} will be a maximum [i].	bases
$p(n)$	Probability density function (pdf) associated with the 1D slithering length (n).	pdf
k_{sm}	($= k_t \varepsilon$) Smolochowski type 3D collision rate limit associated with nitrogen bases A with T or G with C [j].	M^{-1}s^{-1}
k_2	Overall bimolecular renaturation rate in Wetmur-Davidson model [k].	M^{-1}s^{-1}

Notes:

*Onsager radius is defined as the distance between negatively charged phosphate backbones of colliding c-ssDNA chains at which the electrostatic repulsive energy is same as that of the thermal energy.

[a.] Here colliding molecules are of same in size with no charge on them (= $8k_B T/3\eta$ where k_B is the Boltzmann constant, T absolute temperature in degree K and η is the viscosity of solvent). For T = 298K and $\eta \sim 10^{-1}$ kgm^{-1}s^{-1} one obtains $k_t \sim 10^9$ M^{-1}s^{-1}.

[b.] Here l_p is the average bond length associated with c-ssDNA polymer and l_d is the average step size associated with the internal displacement mechanism. Here we consider the average step size since the step size associated with the internal displacement is a random variable.

[c.] Here w_f and w_r are microscopic forwards and reverse transition rates, p_f and p_r are the corresponding transition probabilities and $l_p = 1$ base is the step size.

[d.] Here $w_{f,d}$ and $w_{r,d}$ are microscopic forward and reverse transition rates, $p_{f,d}$ and $p_{r,d}$ are the corresponding transition probabilities and $l_d > l_p$ is the average step size.

[e.] For a linear chain polymer $r_M \simeq \sqrt{M l_p/6}$ where $M = L, l$.

[f.] ($\delta_R = L (1/r_L + 1/r_l)/8$, $\delta_C = (r_L + r_l)^2/4 r_L r_l$.). When $L = l$, then one finds that $\delta_R = L/4 r_L$, $\delta_C = 1$.

[g.] In case of condensed conformational state, $r_C = r_L + r_l$ and in case of relaxed conformational state, $r_R = r_D + r_l$).

[h.] ($= k_{fQ} k_{NZ}/(k_r + k_{NZ})$). Since k_{NZ} is a function of 1D slithering length (n), k_{HQ} also will be a function of n and subsequently we find that $k_{HQ} :\rightarrow k_{HQ}(n) \simeq k_{fQ}/(1 + nL/Y_A^2 + c/Y_E)$. Here Q = C, R depending in the conformational state of c-ssDNA strands. $\bar{k}_{HQ} :\rightarrow \int_0^L k_{HQ} p(n) dn$. Under relaxed conformational state of c-ssDNA strands when ($p Y_E/Y_A) \ll 1$ then as in [Eq 15](#) one obtains $\bar{k}_{HR} \simeq k_{sm}\sqrt{L l_p/c}$ where $k_{sm} = k_t \varepsilon$.

[i.] ($\simeq Y_A^2(c + Y_E)/n Y_E$) Under relaxed conformational state. Solution of $\partial_L k_{HR} = 0$ for L.

[j.] (Where $L = c = 1$). Here we have defined $\varepsilon \simeq \chi_R(k_+/k_r)\sqrt{3/8}$. Upon comparison with the experimental data [6] one finds that $\varepsilon \sim 10^{-3}$.

[k.] Detailed fitting over the experimental data suggested [6] the empirical form as $k_2 = 3.5 \times 10^5 \sqrt{L/c}$. We denote this by k_{HR} in our model.

doi:10.1371/journal.pone.0153172.t001

the spherical solvent shell with surface area $V_M \simeq 4\pi r_M^2$ ($M = L$ for template and $M = l$ for probe) where r_Q is the radius of gyration of the respective c-ssDNA molecule. Under strongly condensed conformational state of c-ssDNA one finds that $V_M < C_M$ and when the DNA polymer is in a relaxed conformational state then one find that $V_M > C_M$. At a coarse grained level one can model the bases of c-ssDNA as a chain of spherical beads with radius r_D. Under relaxed conformational state all these nitrogen base beads are distributed on the surface of the spherical solvent shell that covers the entire c-ssDNA molecule [Fig 2B]. Whereas under condensed conformational state of template c-ssDNA significant fraction of nitrogen base beads will be inaccessible to the inflowing probe c-ssDNA molecules since they are buried inside the matrix of condensed c-ssDNA polymer [Fig 2C].

Calculation of the Nonspecific-Contact Formation Rate

Actually most of the theoretical models derived the scaling over the length of c-ssDNAs mainly from the fact that the radius of gyration (measured in bps) associated with reacting c-ssDNAs scales with their length as $r_M \propto M^\alpha$ ($M = L$ and l for the template and probe strands respectively). Here the value of the exponent $0 < \alpha < 1$ varies depending on the type of polymer and solvent conditions. For an ideal Gaussian chain polymer that is immersed in a theta solvent one finds that $\alpha \sim \frac{1}{2}$. For a linear chain polymer one finds that $r_M = \sqrt{Ml_p/6}$ ($M = L, l$) where l_p is the average bond length [32, 33]. For c-ssDNA we have $l_p \sim 1$ base between nitrogen base beads. Noting these facts the Smolochowski type limiting rate for a diffusion controlled nonspecific contact forming step can be given as follows.

$$k_{fQ} = B_Q J_Q; \ J_Q = (D_l + D_L)/r_Q; \ D_M = k_B T/6\pi\eta r_M; \ Q = \{C, R\}; \ M = \{L, l\} \qquad (2)$$

Here J_Q is the inflowing flux of c-ssDNA molecules, B_Q is the exposed total surface area on which the influx of c-ssDNA molecules J_Q is acting, D_S is the three-dimensional (3D) diffusion coefficient associated with the colliding c-ssDNAs and r_Q is the reaction radius that depends on the conformational state of c-ssDNA. When c-ssDNA strands are relaxed, then all the base beads will be distributed over the surface of the solvent shell that covers the entire c-ssDNA polymer. Since $V_M > C_M$ for $M = \{L, l\}$ there will be several patches on the solvent shell surface without base beads. Noting that a nonspecific contact can be formed only upon collision between probe and base beads of template, one needs to integrate over the surface of the template c-ssDNA polymer that is spread over the surface of spherical solvent shell rather than the entire surface of spherical solvent shell [Fig 2B]. In this situation one finds that $B_R \simeq 2\pi r_R L$ where $r_R = r_D + r_l$ is the reaction radius associated with the collision of the probe c-ssDNA strand on a nitrogen base bead of template c-ssDNA strand. On the other hand, when the c-ssDNA polymer is highly condensed then significant fraction of the nitrogen base beads will be buried inside the matrix of the DNA condensate [Fig 2C].

Unlike the relaxed conformational state, in case of condensed conformational state the surface of the spherical solvent shell that covers entire c-ssDNA strands will be filled with nitrogen base beads. This means that collision between base bead clusters of c-ssDNA strands always result in the formation of nonspecific contacts under condensed conformational state. In this condition one finds that $B_C \simeq 4\pi r_C^2$. Here $r_C = r_L + r_l$ is the reaction radius associated with the collision between the condensed nitrogen base bead clusters of template and probe c-ssDNA strands. Therefore depending on the conformational state of the colliding c-ssDNA molecules the bimolecular collision rate associated with the formation of nonspecific contact between the template and probe molecules can be written as follows.

Case I: Relaxed conformational state of c-ssDNA

$$k_{fR} \simeq 2\pi r_R L J_R; \ J_R \simeq (D_l + D_L)/r_R; \ r_R = r_D + r_l; \ \therefore k_{fR} \simeq k_t L(1/r_l + 1/r_L)/8 \qquad (3)$$

Case II: Condensed conformational state of c-ssDNA

$$k_{fC} \simeq 4\pi r_C^2 J_C; \ J_C \simeq (D_l + D_L)/r_C; \ r_C = r_L + r_l; \ \therefore k_{fC} \simeq k_t (r_L + r_l)^2/4 r_L r_l \qquad (4)$$

Here $k_t \simeq (8 k_B T/3\eta)$ is the Smolochowski type 3D diffusion controlled collision rate limit ($M^{-1}s^{-1}$) associated with the bimolecular collisions between the c-ssDNA molecules. Here k_B is the Boltzmann constant, η is the viscosity of the reaction medium and T is the absolute temperature in degrees K. Further the colliding molecules are assumed to be same in size with no charge on them. In general $k_{fQ} \simeq k_t \delta_Q$ where $\delta_R \simeq L(1/r_l + 1/r_L)/8$ and $\delta_c \simeq (r_L + r_l)^2/4 r_L r_l$. When the colliding c-ssDNA strands are same in size then $\delta_R \simeq L/4 r_L$ and $\delta_c \simeq 1$. Noting that [32–33] for a linear chain polymer $r_L = \sqrt{L l_p/6}$ one obtains $\delta_R \simeq \sqrt{3L/8 l_p}$.

