Studies of DNA Breathing and Helicase Mechanisms by Single Molecule (SM) Fret Between 6-MI and Cy3 in DNA Replication Fork Constructs

ARTICLE *in* BIOPHYSICAL JOURNAL · JANUARY 2014

Impact Factor: 3.97 · DOI: 10.1016/j.bpj.2013.11.446

Invited Review

Fifty Years of DNA "Breathing": Reflections on Old and New Approaches

Peter H. von Hippel,[1,2] Neil P. Johnson,[1] Andrew H. Marcus[1,2,3]

[1] *Institute of Molecular Biology, University of Oregon, Eugene, OR 97403*

[2] *Department of Chemistry, University of Oregon, Eugene, OR 97403*

[3] *Oregon Center for Optics, University of Oregon, Eugene, OR 97403*

Received 2 July 2013; accepted 2 July 2013
Published online 10 July 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22347

ABSTRACT:

The coding sequences for genes, and much other regulatory information involved in genome expression, are located 'inside' the DNA duplex. Thus the "macromolecular machines" that read-out this information from the base sequence of the DNA must somehow access the DNA "interior." Double-stranded (ds) DNA is a highly structured and cooperatively stabilized system at physiological temperatures, but is also only marginally stable and undergoes a cooperative "melting phase transition" at temperatures not far above physiological. Furthermore, due to its length and heterogeneous sequence, with AT-rich segments being less stable than GC-rich segments, the DNA genome 'melts' in a multistate fashion. Therefore the DNA genome must also manifest thermally driven structural ("breathing") fluctuations at physiological temperatures that should reflect the heterogeneity of the dsDNA stability near the melting temperature. Thus many of the breathing fluctuations of dsDNA are likely also to be sequence dependent, and could well contain information that should be "readable" and useable by regulatory proteins and protein complexes in site-specific binding reactions involving dsDNA "opening." Our laboratory has been involved in studying the breathing fluctuations of duplex DNA for about 50 years. In this "Reflections" article we present a relatively chronological overview of these studies, starting with the use of simple chemical probes (such as hydrogen exchange, formaldehyde, and simple DNA "melting" proteins) to examine the local stability of the dsDNA structure, and culminating in sophisticated spectroscopic approaches that can be used to monitor the breathing-dependent interactions of regulatory complexes with their duplex DNA targets in "real time." © 2013 Wiley Periodicals, Inc. *Biopolymers 99: 923–954, 2013.*

Keywords: *hydrogen exchange; formaldehyde probing; protein-nucleic acid interactions; DNA replication; DNA base analogue spectroscopy; single molecule DNA-protein interactions studies; two-dimensional fluorescence spectroscopy*

Correspondence to: Peter H. von Hippel; e-mail: petevh@molbio.uoregon.edu
Contract grant sponsor: NIH
Contract grant numbers: GM-15792 and GM-29158 (to P.H.v.H.)
Contract grant sponsor: American Cancer Society Research Professorship (to P.H.v.H.)
Contract grant sponsor: NSF Chemistry of Life Processes Program
Contract grant number: CHE-1105272 (to A.H.M.)
Contract grant sponsor: Office of Naval Research
Contract grant number: N00014-11-0193 (to A.H.M.)

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Background and Overview

The publication of the Watson-Crick structure of duplex DNA in 1953[1] followed shortly by the isolation and characterization of the first DNA-dependent DNA polymerase by Kornberg and coworkers, and the first DNA-dependent RNA polymerase by Weiss and Gladstone,[2,3] made it apparent that there must be mechanisms to allow these polymerases to gain access to the interior of the DNA duplex. This seemed self-evident because such access was clearly required to make available the coding bases of the single-stranded DNA templates used by these polymerases to align the complementary nucleotide triphosphates (NTPs) for the sequence-specific synthesis required for faithful DNA replication and RNA transcription.

Protein-centric ideas posited that polymerases might recognize and bind to specific target site(s) on the double-stranded DNA exterior (at sites that we now know as replication origins or transcriptional promoters) and then somehow "force" their way into the duplex structure to access the single-stranded (ss) DNA template strands. Nucleic acid-centric approaches built on early experimental studies of double-stranded (ds) DNA melting transitions suggested that the melting of dsDNA should be considered a one-dimensional first order phase transition and that therefore the duplex structure must exhibit thermally-induced fluctuations around its "closed" state at lower temperatures, which ssDNA template-specific binding polymerases might somehow be able to exploit. The question was, did these fluctuations exist, and if so could they be characterized to determine whether proteins requiring access to the dsDNA interior make use of them?

