

Effects of DNA sequence and structure on binding of RecA to single-stranded DNA

R. Bar-Ziv* and A. Libchaber*

Center for Studies in Physics and Biology, The Rockefeller University, New York, NY 10021

Communicated by Robert H. Austin, Princeton University, Princeton, NJ, May 15, 2001 (received for review November 11, 2000)

Fluorescence anisotropy is used to follow the binding of RecA to short single-stranded DNA (ssDNA) sequences (39 bases) at low DNA and RecA concentration where the initial phase of polymerization occurs. We observe that RecA condensation is extremely sensitive to minute changes in DNA sequences. RecA binds strongly to sequences that are rich in pyrimidines and that lack significant secondary structure and base stacking. We find a correlation between the DNA folding free energy and the onset concentration for RecA binding. These results suggest that the folding of ssDNA and base stacking represent a barrier for RecA binding. The link between secondary structure and binding affinity is further analyzed with two examples: discrimination between two naturally occurring polymorphisms differing by one base and RecA binding on a molecular beacon. A self-assembly model is introduced to explain these observations. We propose that RecA may be used to sense ssDNA sequence and structure.

RecA–DNA interaction | DNA secondary structure | fluorescence anisotropy | single nucleotide polymorphism | molecular beacon

Collective interaction of proteins with DNA is fundamental to biological regulation. In some cases, such as gene activation, the interaction is carried out by a handful of different proteins that nucleate on a specific and relatively short DNA sequence to direct synthesis of mRNA (1). In other cases, such as DNA recombination, multiple copies of the same protein interact over hundreds of DNA bases and *polymerize* in a nonspecific manner. So far, it is not clear whether this second type of interaction is specific and if so, what are the mechanisms involved. This question is addressed in this work by using RecA, the protein responsible for recombination in bacteria. A sensitive detection system, fluorescence anisotropy, is used to investigate the initial stages of RecA interaction with its substrate, single-stranded DNA (ssDNA).

A comprehensive review of RecA, its role in recombination, and its DNA binding properties are given in ref. 2. To summarize, on binding ATP or a nonhydrolyzable ATP analog ($[\gamma$ -thio]ATP, ATP γ S), RecA polymerizes on DNA with a site size of three bases per protein (3, 4) and forms a right-handed helix with a 6-RecA pitch, which stretches and rigidifies the DNA (5–7). RecA is also a DNA-dependent ATPase (8, 9); it hydrolyzes ATP when bound to DNA and switches between a high-affinity state, RecA-DNA-ATP, and a low-affinity state, RecA-DNA-ADP (10). Substrates for RecA are both ssDNA and double-stranded DNA (dsDNA). A nucleation barrier characterizes polymerization on dsDNA, whereas binding to ssDNA occurs readily at a high polymerization rate (11, 12). It has generally been considered that RecA binding is not specific, with the exception of a few homopolymers, such as poly(T), to which RecA has enhanced affinity (13, 14). However, recent *in vitro* selection experiments demonstrated that RecA has a preferential affinity for DNA sequences rich in GT composition, some of which are similar to the recombination hotspot χ (15, 16). Differences in affinity are difficult to detect with classical biochemical techniques because the onset of collective binding occurs at low concentration, and, in fact, it was recently shown,

by using surface plasmon resonance, that RecA polymerizes efficiently on certain repeat sequences (17).

Recently, we introduced an approach to study with high resolution RecA polymerization on ssDNA in solution, by using fluorescence anisotropy. This was made possible by using a microdispensing pump, which accurately controlled the addition of small amounts of RecA into the DNA solution over a wide range of concentration (unpublished data). Binding curves as a function of RecA concentration were measured with fluorescent end-labeled ssDNA in the form of poly(T), a polymer that has no secondary structure or base stacking. At low DNA concentration we observed a condensation of RecA on ssDNA. This study showed that the fluorescence anisotropy signal is mostly sensitive to RecA binding in close proximity to the tagged end of the DNA. We examined the effect of ssDNA length on polymerization, and introduced a model of protein assembly that computes the equilibrium distribution of bound RecA domains with good agreement to the data.

In this paper we extend the approach to probe the recognition properties of RecA–ssDNA interaction, simplifying the reactions by using ATP γ S, thus avoiding ATP hydrolysis. To identify high- and low-affinity binding sites, we used a variety of triplet repeats (the size of the binding unit) as substrates, as well as nonrepeat sequences that probe the ability to distinguish small differences. The results obtained suggest simple rules for substrate discrimination. Both secondary structure of self-hybridized sequences and stacked purine-rich sequences represent obstacles for RecA polymerization. When the free energy of ssDNA secondary structure can be computed, we find that RecA concentration at the onset of polymerization depends exponentially on the folding free energy. These elastic and entropic barriers provide a mechanism to discriminate a single base substitution in 63 bases, when such small differences affect local secondary structure or rigidity. Finally, we use a molecular beacon, as a sensitive probe of ssDNA structure, to show that RecA binding in the presence of a well-defined folded state reduces to a two-state system.