Role of Electrostatic Repulsions at the DNA-DNA Interface

While deriving Eqs 3 and 4 we have not considered the electrostatic repulsions between the negatively charged phosphate backbones of c-ssDNA chains and shielding effects of solvent and other ion molecules present at the DNA-DNA interface of c-ssDNA molecules. Upon considering this fact and following the detailed works of Montroll in Ref. [34] we find the expression for the modified bimolecular rate constants as follows.

$$k_{fQ} = k_t \psi_Q \delta_Q; \ \psi_Q \simeq (\kappa/r_Q)/(e^{\kappa/r_Q} - 1); \ \kappa = \zeta_S \zeta_{S'} e^2/\mu k_B T; \ Q = (C, R) \qquad (5)$$

Here κ is the Onsager radius which is defined as the distance between negatively charged phosphate backbones of colliding c-ssDNA chains at which the electrostatic repulsive energy is same as that of the thermal energy ($\sim k_B T$), $\zeta_S e$ and $\zeta_{S'} e$ are the overall charges on the respective c-ssDNA molecules. Since $|\kappa| \geq r_Q$ by definition and $\kappa > 0$ in the present context, one finds that $\psi_Q \sim |\kappa|/r_Q$ for large values of $|\kappa|$ and for $|\kappa| = r_Q$ one obtains $\psi_Q \simeq 0.58$. For a typical value of $|\kappa| \simeq 10 r_Q$ we find that $\psi_Q \simeq 10^{-4}$. When $\kappa < 0$ as in case of site-specific DNA-protein interactions we obtain $\lim_{|\kappa| \to r_Q} \psi_Q \simeq 1.58$. One should also note that $\psi_Q \simeq 1$ only when we have $|\kappa| = 0$. While deriving Eq 5 we have not considered the shielding effects of solvent ions present at the DNA-DNA interface over the electrostatic repulsive forces between the phosphate backbones of c-ssDNA strands. Upon following the Debye theory of kinetic salt effects over diffusion controlled collision rate processes [35], one can rewrite the modified bimolecular rate constant in the presence of other ions in the solvent as follows.

$$k_{fQ} \simeq k_t \chi_Q \delta_Q; \ \chi_Q = \psi_Q \exp(2F\zeta_S \zeta_{S'} \sqrt{\Xi}); \ F \simeq 0.509 \qquad (6)$$

Here Ξ is the overall ionic strength of the aqueous medium and $Q = (C, R)$ as defined in Eqs 3–5 depending on the type of conformational state of colliding c-ssDNA polymers.

Calculation of the Nucleation Time and Nucleation Rate

We learnt from recent computational studies [29] that the colliding c-ssDNA with nonspecific contacts between them undergo several trials of slithering and internal displacement dynamics before reaching the correct-contact and then nucleate the zipping process. These dynamical processes are similar to that of the facilitating 1D diffusional dynamics as in case of site specific DNA-protein interactions. Unlike the overall electrostatic attractive forces acting at the DNA-protein interface, in case of DNA renaturation there is a strong electrostatic repulsive force

acting at the interface of cn-ssDNA that will be shielded by the solvent molecules present at the interface of cn-ssDNA strands. With this background one can model the slithering dynamics as 1D diffusion of one c-ssDNA molecule on the other in the process of searching for the correct-contact. To find the correct-contact on one c-ssDNA the other c-ssDNA needs to try out at least $\lambda = L/n$ stretches of 1D slithering with an average size of n bases. This will ensure that the initial nonspecific contact visits all the possible positions and subsequently the correct contact is formed. The mean first passage time (τ_N) associated with the visit of all the possible positions of c-ssDNAs by the initial nonspecific-contact between them via 1D diffusion can be given as $\tau_N = \lambda \bar{\tau}_c$ where $\bar{\tau}_c \simeq n^2/12D_o$ is the average time that is required by an unbiased 1D random walker to visit n sites of a linear lattice [36–38] confined inside the interval $x \in (0, n)$ starting from anywhere inside the interval as shown in **Appendix A**.

Noting that the nucleation rate is given as $k_N = 1/\tau_N$ one can obtain $k_N \simeq 12D_o/nL$ which means that $k_N \propto 1/L$. This is a reasonable one since the probability of finding the correct contact upon each 3D diffusion mediated nonspecific collisions between c-ssDNA strands is $p_{cc} = 1/L$. One should also note that the probability of finding the correct contact upon each nonspecific collisions will be independent on the repetitiveness of c-ssDNA strands. Here D_o is the one dimensional diffusion coefficient associated with the dynamics of the probe c-ssDNA strand over the template c-ssDNA strand. One should note that D_o includes the contributions from both slithering and internal displacement dynamics which occurs within a cn-ssDNA molecule. Actually one can write down the expression for the 1D diffusion coefficient as follows.

$$D_o = D_{os} + D_{oi}; \; D_{os} \simeq l_p^2(p_f w_f + p_r w_r); \; D_{oi} \simeq l_d^2(p_{f,d} w_{f,d} + p_{r,d} w_{r,d}) \tag{7}$$

Here D_{os} is the diffusion coefficient corresponds to slithering dynamics and D_{oi} corresponds to the internal displacement dynamics where $p_{f/r}$ are the transition probabilities associated with the forward and reverse steps of the random walker with step size $l_p = 1$ base and $w_{f/r}$ are the corresponding transition rates. Similarly in case of internal displacement dynamics l_d is the average step size and $p_{f/r,d}$ are the respective forward and reverse transition probabilities and $w_{f/r,d}$ are the corresponding transition rates. Here we consider the average step size since the step size associated with the inchworm type movements in the internal displacement mechanism is a random variable. For an unbiased searching via both 1D slithering and internal displacements we have $p_{f/r} = p_{f/r,d} = 1/2$.

Although $l_d > l_p$, internal displacement dynamics may not be able to contribute much to the overall searching dynamics under relaxed conformational state of the colliding c-ssDNA strands since it requires looping and segmental motion of c-ssDNA chains. Therefore in general we have $w_{f/r} > w_{f/r,d}$. Further one should note that slithering dynamics is a purely a local one i.e. slithering involves the local movement of bases. This means that the transition rates $w_{f/r}$ and diffusion coefficient D_{os} will be independent of the length of c-ssDNA. Whereas internal displacement involves the movement of bulky segments [29] of the colliding c-ssDNA strands and therefore the corresponding transition rates $w_{f/r,d}$ and diffusion coefficient D_{oi} will be dependent on the size of c-ssDNA strands. Here we can ignore the contribution of D_{oi} to the overall diffusion coefficient D_o only for the relaxed conformational state of c-ssDNA and we cannot ignore it for the condensed conformational state of c-ssDNA since the contribution from internal displacement mechanism will be the dominating one under such conditions.

Calculation of the Zipping Time

Formation of correct-contact will result in the nucleation of zipping process. Upon formation of a nucleation site, the subsequent stochastic zipping of cc-ssDNA can be described by the

following birth-death master equation.

$$\partial_t P(u,t) = k_+ P(u-1,t) + k_- P(u+1,t) - (k_+ + k_-)P(u,t) \qquad (8)$$

Here $P(u,t) = P(u,t|u_0,t_0)$ is the probability of finding the cc-ssDNA with u numbers of correct contacts at time t starting from the nucleation at $t = t_0$ with $u = u_0$ numbers of correct contacts, k_+ (s^{-1}) and k_- (s^{-1}) are the respective average forward and reverse rate constants associated with the microscopic zipping reaction. Here the initial and boundary conditions corresponding to Eq 8 can be written as follows.

$$P(u,t_0) = P(u,t_0|u_0,t_0) = \delta(u-u_0); \quad k_- P(1,t) = k_+ P(0,t); \quad P(\beta+1,t) = 0 \qquad (9)$$

Here $u = 1$ is a reflecting boundary and $u = \beta$ is the absorbing boundary. One can solve the difference equation Eq 8 as follows. By defining equilibrium constant as $K_Z = (k_-/k_+)$, one can find the following expression for the overall mean first passage time associated with complete zipping of β correct contacts ($u = \beta$) of cc-ssDNA starting from the number of correct contacts $u = 1$ as shown in **Appendix B**.

$$\tau_Z = \sum_{u=1}^{\beta} \phi(u) \sum_{w=1}^{u} \left(k_+ \phi(w)\right)^{-1} = \left(K_Z^{\beta+1} - K_Z(\beta+1) + \beta\right)/k_+(1-K_Z)^2 \qquad (10)$$

From this equation we find the limits $\lim_{K_Z \to 1} \tau_Z \simeq \beta(1+\beta)/2k_+$ and $\lim_{K_Z \to 0} \tau_Z \simeq \beta/k_+$ where $\beta = L/l_p$ is a dimensionless quantity which is the total number of correct-contacts between colliding c-ssDNA upon formation of dsDNA. Here L is total length of c-ssDNA in bases and $l_p = 1$ base (1 base $\sim 3.4 \times 10^{-10}$ m). When the forward rate constant associated with formation of dsDNA is much higher than the reverse rate constant then we find the expression for the zipping rate constant as that $k_Z = 1/\tau_Z \simeq k_p/L$ where we have defined $k_p = k_+ l_p$ (bases/s). On the other hand when $k_+ = k_-$ then the zipping will be a pure 1D diffusion process with the phenomenological diffusion coefficient as $D_{\pm} \simeq l_p^2 k_+$ (base^2/s) and subsequently $\tau_Z \simeq L^2/2D_{\pm}$. When the reacting c-ssDNA is repetitive with a sequence complexity of c bases (here we have $c \in (1, L)$), then the observed zipping rate will be proportional to the number of repeats in that template c-ssDNA strand ($\rho = L/c$) and one obtains $\lim_{K_Z \to 0} \tau_Z \simeq c/k_p$ i.e. the total zipping time will be directly proportional to the complexity of the reacting c-ssDNA molecules which is in line with the experimental observations [6–7]. Using these results one can write down the expression for the total time that is required for the overall nucleation and zipping processes (τ_{NZ}) to generate dsDNA as follows.

$$\tau_{NZ} = \tau_N + \tau_Z; \quad \lim_{K_Z \to 0} \tau_Z \simeq c/k_p; \quad \lim_{K_Z \to 1} \tau_Z \simeq Lc/2D_{\pm}; \quad \tau_N = \lambda n^2/12D_o; \quad \lambda = L/n \qquad (11)$$

Here one should note that the zipping process can be thought as a random walk over a linear lattice [36–38] with initial and boundary conditions given in Eq 9. The phenomenological diffusion coefficient associated with such random walk is defined as $D_{\pm} \simeq l_p^2(p_+ k_+ + p_- k_-)$ where $p_{\pm}$ are the transition probabilities associated with the forward and reverse steps in the zipping process. When the zipping is unbiased over forward or reverse steps with $k_+ = k_-$ then $p_{\pm} = 1/2$ and we recover the expression for $D_{\pm} \simeq l_p^2 k_+$ in the limit as $K_Z = 1$.