Before these considerations could go further it was obvious to many of us at the time that some way would have to be found to observe dsDNA fluctuations (or "breathing") experimentally, and then to determine whether at least some of these fluctuations could be appropriately engaged by genome-regulatory proteins in performing their functions. The purpose of this "Reflections" article is to provide a brief overview of the outcomes of these musings, starting with the inception of such studies about 50 years ago and ending at the present day, given that understanding these breathing fluctuations in the context of modern biophysics continues to be as relevant and important now as it seemed to be then.

To that end this article divides naturally into four major sections. In the first we describe some of the early attempts to identify and to characterize such DNA breathing fluctuations by the use of chemical probing approaches. We follow with a more general discussion of how genome-binding proteins might interact with such fluctuations in accessing their binding targets. In the third section we describe how our group and others have more recently exploited spectroscopic approaches using site-specifically placed base analogues to study DNA breathing reactions and related protein reaction pathways at the equilibrium (steady state) level. We conclude with a brief description of how we are currently extending these methods to monitor the actual kinetics (dynamics) of the breathing processes and related protein reaction pathways in an effort to begin to define the transition states of breathing-dependent protein-nucleic acid interactions and reaction mechanisms in "real time."

Perspective and Apology

The "Reflections" presented in this article, at the request of the editors of this journal, mirror the views, and scientific career of the first author, who has been active on issues related to DNA breathing and fluctuations for over 50 years. At the same time this focus has not resulted in what would normally be a general review of the enormous research area that underlies these issues. From that perspective there is clearly an excessive focus on the work that has come out of our own laboratory over the years, and major contributions from other workers have not always been given their due. Rather, together with my two co-author colleagues who have shared some of the more recent parts of this scientific journey, we describe progress in this field largely from the perspective of our own research and our perspectives of how early studies of DNA breathing and protein binding have led us to our present work. We apologize to colleagues elsewhere whose research this approach has inadvertently slighted, and promise to make amends in future overviews of this field that will focus more broadly.

CHEMICAL PROBE STUDIES OF DNA BREATHING FLUCTUATIONS

Early Hydrogen-Deuterium Exchange Experiments

Probably the earliest attempts to characterize structural fluctuations in biological macromolecules by chemical probe methods were made by Linderstrom-Lang and coworkers in Copenhagen, who used hydrogen-deuterium exchange methods to study the stability of globular proteins.[4–6] These experiments were motivated by still earlier experiments with proteolytic enzymes that had shown that globular proteins tended to be very resistant to enzymatic digestion in their native (although not in their more unstructured "denatured") forms, and that when a proteolytic enzyme did succeed in cleaving an initial peptide bond in a globular protein, digestion quickly went to completion in what Linderstrom-Lang termed

a "proteolytic explosion".[7] However, because of their extremely cooperative ("explosive") character, these enzymatic experiments could not provide much quantitative insight into the sources of structural stability of these globular proteins, nor into the distribution of this stability over the protein. (We note that these enzymatic probing techniques, using both proteases and nucleases, did subsequently lead to quantitative studies of relative local macromolecular stability in fibrous proteins such as collagen and in duplex DNA and chromatin. In such studies these probes presumably sense the relative "breathing" stability of various sequences of secondary structure that control binding—in the unfolded form—into the active site of the enzyme. The stability of the secondary structure of fibrous proteins and dsDNA is less cooperative than that of the tertiary structure of tightly folded globular proteins and thus more easily subject to analysis as a function of position along the DNA or protein sequence.) The Danish workers conjectured that the local unfolding reactions that these enzymes were presumably exploiting to successfully attack the peptide bonds of globular proteins might be subject to more quantitative study by using a "chemical probe" approach to characterize breathing fluctuations in globular proteins, and thus turned their attentions to monitoring the rate at which the deuterium isotopes of deuterated water gained access to (and thus exchanged with) the hydrogen atoms of the peptide hydrogen bonds located in the protein interior.

The Linderstrom-Lang group[4–6] used sensitive microdensitometry gradient methods to determine the amounts of deuterium that had been incorporated into the subject proteins during various time courses of deuterium exchange-in or -out, but of necessity included a drying step to remove the excess D_2O from the system. As it subsequently became apparent that the conformation and stability of globular proteins (as well as duplex DNA) depends not only on the sequence of residues along the nascent polypeptide chain, but also on the nature and composition of the aqueous solvent environment, it also became apparent that the use of an intermediate drying step clearly limited the utility of this early method.