Materials and Methods

ssDNA and RecA Protein. In all cases ssDNA oligomers, typically 39 bases long with 5'-fluorescein tags, were used (Midland Certified Reagents, Midland, TX), and purified with reverse-phase HPLC. Triplet repeats are denoted in the text as (NNN)₁₃. Other sequences used are defined as follows: hemoglobin- β (HBB) = ATG GTG CAC CTG ACT CCT GXG GAG AAG TCT GCC GTT ACT, with single point mutation replacing X = A (HBB-N) with X = T (HBB-S), resulting in sickle cell anemia (<http://ncbi.nlm.nih.gov>, GenBank GI = 183944). Single nucleotide polymorphism in p53 (GenBank GI = 4732146) (p53) = GAA

Abbreviations: ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; ATP γ S, $[\gamma$ -thio]ATP.

*To whom reprint requests may be addressed. E-mail: libchbr@mail.rockefeller.edu or barziv@mail.rockefeller.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

GCT CCC AGA ATG CCA GAG GCT GCT CCC CXC GTG
GCC CCT GCA CCA GCG ACT CCT ACA CCG, where X =
C (p53-1) is replaced by X = G (p53-2). A sequence with
minimal secondary structure and equal base composition
(MIX)₃₈ = CTA ATGC CGTA ATGC CGTA ATGC CGTA
ATGC CGTA ATC. The oligonucleotides p53-1 and p53-2 were
purchased with 3'-fluorescein and 5'-trityl to enhance HPLC
purification of the correct length. Molecular beacon (BCN)_{T15} =
(fluorescein-dT)GCGAGT₁₅CTCGC-Dabcyl. DNA oligomer
concentrations were measured at both 492 and 260 nm. RecA
protein (New England Biolabs) binding assays were done in 25
mM Tris·HCl (pH 7.5)/150 mM NaCl/1 mM MgCl₂/1 mM
DTT/100 μM ATPγS. This buffer is similar to the one used in
Biet *et al.* (17).

Anisotropy of RecA-Bound ssDNA. The theoretical principles and
experimental considerations of fluorescence anisotropy are
described in Lakowicz (18). On vertical polarized excitation, light
is emitted with polarization primarily in the vertical direction I_v
but also in the perpendicular direction I_p . The anisotropy is
defined by (18):

$$A = \frac{I_v - I_p}{I_v + 2I_p}.$$

The fluorescent tag has an intrinsic anisotropy, A_0 , which is
reduced due to thermal rotational motion according to:

$$A = \frac{A_0}{1 + (\tau/\theta)},$$

where θ is the rotational correlation time and τ is the fluorescent
lifetime. Large molecules have a longer rotational time than
small molecules (18, 19) and, hence, have a higher fluorescence
anisotropy value. Our recent work showed that the fluorescence
anisotropy of poly(T) is small, and its value increases when RecA
binds close to the labeled end. At high RecA coverage, the
rotation of the bound complex slows down so that the signal
saturates at the intrinsic anisotropy $A_0 = 0.31$ (unpublished data;
see Fig. 1). For the curves presented below, each naked ssDNA
has a different fluorescence anisotropy, A_{DNA} , and we define the
onset of binding R_{onset} as the concentration of RecA at which
 $(A - A_{DNA})/(A_0 - A_{DNA}) = 0.1$.

Measurement Setup. Fluorescence was excited by the 488-nm line
of an air-cooled argon laser (532 Omnicrome, Melles Griot,
Irvine, CA), which was polarized vertically (Glan-Thompson,
Newport, Fountain Valley, CA) and directed to the center of a
quartz cuvette. Light was collected in a "T" format (90° to the
incident beam), projected onto two analyzers oriented parallel
and perpendicular to the polarization vector, and detected by
two sensitive photomultiplier tubes (Thorn EMI Electron Tubes,
Rockaway, NJ). The solution was continually mixed by a stirring
magnet. A simple protein-dispensing pump was constructed by
using a computer-controlled dc motor (860, Newport), a linear
stage, and a Hamilton syringe, dispensing 40–200 μl of RecA
solution into 3 ml of binding buffer. Light intensities were
measured for each of the two polarization directions by using a
photon counter (Hewlett-Packard). The temperature was sta-
bilized by using a water circulation bath at 21.0°C. Binding assays
were obtained by using a computer program (LABVIEW, National
Instruments, Austin, TX) to control the pumping and acquisition.
The protein was dispensed in small aliquots and stirred for
2 min before taking measurements.

Results and Discussion

Base-Stacked Purine-Rich Sequences Hinder RecA Polymerization. We
show that stacked purine-rich sequences (adenine and guanine)

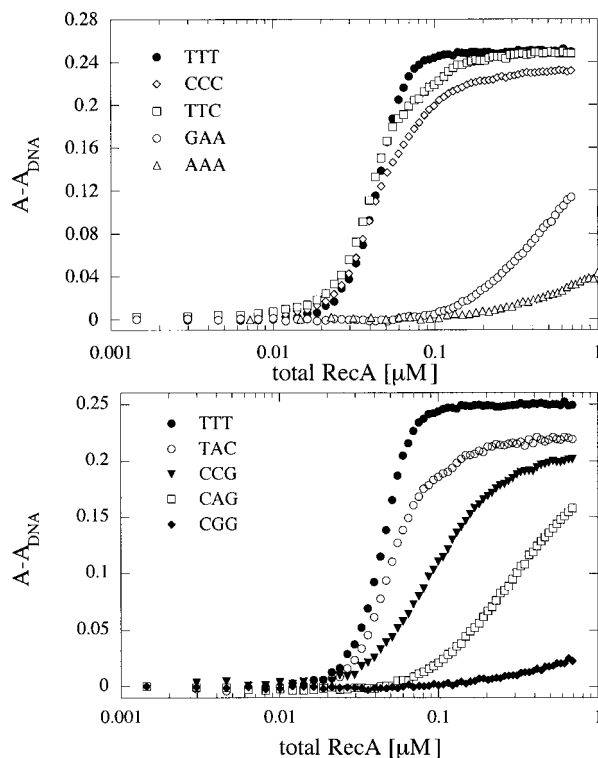


Fig. 1. (Upper) Purines versus pyrimidines. RecA binding curves on ssDNA
triplet repeats, 39 bases long, as designated. The anisotropy of the naked
ssDNA, A_{DNA} , is subtracted from the signal. Concentration of ssDNA is 1 nM
oligomer (13 nM binding sites). (Lower) Effect of ssDNA base pairing. RecA
binding curves on ssDNA triplet repeats with variable G-C base pairing.