Calculation of the Overall Renaturation Rate

Using the expressions for the nucleation and zipping times one can define the overall bimolecular collision rate associated with the complete formation of dsDNA from c-ssDNA in **Scheme III**

of Fig 2 as follows.

$$\bar{k}_{HQ} = \int_0^L k_{HQ} p(n) dn; \quad k_{HQ} = k_{fQ}/(1 + k_r \tau_{NZ}) \simeq k_{fQ}/(1 + nL/Y_A^2 + c/Y_E) \tag{12}$$

Here we have defined two important characteristic lengths $Y_A = \sqrt{12D_o/k_r}$ and $Y_E = k_p/k_r$. The length Y_A describes the distance that is travelled by the initial nonspecific contact before cn-ssDNA dissociates whereas Y_E describes the distance travelled in the zipping reaction before cn-ssDNA dissociates into corresponding c-ssDNA molecules. Generally one observes that $Y_A > Y_E$. The probability density function connected with the 1D slithering lengths n or its weighting function $p(n)$ can be calculated as follows. When the residence times (τ) associated with the dissociation of cn-ssDNA molecules is distributed as an exponential then one finds that $p(\tau) \propto e^{-k_r \tau}$ and subsequently one obtains $p(n) \propto n e^{-(n/Y_A)^2}$ which mainly originates from the fact that within the residence time τ, the distance travelled by the nonspecific contact through 1D diffusion dynamics can be anywhere in the interval $n \in (1, L)$ so that we obtain the transformation rule as $\tau = n^2/12D_o$. With this definition of the residence time of c-ssDNA strands in the cn-ssDNA configuration one can write down the expression for the distribution of slithering lengths n as follows [Fig 4A].

$$p(n) = 2n e^{-(n/Y_A)^2}/Y_A^2 (1 - e^{-(L/Y_A)^2}) \simeq 2n e^{-(n/Y_A)^2}/Y_A^2 \tag{13}$$

Using the expression of $p(n)$ in Eq 13 and expanding k_{HQ} in a Maclaurin series one can obtain the following expression for the overall renaturation rate constant.

$$\bar{k}_{HQ} \simeq (k_{fQ}/(1 + c/Y_E)) \sum_{m=0}^{\infty} \Gamma(m/2 + 1)(-L/Y_A(1 + c/Y_E))^m \tag{14}$$

While deriving this equation without losing generality we have extended the limits of n in the integration towards infinity since both $p(n)$ and k_{HQ} approach zero at this limit. Under certain conditions one can obtain the following approximation for the series in Eq 14.

$$\bar{k}_{HQ} \simeq (k_{fQ} Y_E/c) \sum_{m=0}^{\infty} \Gamma(m/2 + 1)(-\rho Y_E/Y_A)^m; \quad Y_A \gg Y_E; \quad (c/Y_E) \gg 1 \tag{15}$$

Later we will show that these inequality conditions are indeed valid. When the sequence complexity of the template c-ssDNA molecule is high enough and $(p Y_E/Y_A) \ll 1$ then one can write down the leading zeroth order approximation ($m = 0$ in Eqs 14 and 15) of the overall second order rate constant associated with the renaturation of relaxed c-ssDNA chains with equal lengths $L = l$ as follows.

$$\bar{k}_{HR} \simeq k_{fR} Y_E/c = k_{fR} k_p/k_r c = k_t \chi_R k_p L/4 r_L k_r c \propto \sqrt{L}/c\eta; \quad \because r_L \propto \sqrt{L}; \quad k_t \propto 1/\eta \tag{16}$$

Upon substituting the expression $r_L \simeq \sqrt{L l_p/6}$ in Eq 16 one obtains that $\bar{k}_{HR} \simeq k_{sm} \sqrt{L l_p}/c$ in line with the expression for the overall renaturation rate in Wetmur-Davidson model [6]. Here $k_{sm} \simeq k_t \varepsilon$ is the Smolochowski type 3D diffusion controlled collision rate limit corresponding to a situation $L = c = 1$ where we have defined $\varepsilon \simeq \chi_R (k_+/k_r)\sqrt{3/8}$. When the conditions given in Eq 15 are true then Eq 16 suggests that the overall bimolecular collision rate associated with the renaturation reaction is directly proportional to the square-root of the length of c-ssDNA and inversely proportional to both sequence complexity of the reacting c-ssDNA molecules and viscosity of the reaction medium in line with the experimental observations. Experiments suggest that the scaling of renaturation rate with the length of c-ssDNA molecules that is given in Eq 16 is valid [6] only in the range of $L \sim 10^2 - 10^4$ bases. In this context our model suggest that the square root scaling of the renaturation rate on length of reacting
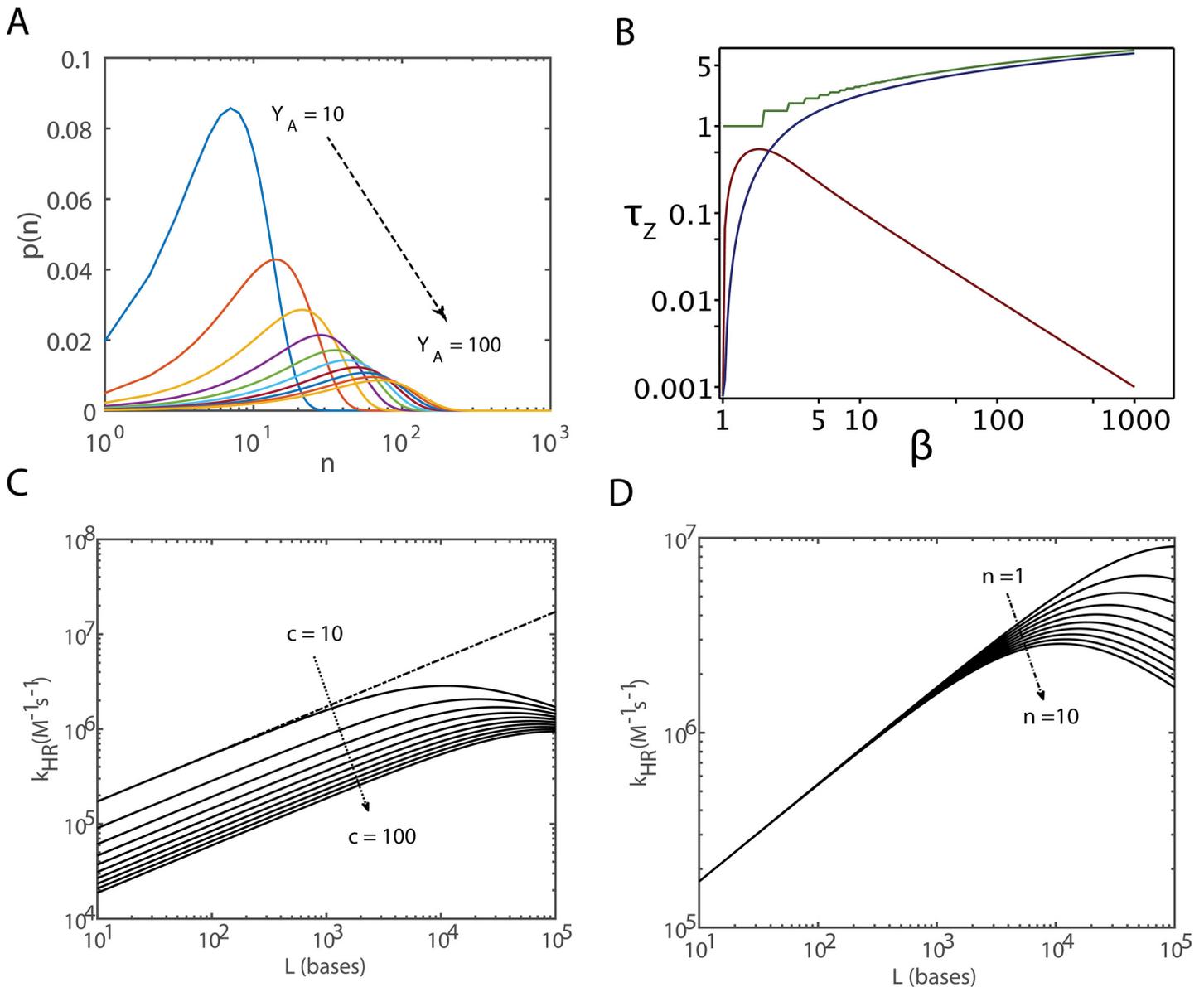
Fig 4. Cooperative effects on DNA renaturation. A. Probability density function associated with the one dimensional slithering length (*n* measured in bases) of cn-ssDNA in the process of searching for the correct-contact as given in Eq 13 for different values of the characteristic length $Y_A = \sqrt{12 D_o / k_r}$ ranging from 10 to 100 bases where D_o (base²s⁻¹) is the one dimensional diffusion coefficient associated with the slithering dynamics and k_r is the dissociation rate constant connected with cn-ssDNA. **B**. Zipping time (τ_Z, measured in seconds) in the presence of cooperative effects. Here sequence complexity (*c*) is same as that of the length (*L*) of c-ssDNA i.e. *c = L*. The number of correct-contacts $\beta = L/l_p$ is a dimensionless quantity where l_p = 1 base and *L* is the length of the reacting c-ssDNA. Green solid line is calculation from Eq 18 and blue solid line is calculation from Eq C9 of **Appendix C**. Here we have set $K_Z \sim 10^{-6}$ and $k_+ \sim 1$ s⁻¹. Red solid line is the derivative of zipping time with respect to β as in Eq 21 which shows that the value of the derivative of overall zipping time with respect to β is < 10⁻² when β > 10². These plots suggest that when K_Z tends towards zero, the overall zipping time of a non-repetitive and long c-ssDNA will be independent of the length of the reacting c-ssDNA molecules. Zipping time of a repetitive c-ssDNA with a sequence complexity of *c* bases increases linearly with *c*. **C**. Variation of the overall renaturation rate k_{HR} with respect to length and complexity of c-ssDNA under relaxed conformational state. Here settings are $k_{fR} \simeq 6 \times 10^5 \sqrt{L}$ M⁻¹s⁻¹, Y_A = 100 bases, Y_E = 1 bases and *n* = 10 bases. **D**. Variation of the overall renaturation rate k_{HR} as in Eq 12 with respect to the length of reacting c-ssDNA strands *L* and 1D slithering distance *n* under relaxed conformational state. Here settings are $k_{fR} \simeq 6 \times 10^5 \sqrt{L}$ M⁻¹s⁻¹, Y_A = 100 bases, Y_E = 1 bases and *c* = 10 bases. In both **C** and **D**, k_{HR} shows a maximum at $L = L_{opt}$. Here Lopt can be obtained by solving $\partial_L k_{HR} = 0$ for *L*. Explicitly one finds that $L_{opt} \simeq Y_A^2 (c + Y_E)/n Y_E$. The dotted line in (**C**) is $k_{HR} \simeq 6 \times 10^5 \sqrt{L}/(1 + c/Y_E)$ which breaks down beyond L_{opt}.

doi:10.1371/journal.pone.0153172.g004

c-ssDNA strands will be valid only when the inequalities given in Eq 15 are true apart from the condition that $(pY_E/Y_A) \ll 1$ which may break down beyond certain values of the copy number (ρ) in the repetitive c-ssDNA.

Role of Cooperativity in Renaturation Kinetics

Scaling results given by Eq 16 works only for highly repetitive and relaxed conformational state of c-ssDNA ($L > c$) and it will break down at $L = c$ since at this point the scaling becomes as $\bar{k}_{HR} \propto 1/\sqrt{L}$ which is not true. The main reason for this observation is that while calculating the zipping rate constant we have not considered the underlying cooperative effects over the zipping process. When c-ssDNA is highly repetitive then the zipping process can take place in parallel for all the $\rho = (L/c)$ number of short repetitive motifs. Under such conditions the cooperative effects will not be noticeable since the enhancement of renaturation process by the parallel-zipping will dominate over the underlying cooperative effects. This means that Eq 16 will be true only for repetitive c-ssDNA.

When the reacting c-ssDNA molecules are non-repetitive and long enough then the probability of formation of an additional correct-contact in the cc-ssDNA molecule that is undergoing zipping reaction will be directly proportional to the existing number of correct-contacts (u) and the probability of breakdown of an existing correct-contact will be directly proportional to the number of overhanging single stranded stretches of cc-ssDNA (β-u). This is true since the existing correct-contacts always stabilize newly formed correct-contacts and overhang single stranded regions of cc-ssDNA always try to destabilize the newly formed correct-contacts. Here one should note that we are dealing with the cooperative effects at a mesoscopic level within an independent and single renaturing cc-ssDNA molecule rather than at macroscopic level where the descriptive parameter of the renaturation process will be the number of mol-bases in c-ssDNA (n_{SS}) or dsDNA (n_{DS}) form rather than the number of correct-contacts in cc-ssDNA (u) as in the current context.

At macroscopic level the rate of change in the number of mol-bases in ssDNA form in the process of zipping will be directly proportional to the number of mol-bases in ssDNA form as well as number of mol-bases in dsDNA form which results in a cooperative sigmoidal type time evolution of the renaturation process [16–17, 28] where the macroscopic kinetic rate equation will be written as $dn_{DS}/dt \propto n_{cc}(n_0 - n_{DS})n_{DS}$. Here n_0 is the initial concentration of mol-bases of c-ssDNA molecules in the system and n_{cc} is the total number of cc-ssDNA molecules in the system. With this background the birth-death master Eq 8 can be rewritten to include the cooperative effects for renaturation of a single cc-ssDNA molecule as follows.

$$\partial_t P(u,t) = k_+(u-1)P(u-1,t) + k_-(\beta-u-1)P(u+1,t) - (k_+u + k_-(\beta-u))P(u,t) \quad (17)$$

Upon solving the backward type equation corresponding to this differential difference equation for appropriate boundary conditions as shown in Eq 9 and **Appendix C** one obtains the following expression for the overall zipping time that is required for the formation of $u = \beta$ numbers of correct-contacts starting from $u = 1$ in the presence of cooperative effects.

$$\tau_Z = \sum_{u=1}^{\beta} K_Z^u(\xi_2 F_1([1,1],[2-\beta],-K_Z^{-1}) + \phi_2 F_1([1,u+1],[u+2-\beta],-K_Z^{-1})) \quad (18)$$

Here $_2F_1$ is the hypergeometric function and we have defined various parameters as follows.

$$\xi = (-1)^u \Gamma(u+1-\beta)/k_-\Gamma(u+1)\Gamma(1-\beta)(\beta-1);$$
$$\phi = \Gamma(u+1-\beta)/K_Z^{u+1}k_+\Gamma(u+2-\beta) \quad (19)$$

The hypergeometric function of type $_2F_1$ is defined as follows.

$$_2F_1([a,b],g,w) = \sum_{m=0}^{\infty} w^m (a)_m (b)_m / m! (g)_m; \quad (h)_q = \Gamma(h+q)/\Gamma(h) \tag{20}$$

One can simplify the complicated expression given by Eq 18 using Fokker-Planck equation especially for large values of β as shown in **Appendix C**. Upon defining $\lim_{K_Z \to 0} \tau_Z = \tilde{\tau}_Z$ one can derive the following approximate expression for the overall zipping time.

$$\tilde{\tau}_Z = (2/k_+) \int_1^{\beta} e^{-2y}(\mathrm{Ei}(1,-2) - \mathrm{Ei}(1,-2y))dy; \quad \mathrm{Ei}(a,z) = \int_1^{\infty} e^{-mz} m^{-a} dm \tag{21}$$

It seems from Eq 21 along with other computational analysis that the dependency of overall zipping time decreases with increasing β in the presence of cooperative effects which can be demonstrated by the following limiting conditions.

$$\partial \tilde{\tau}_Z / \partial \beta \simeq 2e^{-2\beta}(\mathrm{Ei}(1,-2) - \mathrm{Ei}(1,-2\beta))/k_+; \quad \lim_{\beta \to \infty} \partial \tilde{\tau}_Z / \partial \beta = 0 \tag{22}$$

From Eqs 17–22 one finds that when K_Z tends towards zero then in the presence of strong cooperative effects the zipping time of a non-repetitive cc-ssDNA will be independent of the length of the reacting c-ssDNA molecules especially for large values of L as shown in Fig 4B. Based on these observations we recover the observed scaling of the overall bimolecular rate constant on the length of the colliding c-ssDNA molecules as $\bar{k}_{HR} \propto \sqrt{L}/\eta$ for the renaturation of the non-repetitive c-ssDNA strands for which $L = c$ in Eqs 15 and 16 since Y_E will be independent of sequence length and the number of copies will be $\rho = 1$.

Discussion

Two-Step DNA Renaturation Model of Wetmur and Davidson

Understanding the mechanism of renaturation of c-ssDNA is one of the central topics in molecular biology and biological physics. Wetmur and Davidson [6] developed their model by assuming that the renaturation rate is directly proportional to the total phosphate concentration which is in turn directly proportional to the total number of mol-bases in ssDNA or dsDNA form. According to their model the overall second order rate (k_2) associated with the renaturation of repetitive c-ssDNA can be written as $k_2 \propto k_N p$ where $\rho = L/c$ is the copy number of repetitive motifs in the entire c-ssDNA polymer. Here L is the length of c-ssDNA and c is the sequence complexity and the nucleation rate was assumed to scale with L as $L^{-\alpha}$ where $\alpha = 1/2$ due to the excluded volume effects associated with the interpenetration of c-ssDNA molecules that is essential for the nucleation reaction. As a result one obtains the scaling as $k_2 \propto \sqrt{L}/c$ where the proportionality constant was assumed to be the Smolochowski type bimolecular collision rate limit (k_{sm}) i.e. $k_2 = k_{sm}\sqrt{Ll_p}/c$.