hinder RecA polymerization, whereas pyrimidine-rich se-
quences (cytosine and thymine) form high-affinity substrates.
Fig. 1 Upper presents binding curves of RecA to (i) singlet
repeats: poly(T), poly(C), poly(A), and (ii) two triplet repeats,
one with pyrimidines, (TTC)₁₃, and one with purines, (GAA)₁₃.
Poly(T) and poly(C) are flexible homopolymers of pyrimidines
and, hence, pose no elastic resistance to RecA binding. For both
sequences, the onset of polymerization is at low RecA concen-
tration, ≈30 nM, with reduced sharpness for poly(C), indicating
less cooperativity (20). In contrast, poly(A) and poly(G) are rigid
purine homopolymers. As a result, the polymerization onset for
poly(A) is shifted to high concentration, 0.6 μM, reaching no
saturation even at 3 μM RecA. Unfortunately, pure poly(G)
cannot be synthesized. The low-affinity substrate is kept when
purines are mixed, as in (GAA)₁₃, with a slight enhancement
with respect to poly(A). Mixing pyrimidines, (TTC)₁₃, leaves
RecA affinity high, but reduces the sharpness of the condensa-
tion. This may be related to a reduced diffusion, and hence
cooperativity (20), of RecA monomers along a (TTC)₁₃ chain,
whereas on poly(T) they diffuse freely. Finally, to verify that
stacking is indeed important we tested a repeat with 50%
purines, (AC)₂₁, for which stacking of the purines is broken (not
shown). Essentially we find that (AC)₂₁ behaves like poly(C).

Secondary Structure of ssDNA Hinders RecA Polymerization. Second-
ary structures caused by self-hybridization of ssDNA are a major
limitation to RecA polymerization. Fig. 1 Lower shows binding
curves for a selected number of triplet repeats, which have mixed
base composition and various degrees of secondary structure:
(TAC)₁₃, (CCG)₁₃, (CAG)₁₃, (CGG)₁₃, and poly(T) — as a
reference. They are presented in a decreasing order of affinity
to RecA. (TAC)₁₃ is a flexible polymer with 2/3 pyrimidines (TC)

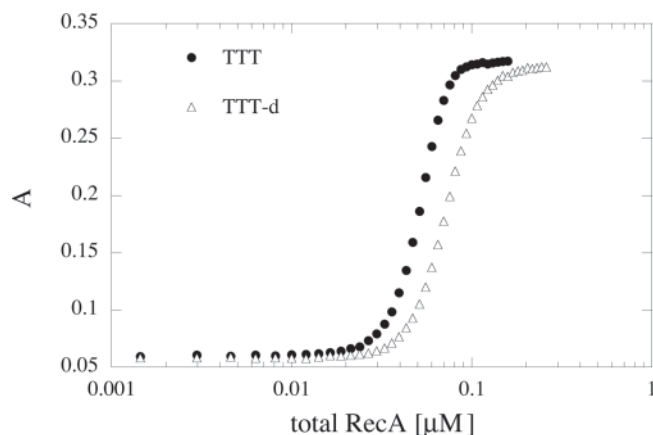


Fig. 2. Effect of “lattice defect.” Introducing a weak binding site AAA in the middle of poly(T) (TTT-d = $T_{18}A_3T_{18}$), forms a weak base pairing hairpin structure, shifting the onset of binding.

with very weak secondary structure, and thus forms a high-affinity substrate. (CCG)₁₃ is also flexible, with an equivalent pyrimidine content (CC), but has secondary structure because of G-C base pairing. As a result, RecA binding is weaker and less cooperative. (CAG)₁₃ is purine rich (AG), with more base stacking and tighter secondary structure than (CCG)₁₃ and, indeed, the affinity is further reduced. Last, (CGG)₁₃ is purine rich with a tight secondary structure, and is, hence, the lowest-affinity substrate. Finally, we embedded one “defect”—a weak binding site (AAA)—in the middle of a poly(T) sequence. The defect induces base pairing of TTT-AAA to form a weakly folded loop. This additional secondary structure decreases the overall affinity of RecA for this sequence, as shown in Fig. 2. These triplet repeats are only a fraction of all of the possible repeats, but they clearly show that secondary structures resulting from base pairing present a barrier for RecA binding.