In Wetmur-Davidson model it was assumed that $k_{sm} = k_t$. Here one can identify that $k_2 = \bar{k}_{HR}$ and $k_{sm} = (k_t \varepsilon)$ of our model Eq 16 particularly for a relaxed conformational state of c-ssDNA that also includes the contributions from the electrostatic repulsions at the interface of colliding c-ssDNA molecules. Since in this model the nucleation is combined with nonspecific-contact formation step one finds that the nucleation rate is inversely proportional to the square root of the length of c-ssDNA strands. The main arguments for this scaling result put forth by Wetmur and Davidson were viz. (**a**) the radius of gyration of c-ssDNA is directly proportional to its length and (**b**) the reaction radius associated with the collision between c-ssDNA molecules is independent of the radius of gyration of c-ssDNA chains since both these strands can interpenetrate freely upon their collision. Though the assumption (**a**) is a right one for Gaussian chain polymers there

are several questions associated with the assumption (**b**) since the reaction radius always depends on the sum of the radii of gyration of reactant molecules. In this context our detailed model clarifies the origin of such scaling in the renaturation phenomenon. It is clear from our theory that the rate constant associated with the formation of initial nonspecific-contact is directly proportional to the square-root of the length of the reacting c-ssDNA molecules.

One should note that it is very difficult to identify and isolate nucleated cc-ssDNA molecules since they are indistinguishable from the zipping cc-ssDNA molecules. Therefore it is more appropriate to combine the nucleation step with the zipping step rather than with the nonspecific-contact formation step as in case of Wetmur-Davidson model. Other issues in their model are such as the breakdown of scaling at $L = c$ arises because the underlying cooperative effects in the long and non-repetitive c-ssDNA were not considered in their model as in case of our [Eq 16](). Further upon extrapolating towards the limit $L = 1$ base (so that the sequence complexity becomes as $c = 1$ base) Wetmur-Davidson model predicted that $k_2 = k_t$. However experimental observations suggested that the extrapolated bimolecular collision rate constant associated with the renaturation reaction at $L = c = 1$ was ~10^3 times lower than the Smolochowski type diffusion controlled bimolecular collision rate limit (k_t). On this basis they in turn discarded the possibility of diffusion control in the kinetics of renaturation of c-ssDNA molecules.

In this context [Eq 16]() of our model suggests an approximate expression for the extrapolation intercept as $k_{sm} \simeq k_t \varepsilon$ from which we can deduce that $\varepsilon \simeq \chi_R (k_+/k_r) \sqrt{3/8} \simeq 10^{-3}$. One should note that **scheme I** of [Fig 1]() is still valid with zero nucleation and zipping times and the steric factor ε mainly accounts for the geometric constraints associated with the bond formation between nitrogen bases A-T or G-C of the colliding single nucleotides in the limit $L = 1$ and $c = 1$ apart from the electrostatic repulsions due to the negatively charged phosphate groups. Here one should note that in our model the square-root dependency of renaturation rate on the length of c-ssDNA molecules mainly originates from the fact that the radius of gyration of c-ssDNA molecules scales with length as $r_L \propto L^\alpha$ where $\alpha = \frac{1}{2}$ which is valid only for a Gaussian type polymers in a theta solvent. It seems that the error introduced by this assumption in the exponent is within the experimental error range [6].

Temperature Dependency of DNA Renaturation Rate

Although the dissociation rate (here it is k_r) constant increases exponentially with temperature [39–42] there are several controversies exist on the dependency of renaturation rate constant on increasing temperature. Some experimental studies established [39–40] a decrease in the renaturation rate constant with increase in temperature and some other studies have shown an increase in the renaturation rate constant with increase in temperature [8, 41]. In general it seems that the temperature dependency of the renaturation rate constant is of non-Arrhenius one and non-monotonic type [42]. Simulation studies suggested that there exists an optimum temperature at which the renaturation rate constant is a maximum [29]. In our model the non-specific-contact formation, nucleation and zipping steps are all influenced by the rise in temperature in a complicated way.

Actually in [Eq 16](), the Smolochowski type bimolecular collision rate constant depends on temperature as $k_t = (8k_B T/3\eta)$ where we assume that viscosity of the medium is not changing much in the range of temperature variation and the dissociation rate scales with temperature as $k_r = k_r^0 \exp(-\omega/k_B T)$ in line with transition state theory where ω is the free energy barrier associated with the dissociation of cn-ssDNA complex. The rate constant associated with the microscopic zipping ($k_p \sim l_p k_+$) is connected with the microscopic diffusion coefficient $D_\pm \sim l_p^2 k_+$ associated with the zipping reaction. Apart from these the dimensionless parameter χ_R

corresponding to the overall electrostatic repulsions and the shielding effects of solvent ions at the interface of cn-ssDNA molecules also depends on the temperature as given in Eqs 5 and 6. It seems that the non-Arrhenius type kinetic behaviour arises as a consequence of a complicated interplay between increase in the rate of nonspecific-contact formation and combined effects of increase in the dissociation rate constant and microscopic zipping rate constant as the temperature increases from low to high values. Since $\bar{k}_{HR} \propto k_B T \exp(\omega/k_B T)$ one finds that $\ln \bar{k}_{HR}$ will be a maximum approximately at $T \sim \omega/k_B$ as observed in the simulation studies [29].

Comparison with Experimental Data on DNA Renaturation

Under relaxed conformational state, the 3D diffusion mediated nonspecific contact formation rate scales with the size of c-ssDNA strands as $k_{fR} \simeq k_t \chi_R \sqrt{3L/8l_p}$. Since the negatively charged phosphate backbones of c-ssDNA repel each other the Onsager radius (κ) associated with the collision of c-ssDNA strands will be much higher than the reaction radius (r_R) under relaxed conformational state. When $\kappa \sim 10 r_R$ then one obtains $\chi_R \sim 10^{-3}$. Noting that $k_t \sim 10^9$ M^{-1}s^{-1} and $l_p \sim 1$ base one finds that $k_{fR} \simeq 6 \times 10^5 \sqrt{L}$ M^{-1}s^{-1}. In the calculation of k_t we have used $T = 298$K and viscosity coefficient $\eta \sim 10^{-1}$ kgm^{-1}s^{-1} for aqueous conditions. Single molecule experiments on DNA polymer in aqueous solution suggested [43–44] a 3D diffusion coefficient as $D_L \simeq D_L^o (L/l_p)^{-\alpha}$ where $D_L^o \simeq 1.4 \times 10^{10}$ base^{2}s^{-1} and $\alpha \sim 0.59$–0.71 depending on the condition of the aqueous medium. Furthermore the 3D diffusion coefficient seems to be several orders of magnitude less under crowded cytoplasmic environment [43]. Therefore the 1D diffusion coefficient associated with the slithering dynamics will be 10–10^2 times slower than the 3D one since the local dynamics of individual bases which are involved in the slithering dynamics will be significantly restricted by the adjacent bases apart from the reduced degrees of freedom. In this background one can approximate the 1D diffusion coefficient as $D_o \simeq 10^9$ base^{2}s^{-1} under relaxed conformational state of c-ssDNA strands. Using this one can arrive at an empirical expression for the nucleation rate as $k_N \simeq (10^9/nL)$.

From experimental studies [41] one finds the zipping rate as $k_p \sim 10^6$ bases/s which means that $k_+ \sim 10^6$ s^{-1}. Using these values one obtains the zipping rate as $k_Z \simeq (10^6/c)$ s^{-1}. Noting from experimental observations that $\varepsilon \simeq \chi_R (k_+/k_r)\sqrt{3/8} \sim 10^{-3}$ and using the values of k_+ and χ_R one can obtain the value of the dissociation rate as $k_r \sim 10^6$ s^{-1}. Using the numerical values of (D_o, k_p and k_r) one finds that $Y_A \sim 10^2$ bases and $Y_E \sim 1$ bases. Computational studies on the renaturation of short fragments of c-ssDNA strands suggested [29] the most probable value for the 1D slithering length as $n \sim (4$–$10)$ bases. For the purpose of calculations we use $n = 10$ bases. From Eq 12 one finds that the overall renaturation rate k_{HR} will be a maximum at $L = L_{opt}$ where $L_{opt} \simeq Y_A^2 (c + Y_E)/nY_E$ (Fig 4C and 4D). This can be obtained by solving $\partial_L k_{HR} = 0$ for L. Here k_{HR} scales with L as $k_{HR} \propto \sqrt{L}$ whenever $L < L_{opt}$ since the nonspecific contact formation step will be the dominating component under such conditions. When $L > L_{opt}$ then the scaling with L becomes as $k_{HR} \propto 1/\sqrt{L}$ since the nucleation and zipping steps will be the bottlenecks under such conditions. Upon substituting $L = L_{opt}$ into k_{HR} one can obtain the maximum achievable bimolecular renaturation rate as $\max(k_{HR}) \simeq k_t \chi_R Y_A \sqrt{3/32nl_p}(1 + c/Y_E)$ for the renaturation of repetitive c-ssDNA strands. When the sequence complexity c is much higher than the characteristic length Y_E then one can deduce that $\max(k_{HR}) \propto Y_A \sqrt{Y_E/nc}$ and the maximum achievable renaturation rate seems to be $\max(k_{HR}) \simeq 3 \times 10^7/\sqrt{nc}$ M^{-1}s^{-1}. Upon substituting the numerical values of Y_A and Y_E into the expression for L_{opt} one obtains $L_{opt} \sim 10^4 c/n$. For a sequence of c-ssDNA strands with a complexity of $c \sim n$ one finds that $L_{opt} \sim 10^4$ bases. From

Eq 12 we find that the square root scaling of the renaturation rate on L will be valid only up to $\sim 10^4$ bases in line with the experimental observations [6, 27].

Upon substituting the numerical values of (Y_A, Y_E and k_{fR}) into the expression for the renaturation rate k_{HR} one finds that $k_{HR} \simeq 6 \times 10^5 \sqrt{L}/(1 + 10^{-4} nL + c)$. Interestingly one should note that this functional form will behave as $k_{HR} \simeq 6 \times 10^5 \sqrt{L}/(1 + c)$ whenever $L < (Y_A^2 c/nY_E)$. When $L < L_{opt}$ and $c > 1$ then one obtains that $k_{HR} \simeq 6 \times 10^5 \sqrt{L}/c$. Remarkably this expression for the renaturation rate k_{HR} is much close to the experimentally obtained fitting function of Wetmur-Davidson for the overall bimolecular renaturation rate as $k_2 \simeq 3.5 \times 10^5 \sqrt{L}/c$ for the experimentally measured range of L from 10^2 to 10^4 bases (Eq. 20 in Ref. [6]). It is still not clear [27] whether this scaling relationship will be valid beyond this range of L or not. In this context our theory based on **Scheme III** predicts that the square root scaling of the overall renaturation rate on the length of c-ssDNA will break down beyond L_{opt}. When the inequality conditions given in Eq 15 are not true then one can also show that the average renaturation rate $\bar{k}_{HR}$ will be a maximum at $L = \bar{L}_{opt}$ where $\bar{L}_{opt}$ is the solution of $\partial_L \bar{k}_{HR} = 0$ for L as follows.

$$\partial_L \bar{k}_{HR} = \int_0^L ((n - 2L(n/Y_A)^2)e^{-(n/Y_A)^2}/g_n)dn + 2L^2 e^{-(L/Y_A)^2}/g_L = 0; \quad g_b = (1 + bL/Y_A^2 + c/Y_E)$$

Here the subscript b in g_b can take (n, L). Detailed analysis suggests an approximation as $\bar{L}_{opt} \simeq Y_A^2(c + Y_E)/\bar{n}Y_E$ where $\bar{n} \simeq \int_0^\infty np(n)dn \simeq \sqrt{\pi}Y_A/2$. Clearly we find that $\bar{L}_{opt} < L_{opt}$. This is because the probability density function associated with the 1D slithering lengths as defined in Eq 13 will be valid only when the sequence complexity c is much higher than the maximum possible 1D slithering length Y_A. When $c < Y_A$ then the slithering dynamics associated with a nonspecific contact present in between the cn-ssDNA strands can progress only for c number of steps. Beyond this point either dissociation of cn-ssDNA strands or relocation of the nonspecific contact to some other position of cn-ssDNA is necessary. Therefore under such conditions one can replace the probability density with a delta function as $p(n) \simeq \delta(n-c)$. This means that we need to substitute as $\bar{n} \sim c$ in our calculations whenever $c < Y_A$ and one obtains $\bar{L}_{opt} \simeq Y_A^2/Y_E \sim 10^4$ in the present context.

Justifications for the Three-Step Model

In **Scheme III** the nucleation rate is inversely proportional to the length and zipping rate in inversely proportional to the complexity of the c-ssDNA strands. Therefore it is reasonable to cluster both nucleation and zipping steps together and assume that they are the rate limiting ones compared to the nonspecific contact formation step. Unlike the rate of nucleation and zipping the nonspecific contact formation rate increases with the length of c-ssDNA in a square root manner. We substantiate the three steps of renaturation by the following reasons viz. (**a**) the underlying microscopic processes are clearly dissimilar in cases of nonspecific contact formation, nucleation and zipping. Here nonspecific contact formation is a pure 3D diffusion mediated collision process. Whereas nucleation involves a combination of 1D and 3D diffusion. The zipping step is pure 1D diffusion like process which progresses from a stable nucleus, and (**b**) the scaling relationships associated with the corresponding rates on the length of reacting c-ssDNA and sequence complexity are different from each other. Clearly nonspecific contact formation, nucleation and zipping are all phenomenologically distinct processes which substantiate our three-step model. Since the nucleation and zipping and parts of a continuous process and a nucleated cc-ssDNA molecule is indistinguishable from the zipping one we have

combined the nucleation step with the zipping step for the purpose of computing the rate associated with the overall nucleation-zipping. Here one should note that cn-ssDNA which is the product of nonspecific contact formation step is distinct from nucleated/zipping cc-ssDNA molecule.

Comparison with Earlier Diffusion Based Models

Since 3D diffusion based models predict the scaling of overall renaturation rate on the length and sequence complexity of c-ssDNA strands as $k_{HR} \propto L/c$ one can rule out the possibility of **Scheme I**. Two step models as in **Scheme II** correctly predict the scaling of renaturation rate on the length and complexity of the reacting c-ssDNA strands. However models based on **Scheme II** such as the Wetmur-Davidson model [6] fail when $L = c$ apart from their inability to explain the discrepancy in the intercept value of the bimolecular renaturation rate at $L = c = 1$. Models based on transition rate theory cannot explain the inverse viscosity dependence of the overall renaturation rate.

Sikorav et.al [27] suggested a Kramer's type expression for the bimolecular nucleation rate constant where the scaling dependency of the overall bimolecular collision rate associated with renaturation on the length of c-ssDNA mainly originates from the entropic component of free energy barrier. However in this model the reaction coordinate and origin of free energy barrier associated with the nucleation and zipping are not clearly defined. Further the exact mechanism of formation of nucleation sites is not clearly explained. On the other hand as correctly pointed out by them, one cannot explain all the experimental observations related to the entire process of renaturation of c-ssDNA molecules with purely diffusion-controlled formalism or transition state theoretical framework. From our model we can conclude that the nonspecific contact formation step is a pure three dimensional diffusion controlled collision rate processes whereas both nucleation and zipping steps involve a sequence of several microscopic crossings of free-energy barriers as well as one dimensional diffusion type slithering dynamics on a linear lattice.

Effects of Conformational State of DNA on the Renaturation Rate

Condensed conformational state of c-ssDNA polymers is one more cause for the breakdown of the scaling of renaturation rate on the length of c-ssDNA that is given in Eq 16. When the colliding c-ssDNA molecules are in condensed conformational state then the rate constant associated with the nonspecific-contact formation step will be independent of the length of c-ssDNA when $L = l$. Under such conditions the overall second order rate constant associated with the renaturation of repetitive c-ssDNA chains will be inversely proportional to the sequence complexity. Here one should note that while deducing these facts we have not considered the condensation of both c-ssDNA molecules together which is known to enhance the overall renaturation rate over several orders of magnitude as in case of renaturation in the phenol-water interface [45]. Under such co-condensation of both strands of c-ssDNA the rate of nonspecific-contact formation is very large and the dissociation rate will be very small and the rate limiting steps are the nucleation and zipping ones.

Conclusions

Renaturation (or hybridization) of complementary single strands of DNA is an important phenomenon in molecular biology and biological physics. Understanding the kinetic mechanism of renaturation is very much useful to further understand the winding-unwinding dynamics of double stranded DNA under both *in vitro* and *in vivo* conditions. Here we have developed a stochastic dynamics based model on the DNA renaturation phenomenon to explain various

scaling behaviours of renaturation rate. According to our model there are at least three steps in the renaturation process viz. nonspecific-contact formation, stochastic nucleation and zipping. Most of the earlier two-state models combined nucleation with nonspecific-contact formation step. We argue that it is considerably meaningful when we combine the nucleation with the zipping since nucleation is the initial step of zipping. Nonspecific-contact formation step is a pure three-dimensional diffusion controlled collision process. On the other hand nucleation involves several rounds of one-dimensional slithering dynamics of one single strand of DNA on the other complementary strand in the process of searching for the correct-contact and initiate nucleation. Upon nucleation, the stochastic zipping follows to generate a fully renatured double stranded DNA.

It seems that the square-root dependency of the overall renaturation rate constant on the length of reacting single strands originates mainly from the geometric constraints in the diffusion controlled nonspecific-contact formation step. On the other hand the inverse scaling of the renaturation rate with the sequence complexity originates from the stochastic zipping which involves several rounds of crossing of free-energy barrier at microscopic level. When the sequence of renaturing single strands of DNA is repetitive with less complexity then the cooperative effects will not be noticeable since the parallel zipping will be a dominating enhancement factor. However for DNA strand with high sequence complexity and length one needs to consider the cooperative effects both at microscopic and macroscopic levels to explain various scaling and kinetic behaviours of the overall renaturation rate.