Folding of ssDNA as a Nucleation Barrier and Implication for Base Stacking. The previous sections presented a trend by which secondary structures induced by base pairing and base stacking hinder RecA polymerization. The fluorescence anisotropy of ssDNA reflects its rigidity, as shown in Fig. 3 (Upper). When ssDNA is folded into a rigid state, its anisotropy is no longer that of a flexible polymer, and this anisotropy increases with rigidity, as expected for globular molecules (18). Can we condense the observed sequence-dependent binding into a more quantitative unified picture? If we consider that RecA polymerizes on ssDNA by unzipping self-hybridization and by unfolding or stretching stacked bases, we can attempt to relate RecA polymerization onset to the secondary structure free energy of ssDNA, whenever the latter is computable. The folding secondary structure and free energy of a ssDNA sequence that has some appreciable self-hybridization can be readily computed, by using, for example, the program *mfold* (ref. 21; <http://mfold2.wustl.edu/mfold>). This program does not compute the free energy of a stacked sequence, such as poly(A), because stacking interactions are not well understood, as we discuss below. In Fig. 3 Lower we plot RecA onset concentration (R_{onset} ; see *Materials and Methods*) as a function of the computed ssDNA folding free energy ΔG for all our 36- to 39-base-long sequences, excluding those that do not self-fold into stable structures (see below). From this plot we obtain a clear correlation: the tighter the folded structure, the harder it is for RecA to bind. In fact, the data fit well to an exponential function: $R_{\text{onset}} \approx \exp(0.09\Delta G)$. This result suggests that the binding onset of RecA is activated with a nucleation barrier that is directly related to the folded state of ssDNA. In

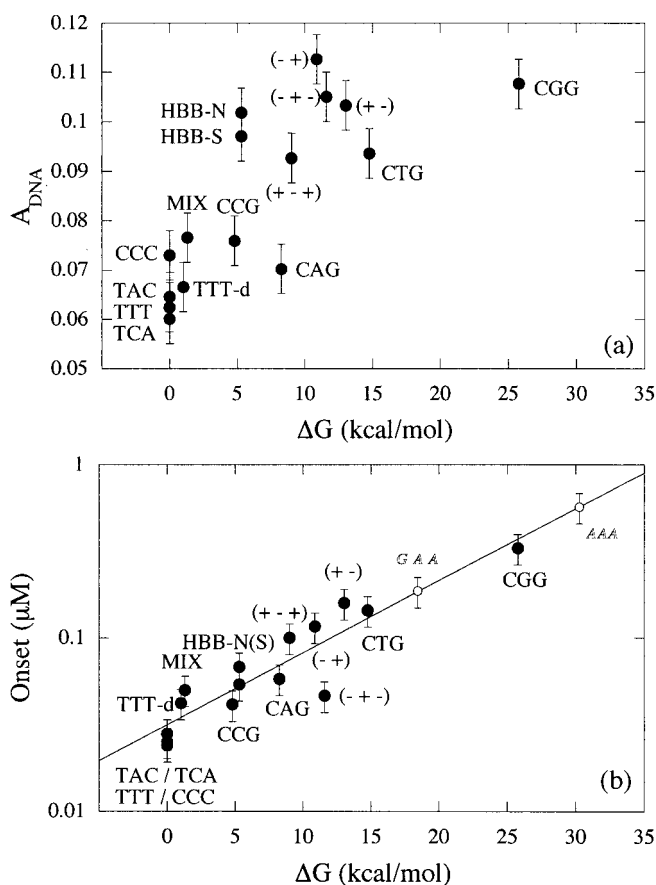


Fig. 3. (Upper) Anisotropy of naked DNA. Anisotropy of ssDNA as a function of computed folding free energy for various sequences: triplet repeats, as designated; HBB-N(S), as described in the text; alternating sequences Fig. 5: (+-), (TTC-CGG)₆; (-+), (CGG-TTC)₆; (+++), (TTC)₃(CGG)₆(TTC)₃; (---), (CGG)₃(TTC)₆(CGG)₃. TTT-d = $T_{18}A_3T_{18}$. Poly(C), poly(T), (TAC)₁₃, and (TCA)₁₃ do not fold into stable structures, and we set $\Delta G = 0$. (Lower) RecA binding to folded ssDNA. Measured onset of RecA binding versus computed folding free energy (●). Solid line, exponential fit: $R_{\text{onset}} = 0.03\exp(0.09\Delta G)$. Free energy values for poly(A) and (GAA)₁₃ were included in the graph by extrapolating the fit (○).

the plot we set $\Delta G = 0$ for poly(T) and poly(C) because they do not fold and stack only weakly. Likewise, (TAC)₁₃ and (TCA)₁₃ do not fold into stable structures, and we set $\Delta G = 0$.

The correlation between RecA binding onset and ssDNA structure may have implications for base stacking. Although it is generally known that stacking interactions of adjacent bases order according to purine–purine > purine–pyrimidine > pyrimidine–pyrimidine, there are significant variations in reported base-stacking parameters. For example, values for AA stacking free energy vary between 1 and 6 kcal/mol (1 kcal = 4.18 kJ) (22). Alternatively, some sources provide stacking interaction parameters only in the context of dsDNA (23). We give an order of magnitude estimate and compare with RecA onset: 1 kcal/mol per AA pair implies 38 kcal/mol for a 39-mer poly(A) and, if we extrapolate our fit in Fig. 3 Lower to include the onset of RecA binding to poly(A), we deduce a value of 30 kcal/mol, which is in reasonable agreement. We therefore propose to use the correlation between self-hybridization folding free energy and RecA onset to estimate the free energy of sequences for which we cannot compute the stacking free energy by extrapolating the fit to read off the free energy at a measured RecA onset. We have included the extrapolated free energy values for poly(A)