Appendix

A. Nucleation via 1D Diffusion over Linear Lattice

The searching of the non-specifically bound cn-ssDNA strands for the correct-contacts on each other can be modelled as 1D random walk on a linear lattice. The stochastic differential equation associated with such an unbiased random walk on a linear lattice can be written as follows [36–38].

$$dx/dt = \sqrt{2D_o}\Gamma_t; \ x \in (0, n); \ \langle\Gamma_t\rangle = 0; \ \langle\Gamma_t\Gamma_{t'}\rangle = \delta(t - t') \tag{A1}$$

Here x is the position of the random walker on the linear lattice, D_o is the one dimensional diffusion coefficient associated with the dynamics of the random walker, Γ_t is the Gaussian white noise with mean and variance as given in Eq A1. The probability density function associated with the dynamics of such random walker obeys the following forward Fokker-Planck equation (FPE) with initial condition.

$$\partial p(x, t|x_0, t_0)/\partial t = D_o \partial^2 p(x, t|x_0, t_0)/\partial x^2; \ p(x, t0|x_0, t_0) = \delta(x - x_0) \tag{A2}$$

The boundary conditions associated with Eq A2 are $p(0,t|x_0,t_0) = p(n,t|x_0,t_0) = 0$. Here $p(x,t|x_0,0)$ is the probability of observing the random walker at position x at time t with the condition that the random walker was at position x_0 at $t = t_0$. Settings $t_0 = 0$, the general solution to Eq A2 for the appropriate initial and boundary conditions can be obtained by the method of Eigen function expansion using biorthogonal set as follows [36–38].

$$p(x, t|x_0, 0) = (2/n)\sum_{k=0}^{\infty} \exp(-k^2\pi^2 D_o t/n^2)\sin(k\pi x_0/n)\sin(k\pi x/n) \tag{A3}$$

The mean first passage time (MFPT) associated with the escape of a random walker obeying Eq A3 from the interval $x \in (0, n)$ starting from an arbitrary lattice point x inside the interval

obeys the following backward type Fokker-Planck equation.

$$D_o(d^2\tau_x/dx^2) = -1; \ \tau_0 = \tau_n = 0; \ \tau_x = (n^2 - x^2)/2D_o; \ \bar{\tau}_c = (1/n)\int_0^n \tau_x dx \simeq n^2/12D_o \quad (A4)$$

Since the random walker can enter initially anywhere in interval $x \in (0, n)$ of linear lattice with equal probabilities one needs to average the computed MFPT τ_x over all the values of initial positions x. As in Eq A4 we find the initial position averaged value as $\bar{\tau}_c \simeq n^2/12D_o$. This is approximately the time that is required by a random walker to visit all the sites of a linear lattice confined inside the interval $x \in (0, n)$ starting from anywhere inside the interval. This is evident from the following arguments. When we introduce reflecting boundaries at $x = 0$ as well as $x = n$ then $[\partial_x p(x,t|x_0,0)]_{x=0} = [\partial_x p(n,t|x_0,0)]_{x=n} = 0$ are the corresponding boundary conditions. The probability density function associated with the dynamics of such a random walker confined inside those reflecting boundaries can be given as follows.

$$p(x,t|x_0,0) = 1/n + (2/n)\sum_{k=1}^{\infty} \exp(-k^2\pi^2 D_o t/n^2)\cos(k\pi x_0/n)\cos(k\pi x/n) \quad (A5)$$

From this equation one can conclude that $p(x,t|x_0,0) \simeq 1/n$ whenever $t \geq (n^2/\pi^2 D_o)$ which is close to the initial position averaged mean first passage time $\bar{\tau}_c$. In other words in the presence of reflecting boundaries at both the ends of the linear lattice, the probability of observing the random walker anywhere within those boundaries will be equal when $t > \bar{\tau}_c$.

B. Zipping of Repetitive c-ssDNA Sequences

The zipping of cc-ssDNA strands will also be a stochastic process which can be described by the following birth-death master equation.

$$\partial_t P(u, t) = k_+ P(u - 1, t) + k_- P(u + 1, t) - (k_+ + k_-)P(u, t) \quad (B1)$$

Here $P(u, t) = P(u,t|u_0,t_0)$ is the probability of finding the cc-ssDNA with u numbers of correct contacts at time t starting from the nucleation at $t = t_0$ with $u = u_0$, k_+ (s^{-1}) and k_- (s^{-1}) are the respective average forward and reverse rate constants associated with the microscopic zipping reaction. Here the initial and boundary conditions corresponding to Eq B1 can be written as follows.

$$P(u, t_0) = P(u, t_0|u_0, t_0) = \delta(u - u_0); \ k_- P(1, t) = k_+ P(0, t); \ P(\beta + 1, t) = 0 \quad (B2)$$

The mean first passage time associated with the complete zipping of cc-ssDNA strands obeys the following backward type master equation with similar boundary conditions.

$$k_+ U(u) - k_- U(u - 1) = -1; \ U(u) = \tau(u + 1) - \tau(u); \ \tau(\beta + 1) = 0; \ \tau(-1) = \tau(0) \quad (B3)$$

Here $u = 1$ is a reflecting boundary and $u = \beta$ is the absorbing boundary. One can solve the difference equation Eq B2 as follows. By defining equilibrium constant as $K_Z = (k_-/k_+)$, Eq B2 can be rewritten in the following form.

$$k_+ u\phi(u)[\Omega(u) - \Omega(u - 1)] = -1; \ \phi(u) = \prod_{w=2}^{u} K_Z; \ \Omega(u) = U(u)/\phi(u) \quad (B4)$$

Upon solving [36–38] this difference equation for the boundary conditions given in Eq B3 we find the following expression for the overall mean first passage time associated with complete zipping of β correct contacts ($u = \beta$) of cc-ssDNA starting from the number of correct

contacts $u = 1$.

$$\tau_Z = \sum_{u=1}^{\beta} \phi(u) \sum_{w=1}^{u} (k_+ \phi(w))^{-1} = (K_Z^{\beta+1} - K_Z(\beta + 1) + \beta)/k_+(1 - K_Z)^2 \qquad (B5)$$

C. Cooperative Effects on Non-Repetitive c-ssDNA Sequences

In the presence of cooperative effects the probability of formation of an additional correct contact in cc-ssDNA will be directly proportional to the already exiting number of correct contacts. Similarly the probability associated with the breaking of a correct contact will be directly proportional to the already exiting single stranded overhangs of cc-ssDNA. In the background the birth-death master equation described by Eq B1 can be rewritten to include the cooperative effects for the renaturation of a nonrepetitive single cc-ssDNA as follows.

$$\partial_t P(u, t) = k_+(u - 1)P(u - 1, t) + k_-(\beta - u - 1)P(u + 1, t) - (k_+ u + k_-(\beta - u))P(u, t) \quad (C1)$$

The mean first passage time $\tau(u)$ associated with evolution of the system from correct-contact $u = 1$ to complete dsDNA form with correct-contacts $u = \beta$ can be written as follows where $U(u)$ and other boundary conditions are defined as in Eq B3.

$$k_+ u U(u) - k_-(\beta - u)U(u - 1) = -1; \ k_+ u \phi(u)[\Omega(u) - \Omega(u - 1)] = -1 \qquad (C2)$$

Here $\Omega(u) = U(u)/\phi(u)$ and the function $\phi(u)$ is defined as follows.

$$\phi(u) = \prod_{w=1}^{u} K_Z(\beta - w)/w = (-K_Z)^u \Gamma(u + 1 - \beta)/\Gamma(u + 1)\Gamma(1 - \beta); \ K_Z = k_-/k_+ \quad (C3)$$

Upon solving the difference equation Eq C2 for appropriate boundary conditions one obtains the following expression for the overall zipping time that is required for the formation of $u = \beta$ numbers of correct-contacts starting from $u = 1$ in the presence of cooperative effects.

$$\tau_Z = \sum_{u=1}^{\beta} K_Z^u (\xi {}_2F_1([1, 1], [2 - \beta], -K_Z^{-1}) + \phi {}_2F_1([1, u + 1], [u + 2 - \beta], -K_Z^{-1})) \qquad (C4)$$

Here ${}_2F_1$ is the hypergeometric function and we have defined various parameters as follows.

$$\xi = (-1)^u \Gamma(u + 1 - \beta)/k_- \Gamma(u + 1)\Gamma(1 - \beta)(\beta - 1);$$
$$\phi = \Gamma(u + 1 - \beta)/K_Z^{u+1} k_+ \Gamma(u + 2 - \beta) \qquad (C5)$$

The hypergeometric function of type ${}_2F_1$ is defined as follows.

$${}_2F_1([a, b], g, z) = \sum_{m=0}^{\infty} z^m (a)_m (b)_m / m!(g)_m; \ (h)_q = \Gamma(h + q)/\Gamma(h) \qquad (C6)$$

To simplify the complicated expression for τ_Z in Eq C3 particularly for sufficiently large values of β one can approximate Eq C3 by the following continuous type Fokker-Planck equation (FPE) [36–38].

$$\partial_t P(u, t) = -\partial_u(A(u)P(u, t)) + \partial_u^2(B(u)P(u, t))/2 \qquad (C7)$$

Here the drift and diffusion coefficients can be written as follows.

$$A(u) = k_+ u - k_-(\beta - u); \ B(u) = k_+ u + k_-(\beta - u) \qquad (C8)$$

Eqs C7 and C8 suggest that in the presence of cooperative effects the phenomenological diffusion coefficient associated with the zipping dynamics ($D_\pm$) will be dependent on the number of correct-contacts. Using the backward type FPE corresponding to Eq C7 one can obtain the

mean first passage time associated with the evolution of the system starting from $u = 1$ to $u = \beta$ as follows.

$$\tau_Z \simeq (2/k_+) \int_1^\beta (\mathrm{H}(y)/\Phi(y)) dy; \; \mathrm{H}(y) = \int_1^y (\Phi(w)/(w + K_Z(\beta - w))) dw \quad (\mathrm{C9})$$

In this equation various functions and parameters are defined as follows.

$$\Phi(q) = \exp(2 \int_1^q p(w) dw); \; p(w) = A(w)/B(w) = (w - K_Z(\beta - w))/(w + K_Z(\beta - w)) \quad (\mathrm{C10})$$

Computational analysis of Eqs C7 and C9 suggests that in the limit as K_Z tends towards zero, the overall zipping time τ_Z approximately scales with β as $1\text{-}e^{-2\beta}$. Upon defining the limit as $\lim_{K_Z \to 0} \tau_Z = \tilde{\tau}_Z$ one can derive the following expression for the overall zipping time.

$$\tilde{\tau}_Z = (2/k_+) \int_1^\beta e^{-2y}(\mathrm{Ei}(1, -2) - \mathrm{Ei}(1, -2y)) dy; \; \mathrm{Ei}(a, z) = \int_1^\infty e^{-mz} m^{-a} dm \quad (\mathrm{C11})$$

This equation suggests that for a sufficiently large value of β, in the presence of cooperative effects the overall zipping time will be almost independent of β since $\lim_{\beta \to \infty} \partial_\beta \tilde{\tau}_Z = 0$.

Author Contributions

Conceived and designed the experiments: RM. Performed the experiments: RM. Analyzed the data: RM GN. Contributed reagents/materials/analysis tools: RM. Wrote the paper: RM.

References

1. Watson JD, Crick FHC. Molecular structure of nucleic acids. Nature 1953; 171: 737–738. PMID: 13054692

2. Benjamin L. Genes VII. Oxford University Press, London, 2000.

3. Pinheiro AV, Han D, Shih WM, Yan H. Challenges and opportunities for structural DNA nanotechnology. Nat. Nanotechnol. 2011; 6: 763–772. doi: 10.1038/nnano.2011.187 PMID: 22056726

4. Bath J, Turberfield AJ. DNA nanomachines. Nat. Nanotechnol. 2007; 2: 275–284. doi: 10.1038/nnano.2007.104 PMID: 18654284

5. Qian L, Winfree E. Scaling up digital circuit computation with DNA strand displacement cascades. Science. 2011; 332: 1196–1201. doi: 10.1126/science.1200520 PMID: 21636773

6. Wetmur JG, Davidson N. Kinetics of renaturation of DNA. J. Mol. Biol. 1968; 31: 349–370. PMID: 5637197

7. Studier FW. Effects of conformation of single-stranded DNA on renaturation and aggregation. J. Mol. Biol. 1969; 41: 199–209. PMID: 5801474

8. Craig ME, Crothers DM, Doty P. Relaxation kinetics of dimer formation by self-complementary oligonucleotides. J. Mol. Biol. 1971; 62: 383–392. PMID: 5138338

9. Davidson EH, Hough BR, Klein WH, Britten RJ. DNA sequence organization in the mollusc aplysia californica. Cell. 1975; 4, 217–238.

10. Manning GS. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q. Rev. Biophys. 1978; 11: 179–246. PMID: 353876

11. Manning GS. On the application of polyelectrolyte limiting laws to the helix-coil transition of DNA. V. ionic effects on renaturation kinetics. Biopolymers. 1976; 15: 1333–1343. PMID: 949538

12. Anderson MLM. Nucleic Acid Hybridization (Bios Scientific, Oxford), 1999.

13. Bloomfield VA, Crothers DM, Tinoco Jr, I. Physical Chemistry of Nucleic Acids; Harper and Row: New York, 1974.

14. Rau DC, Klotz LC. A more complete kinetic theory of DNA renaturation J. Chem. Phys. 1975; 62: 2354–2366.

15. Cantor CR, Schimmel PR. Biophysical Chemistry. Part III. The Behavior of Biological Macromolecules; W. H. Freeman and Company: New York, 1980.

16. Murugan R. Revised theory on DNA renaturation kinetics and its experimental verification. Biochem. Biophys. Res. Comm. 2002; 293: 870–873. PMID: 12054552

17. Murugan R. A stochastic model on DNA renaturation kinetics. Biophys. Chem. 2003; 104: 535–541. PMID: 12914901

18. Santa Lucia J. Jr and Hicks D. The thermodynamics of DNA structural motifs. Annu. Rev. Biophys. Biomol. Struct. 2004; 33: 415–440. PMID: 15139820

19. Gao Y, Wolf LK, Georgiadis RM. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. Nucleic Acids Res. 2006; 34: 3370–3377. PMID: 16822858

20. Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K, Herschlag D, and Block S. M. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. Science. 2006; 314: 1001–4. PMID: 17095702

21. Zhang DY, Winfree E. Control of DNA strand displacement kinetics using toehold exchange. J. Am. Chem. Soc. 2009; 131: 17303–17314. doi: 10.1021/ja906987s PMID: 19894722

22. Dirks RM, Bois JS, Schaeffer JM, Winfree E, Pierce NA. Thermodynamic analysis of interacting nucleic acid strands. SIAM Rev., 2007; 29: 65–88.

23. Sambriski EJ, Schwartz DC, de Pablo JJ. Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. Proc. Natl Acad. Sci. USA. 2009; 106: 18125–18130. doi: 10.1073/pnas.0904721106 PMID: 19815517

24. Sambriski EJ, Ortiz V, de Pablo JJ. Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways. J. Phys. Condens. Matter, 2009; 21: 034105. doi: 10.1088/0953-8984/21/3/034105 PMID: 21817250

25. Sambriski EJ, Schwartz DC, de Pablo JJ. A mesoscale model for DNA and its renaturation. Biophys J. 2009; 96:1675–1690. doi: 10.1016/j.bpj.2008.09.061 PMID: 19254530

26. Araque JC, Panagiotopoulos AZ, Robert MA. Lattice model of oligonucleotide hybridization in solution. I. Model and thermodynamics. J. Chem. Phys. 2011; 134: 165103–165116. doi: 10.1063/1.3568145 PMID: 21528982

27. Sikorav JL, Orland H, Braslau A. Mechanism of thermal renaturation and hybridization of nucleic acids: Kramers' process and universality in Watson-Crick base pairing. J. Phys. Chem. B, 2009; 113: 3715–3725. PMID: 19673131

28. Niu L, Yanh Z, Zhou J, Mao C, Liang H, Liang D. Mechanism of DNA assembly as revealed by energy barriers. Chem. Commun. 2015; 51: 7717–7720.

29. Ouldridge TE, Sulc P, Romano F, Doye JPK, Louis AA. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. Nucleic. Acids. Res. 2013: 41: 8886–8895. doi: 10.1093/nar/gkt687 PMID: 23935069

30. Murugan R. Generalized theory of site-specific DNA-protein interactions. Phys. Rev. E. 2007; 76: 011901–011909.

31. Murugan R. Theory of site-specific DNA-protein interactions in the presence of conformational fluctuations of DNA binding domains. Biophys. J. 2010; 99: 353–60. doi: 10.1016/j.bpj.2010.04.026 PMID: 20643052

32. de Gennes PG. J. Phys. (Paris) 1976; 37: 1445.

33. de Gennes PG. Scaling Concepts in Polymer Physics (Cornell University Press, Ithaca) 1979.

34. Montroll EW. A note on the theory of diffusion controlled reactions with application to the quenching of fluorescence. J. Chem. Phys. 1946; 14: 202–212.

35. Debye P. Reaction rates in ionic liquids. Trans. Electrochem. Soc (A paper presented at the Eighty-second general meeting, held at Detroit, Mich., October 9) 1942; 265–272.

36. Gardiner CW. Handbook of Stochastic Methods (Springer, Berlin), 2002.

37. Risken H. Fokker Plank Equations (Springer, Berlin), 1996.

38. Van Kampen NG. Stochastic Processes in Physics and Chemistry (North-Holland), 1994.

39. Morrison LE, Stols LM. Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution. Biochemistry. 1993; 32: 3095–3104. PMID: 8457571

40. Porschke D, Uhlenbeck OC, Martin FH. Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs. Biopolymers. 1973; 12: 1313–1335.

41. Porschke D, Eigen M. Co-operative non-enzymatic base recognition. J. Mol. Biol. 1971; 62: 361–381. PMID: 5138337

42. Chen C, Wang W, Wang Z, Wei F, Zhao XS. Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization. Nucleic Acids Res. 2007; 35: 2875–2884. PMID: 17430963

43. Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS. Size-dependent DNA mobility in cytoplasm and nucleus. J. Biol. Chem. 2000; 275:1625–9. PMID: 10636854

44. Robertson RM, Laib S, Smith DE. Diffusion of isolated DNA molecules: Dependence on length and topology. Proc. Natl Acad. Sci. USA.2006; 103: 7310–7314. PMID: 16648255

45. Goldar A, Sikorav JL. DNA renaturation at the water-phenol interface Eur. Phys. J. E. 2004; 14: 211–239.