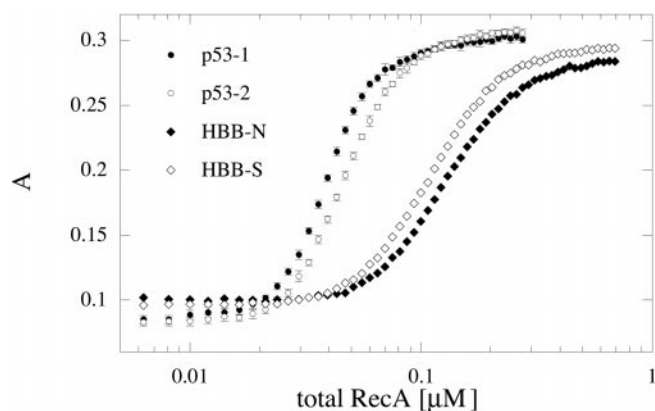


Fig. 4. Detecting single base substitution. RecA binding to hemoglobin-[β] HBB-N and HBB-S sequences, which differ by 1 base in 39, and to p53-1 and p53-2, which differ by 1 base in 63 (see text).

and (GAA)₁₃ in Fig. 3 *Lower* (with different symbols) as examples.

Single Base Substitution. If the onset of binding is exponentially sensitive to secondary structure, can RecA polymerization discriminate a single base substitution? Here we show two examples in which a single base substitution can be detected, one where the base substitution changes the folding and one where it affects local rigidity. We chose known naturally occurring substitutions for the purpose of illustrating the sensitivity, and both examples are not related to RecA function *in vivo*. The first example is a naturally occurring polymorphism of p53 (codon 72 Arg/Pro), and we designed a 63-mer with the single base change in the middle (p53-1, p53-2, see *Methods and Materials*). This base substitution changes the secondary structure and its folding free energy (21) from 5.7 kcal/mol (p53-1) to 7.5 kcal/mol (p53-2). This change is reflected in the binding; RecA polymerizes better on p53-1, as shown in Fig. 4, with a discrimination between the two sequences that is within the error bars.

The second case is a single substitution that does not induce a computable change in secondary structure. We selected the single base mutation occurring in hemoglobin-[β] subunit and designed two sequences, HBB-N and HBB-S (see *Methods and Materials*). Fig. 4 clearly shows that RecA binding to the normal sequence (HBB-N) is weaker than to the mutated one (HBB-S). Interestingly, the computation of folded structure does not include stacking and hence the folding free energy of the two sequences is identical (Fig. 3 *Lower*). However, on close inspection of the sequences, we can rationalize the discrimination. The mutation occurs within a rigid segment of 9 purine bases. Replacement of a base with a pyrimidine reduces this rigidity, and we expect enhanced RecA binding, as observed.

Alternating Sequences: Design of Recognition Elements. One can construct sequences that distribute rigidity and secondary structure. Cooperative monomer-monomer interaction on the DNA determines the average length of the RecA-binding unit. This average length corresponds to an effective Hill coefficient, which, in turn, determines the slope of the binding curve, or the degree of nonlinearity. By using simple triplet repeats, we cannot impose any particular form of cooperativity. However, the classification of high and low affinity binding sites, as well as the role of secondary structure, provides a basis for rational design of ssDNA recognition elements, in which the binding affinity and degree of cooperativity in RecA binding can be tuned. An example of such a design is presented in Fig. 5. With TTC and CGG as good (+) and bad (-) sites, these linear sequences are

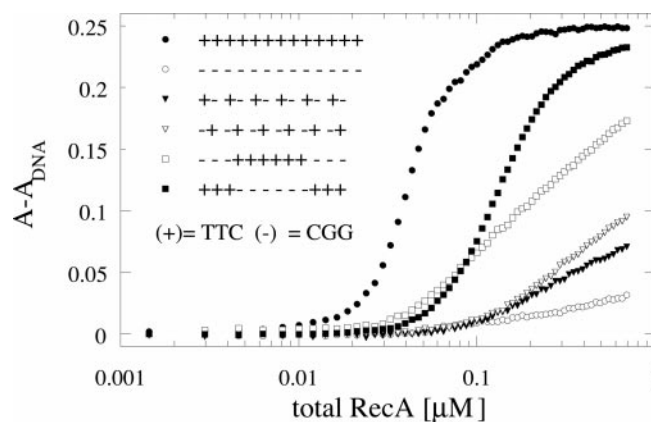


Fig. 5. Alternating sequences. RecA binding to mixed sequences, which distribute secondary structure and rigidity, as designated.

“bipolar.” In the (TTC)₁₃ repeat, each bound monomer enhances the binding of its nearest neighbor, whereas in an alternating TTC-CGG sequence [denoted (+-) and (-+)], every bound monomer antagonizes its nearest neighbor. The sequences (+-+) and (-+-) have “domain walls” between uniform regions of good and bad sites. Each of these sequences folds into a particular structure and, hence, this design relates induced cooperativity to secondary structure. The change in onset and in cooperativity follows the general trend of Fig. 3 (*Lower*).

A Self-Assembly Model. What is the origin of the activation barrier (Fig. 3 *Lower*)? Consider the binding of RecA to ssDNA that has a local self-hybridized folded state with a free energy ΔG . Recently, we introduced a simple adsorption model in which monomers in solution self-assemble on DNA (unpublished work). This model, which is similar to micelle formation, explained the condensation of RecA on poly(T), its length dependence, for both 5' and 3' labeling. Here we modify the model to include a folded state of ssDNA. Because RecA polymerizes in the direction 5' to 3' in equilibrium, there will be more domains close to the 3'-end. For short DNA substrates, as in our case, we consider only a single RecA domain per DNA bound to the 3'-end, and compute the distribution of domain lengths. Each DNA polymer of length L is uniquely defined by the length of the bound RecA domain, l . The system is at multiple chemical equilibria among the species: l -domains, free RecA, and free DNA, at concentrations D_l , R_0 , and D_0 , respectively. The free energy of the system is the sum of the adsorption energy (first term) and the ideal gas entropies of each species (second and third terms), given in units of $k_B T$, by

$$f = \sum_{l=1}^L D_l(-\varepsilon l + \delta_l) + \sum_{l=0}^L D_l(\log D_l - 1) + R_0(\log R_0 - 1). \quad [1]$$

Here, ε is a sequence-averaged monomer adsorption energy and δ_l is the “end-cap” energy, so that the energy of an adsorbed RecA at the end of a domain is $\varepsilon - \delta_l$. Why do we subtract δ_l ? In the absence of any ssDNA folding, the end-cap energy is constant independent of RecA domain length. However, if the DNA sequence has local folding, RecA binding will be penalized by the energy needed to overcome this fold. This implies that above some effective domain length, l_{eff} , the end-cap energy increases by an amount ΔG and, hence, we write

$$\delta_l = \begin{cases} \delta & \text{if } l < l_{\text{eff}}; \\ \delta + \Delta G & \text{if } l_{\text{eff}} \leq l \leq L. \end{cases} \quad [2]$$

Conservation of total RecA and DNA molecules, R_T and D_T , implies $R_0 + \sum_l D_l = R_T$ and $D_0 + \sum_l D_l = D_T$. Minimizing the free energy and using these conservation laws, we obtain an exponential equilibrium distribution of domain lengths:

$$D_l = D_0 \exp((\varepsilon - \mu)l - \delta_l), \quad [3]$$

where $\mu = -\log(R_0)$ is the chemical potential of the free RecA. The mean anisotropy measured is then an ensemble average over the RecA occupancies

$$A = \sum_l A_l D_l, \quad [4]$$

where A_l is the anisotropy signal of a DNA molecule on which l RecA monomers are bound. The fluorescence anisotropy signal increases significantly only when RecA binds close to the labeled end, implying that for 5'-labeling, the onset is determined primarily by the long RecA domains (unpublished data). Therefore, to estimate the onset concentration we consider only one term in Eq. 4, $l = l_{\text{eff}}$ and set $D_{l_{\text{eff}}} \sim \alpha$, with α being a small fractional occupancy ≈ 0.1 . From Eqs. 3 and 2, and because the RecA concentration is not depleted, $R_0 \approx R_T$, we obtain:

$$R_{\text{onset}} \approx \alpha^{1/l_{\text{eff}}} e^{-\varepsilon + \delta/l_{\text{eff}}} \exp(\Delta G/l_{\text{eff}}).$$

We thus recover the exponential shift of the onset with the DNA folded free energy. Comparing the factor in our exponential fit (Fig. 3 Lower) we obtain $l_{\text{eff}} \approx 7$ monomers.

RecA Polymerization on a Molecular Beacon: A Two-State System. Lastly, we use a molecular beacon, a sensitive ssDNA hairpin probe that emits light according to its conformation, to obtain further insight into the mechanism of RecA binding to folded DNA. The beacon comprises a loop and a stem with a fluorophore and its quencher at the opposite ends of the stem (24). When the beacon is closed, in the hairpin conformation, fluorescence is quenched because of the proximity of the fluorophore and the quencher.

Fig. 6 shows that RecA binding unfolds the hairpin structure of a beacon, $(\text{BCN})_{T15}$ (see *Materials and Methods*), resulting in a 28-fold increase of total fluorescence as well as an increase in fluorescence anisotropy. Note that at low RecA concentration the anisotropy signal is larger than the intensity one, a surprising result that is explained below. The combined signal of the beacon provides more information, as compared with end-labeled ssDNA without a quencher. This will suggest that the constraint set by the closed stem on the self-assembly of RecA effectively reduces the beacon to two states: (i) closed with few RecA monomers bound to the loop, and (ii) open and almost fully covered with RecA. We will also obtain an estimate for the DNA folding energy.

Closed beacons may have RecA domains no longer than a length L_* , above which the stem must unzip due to the rigidification of ssDNA by the RecA domain. L is the maximal number of monomers. As explained above, Eq. 3, the distribution of RecA domains bound to ssDNA is exponential in domain length l , $D_l \sim \exp((\varepsilon - \mu)l - \delta_l)$, where μ , ε , and δ_l are the RecA chemical potential, adsorption, and end-cap energies. For the short domains $l < L_*$, the end-cap energy is a constant, $\delta_l = \delta$, whereas for $l > L_*$, it must include the stem folding energy, $\delta_l = \delta + \Delta G_f$. This creates a gap in the length distribution, breaking it into two piece-wise exponentials that separate open and closed beacons, as sketched in the bottom of Fig. 6 for the region where half of the beacons are open.

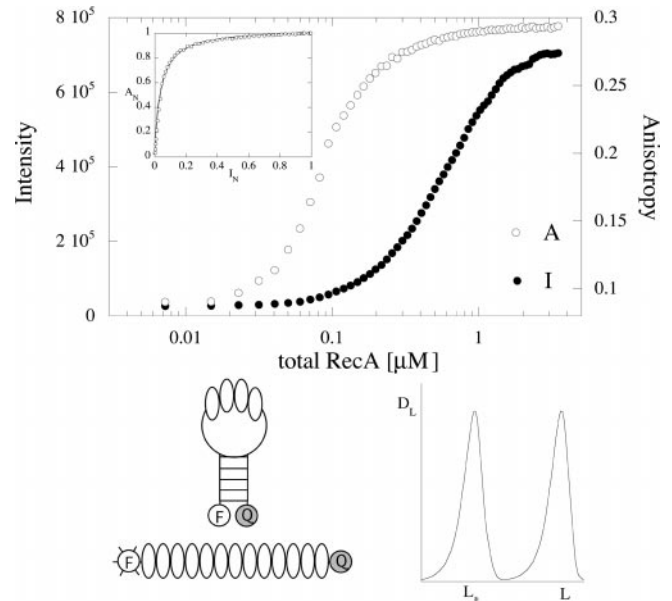


Fig. 6. Molecular beacon. (Upper) Fluorescence anisotropy and total intensity of RecA binding to a beacon with a 15-T loop, $(\text{BCN})_{T15}$. The total intensity of the beacon increases by 28-fold on binding. (Inset) Normalized beacon anisotropy as a function of its intensity (circles) and a two-state model (solid line). (Lower Left) of RecA Polymerization (ellipses) opens the beacon. F, fluorophore; Q, quencher. (Lower Right) sketch of RecA domain length distribution in the vicinity of the half-intensity point, where 50% of the beacons are open.

If closed beacons emit less light than open ones by a factor $\eta \gg 1$, it is straightforward to show that the total intensity and the mean anisotropy are a weighted ensemble average over the intrinsic anisotropies A_l (18):

$$I = \frac{1}{\eta} \sum_{l < L_*} D_l + \sum_{l > L_*} D_l \quad [5]$$

and,

$$A = \frac{\sum_{l < L_*} A_l D_l + \eta \sum_{l > L_*} A_l D_l}{\sum_{l < L_*} D_l + \eta \sum_{l > L_*} D_l}. \quad [6]$$

The distribution D_l is normalized so that $\sum_l D_l = 1$, and for $\eta = 1$ we get $I = 1$, $A = \sum_l A_l D_l$, as the simple ensemble average (Eq. 4). Close to the half-intensity point most of the closed beacons have L_* bound monomers and most of the open ones have L monomers, effectively reducing the system to two states, D_{L_*} and D_L , with $D_{L_*} + D_L \approx 1$. We therefore reduce (Eqs. 5 and 6) to two states, and define the normalized total intensity and anisotropy of the beacon:

$$I_N = \frac{I - I_{L_*}}{I_L - I_{L_*}} \approx D_L, \quad [7]$$

and

$$A_N = \frac{A - A_{L_*}}{A_L - A_{L_*}} \approx \frac{\eta D_L}{(\eta - 1) D_L + 1}. \quad [8]$$

Combining Eqs. 7 and 8, we obtain:

$$A_N \approx \frac{\eta I_N}{1 + (\eta - 1) I_N}. \quad [9]$$

The *Inset* of Fig. 6 shows that, indeed, the beacon fits well to this two-state model with the measured signal-to-noise ratio, $\eta = 28$, and no adjustable parameters. Note that Eq. 9 correctly predicts that the anisotropy precedes the intensity, $A_N > I_N$.

Finally, we estimate the beacon folding free energy. From the intensity and anisotropy signals we find the RecA half-point concentrations ($I_N = 1/2$, $A_N = 1/2$) R_I and R_A , where $D_{L_s} = D_L$ and $D_{L_s} = \eta D_L$, respectively. When our exponential distribution D_I is used, these two relations yield an expression for the beacon folding free energy ΔG_f :

$$\Delta G_f = \frac{\log \eta}{\log(R_I/R_A)} (\log R_I + \varepsilon).$$

With $\varepsilon = 15.5k_B T$ measured for poly(T) (unpublished data) we obtain $\Delta G_f = 4.3k_B T$, which is smaller by a factor 2 than the folding computation.

Summary and Conclusions

This paper addresses the issue of collective binding of RecA protein to DNA and the physical mechanisms that modulate the specificity of this interaction. We presented experiments of RecA polymerization on ssDNA oligomers using fluorescence anisotropy with improved resolution due to a microdispensing system. This sensitive system allowed us to probe the onset of binding at a low concentration regime in which RecA condenses on ssDNA. We have previously shown that the onset is very sensitive to various parameters, such as sequence length and salt (unpublished data). Here we focused on RecA sequence-recognition properties. First, RecA binding sites of low and high affinity were identified by using triplet repeats in the presence of ATP γ S, a nonhydrolyzable analog of ATP. We find that RecA binds well to sequences rich in pyrimidines and that ssDNA rigidity of stacked purine sequences and secondary structure caused by self-hybridization all hinder polymerization. We find a correlation between the onset concentration for polymerization and the free energy of ssDNA folding: the onset of polymerization on all sequences depends exponentially on the folded state free energy. This result is consistent with a simple nucleation model in which RecA domains must unfold DNA secondary structure, which poses an activation barrier for polymerization. A similar idea has recently been proposed in single-molecule experiments that have shown that RecA binds well to stretched DNA, thus implying an important role for thermal fluctuations in the binding (26). The correlation between folding free energy and polymerization onset does not exclude se-

quence-dependent specificity, especially to purine-rich stacked sequences.

The observation that the onset of binding is exponentially sensitive to ssDNA structure and sequence led us to consider the limit of a single base substitution and how it might affect the binding. Two known, naturally occurring substitutions were studied as examples, a single point mutation (1 in 39 bases: sickle cell anemia) and a single nucleotide polymorphism (1 in 63 bases: p53). In both cases, RecA binding curves on the normal and the altered sequences were distinguishable and followed the general trend that rigidity and secondary structure pose resistance to binding. In the first case, the base change affects the local rigidity, whereas in the second case it directly alters the ssDNA folding. We also studied the binding of RecA to a molecular beacon, as an example of a well-defined secondary structure. The barrier for binding imposed by the hairpin reduces the problem to a two-state system.

These measurements lead us to conclude that RecA polymerization, at its onset, is a sensitive detector of DNA sequence and structure. This conclusion may have importance both for the basic understanding of RecA-ssDNA interaction and for the purpose of molecular computation, in which recognition elements of DNA may be useful. The significance of the present study to RecA-DNA recognition *in vivo* requires further work under various conditions, in particular in the presence of ATP. Nevertheless, the identification of simple rules for DNA sequence discrimination and the activation picture presented may provide a basis for identification of possible recombination hotspots. It is instructive to compare our results with experiments of *in vitro* molecular selection that showed that RecA has a preferential affinity for DNA sequences rich in G and T composition (15), some of which are similar to the χ sequence (16). We have looked at the selected sequences (15) and, in fact, the best RecA substrates have minimal secondary structure and lack significant base stacking, which follows the same trends that we have found independently. Specifically, one of best selected sequences, CLONE1 (GCG TGT GTG GTG GTG TGC) (15), folds into an unstable large loop held by the G-C base pairs of the ends and any possible stacking of purines is broken by intervening pyrimidines.

We thank G. Bonnet, M. Calame, B. Dubertret, N. Goddard, D. Lubensky, M. Morpurgo, D. Nelson, B. Shraiman, and D. Thaler for helpful discussions and suggestions. We are especially grateful to K. Adzuma for his support and to T. Tlusty for his help with the theoretical model. This work was supported by the Mather Foundation. R.B.-Z. acknowledges a Burroughs Wellcome Fellowship.

1. Ptashne, M. & Gann, A. (1998) *Curr. Biol.* **8**, R812–R822.
2. Roca, R. & Cox, M. (1997) *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 129–223.
3. Egelman, M. H. & Stasiak, A. (1986) *J. Mol. Biol.* **191**, 677–697.
4. Takahashi, M., Kubitsa, M. & Norden, B. (1989) *J. Mol. Biol.* **205**, 137–147.
5. Stasiak, A., DiCapua, E. & Koller, T. (1981) *J. Mol. Biol.* **151**, 557–564.
6. Leahy, M. C. & Radding, C. M. (1986) *J. Biol. Chem.* **261**, 6954–6960.
7. Egelman, E. H. & Stasiak, A. (1988) *J. Mol. Biol.* **200**, 329–349.
8. Ogawa, T., Wabiko, H., Tsurimoto, T., Horii, T., Masukata, H. & Ogawa, H. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 909–915.
9. Roberts, J. W., Roberts, C. W., Craig, N. L. & Phizicky, E. M. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 917–920.
10. Menetski, J. P. & Kowalczykowski, S. C. (1985) *J. Mol. Biol.* **181**, 281–295.
11. Kowalczykowski, S. C., Clow, J. & Krupp, R. A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3127–3131.
12. Pugh, B. F. & Cox, M. M. (1988) *J. Mol. Biol.* **203**, 479–493.
13. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1981) *J. Biol. Chem.* **256**, 8835–8844.
14. Cazenave, C., Chabbert, M., Toulme, J.-J. & Helene, C. (1984) *Biochim. Biophys. Acta* **781**, 7–13.
15. Tracy, R. B. & Kowalczykowski, S. C. (1996) *Genes Dev.* **10**, 1890–1903.
16. Anderson, D. G. & Kowalczykowski, S. C. (1997) *Cell* **90**, 77–86.
17. Biet, E., Sun, J. S. & Dutreix, M. (1999) *Nucleic Acids Res.* **27**, 596–600.
18. Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy* (Kluwer, Boston), 2nd Ed.
19. Berg, H. C. (1993) *Random Walks in Biology* (Princeton Univ. Press, Princeton), Expanded Ed.
20. McGhee, J. D. & von Hippel, P. H. (1974) *J. Mol. Biol.* **86**, 469–489.
21. SantaLucia, J., Jr. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1460–1465.
22. Fireman, R. A. & Honig, B. (1995) *Biophys. J.* **69**, 1528–1535.
23. Cantor, C. R. & Smith, C. L. (1993) *Genomics* (Wiley, New York), p. 73.
24. Tyagi, S. & Kramer, F. R. (1996) *Nat. Biotechnol.* **14**, 303–308.
25. Bonnet, G., Krichevsky, O. & Libchaber, A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8602–8606.
26. Leger, J. F., Robert, J., Bourdieu, L., Chatenay, D. & Marko, J. F. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12295–12299.