Hydrogen-Tritium (H-T) Exchange Experiments

This problem engaged the attention of Walter Englander, then a postdoctoral fellow with William Harrington at NIH and subsequently with the first author (P.H.v.H.) at Dartmouth Medical School. Englander realized that he should be able to monitor the rates of exchange-in or -out of the tritium isotope of tritiated water (THO) with the hydrogens of the interior hydrogen bonds of the macromolecule, and that these rates should be interpretable as quantitative chemical probes of solvent access to various parts of the macromolecular interior in such experiments. He also appreciated that one could then employ the newly developed Sephadex gel filtration columns to quantitatively separate tritium-labeled macromolecules from their solvent environment, and thus from the excess tritiated water remaining in the solvent, without exposing the macromolecules to a drying step. Furthermore, tritium would serve as a trace label, thus avoiding both conformational changes in the macromolecule due to "bulk" isotope effects and stability perturbations resulting from replacing the H_2O environment with D_2O.

He built on this idea to develop the tritium-Sephadex method that was first used to study breathing fluctuations in solution and at roughly physiological concentrations in both proteins and DNA. In early articles Englander[8] demonstrated that this approach could be used to study protein stability and local "breathing" reactions, and Printz and von Hippel[9,10] showed that the method could also be used to study breathing fluctuations in DNA. It is an interesting coincidence that these first studies on DNA breathing fluctuations were performed in 1963, the same year as the founding of "Biopolymers," whose 50th anniversary this special issue on Nucleic Acid Structure and Function serves to commemorate.

Probing Duplex DNA Breathing Fluctuations with H-T Exchange

These initial tritium exchange experiments showed that the approach could be used to label only the amino and imino hydrogens of the "interior" Watson-Crick (W-C) hydrogen bonds of the DNA duplex, presumably because the exchangeable hydrogens of the "exterior" amino and imino groups of the bases were fully exposed to the solvent environment through the dsDNA grooves, and thus exchanged with solvent H_2O too rapidly to be resolved. The positions of these hydrogens in A-T and G-C base pairs (bps) are shown in Figure 1, and the first H-T exchange experiments on duplex (calf thymus) DNA are shown in Figure 2a, where the logarithm of tritium per bp remaining in the duplex DNA is plotted as a function of time after separating the tritiated dsDNA from its solvent environment (the tritium "exchange-out" time). Figure 2b also shows that the exchangeable hydrogens of the W-C hydrogen bonds of a variety of different types of preparations of calf thymus dsDNA exchange-out essentially as a single exponential class, with some curvature at long-times suggesting that some of the least accessible hydrogens might be exchanging slightly more slowly.

The important points to note in Figure 2a are that over 80% of the W-C hydrogens (both amino and imino) exchanged-out essentially as a single kinetic class, and that the total number of slowly-exchangeable hydrogens present in dsDNA samples as a function of exchange-out time corresponded to the number of inter-base Watson-Crick (W-C)

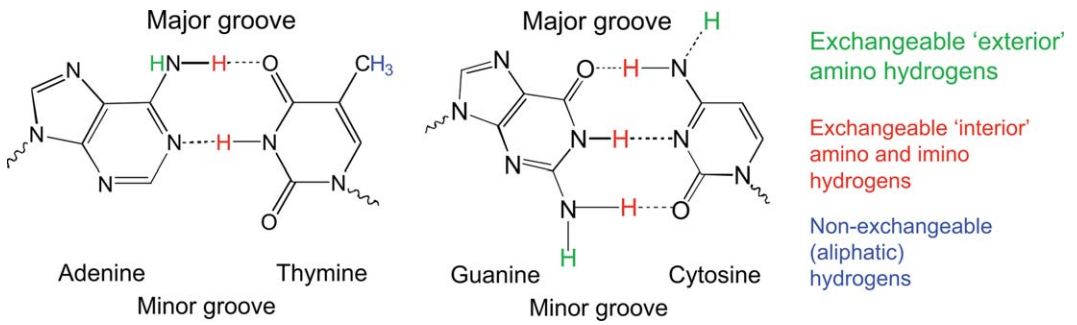


FIGURE 1 Structures of the Watson-Crick base pairs showing the hydrogen moieties that can and cannot undergo exchange.

hydrogen bonds calculated from the base composition of the DNA. The same behavior is seen in Figure 2b, which plots the exchange-out curve for intact calf thymus DNA (42 mole % G + C) with comparable curves for *Micrococcus Lysodeikticus* DNA (72 mole % G + C) and *Clostridium perfringens* DNA (31 mole % G + C). All three exchange-out curves extrapolated back to intercepts corresponding to the expected values of W-C hydrogens per base pair, and all three exchange-out curves otherwise were essentially parallel, showing that a common exchange-out rate is also observed for DNAs of very different base composition and sequence.

Modeling the H-T Exchange Reaction with a Two-Step Mechanism

Having established that the tritium-Sephedex exchange method could accurately measure the number of hydrogen bonds present in the "interior" of the dsDNA, while the exchangeable hydrogens on the surface of the duplex exchanged too rapidly with solvent water to be resolved, it was clear that slopes of the "exchange-out" curves for the tritium probe must reflect, in some way, the kinetics of the transient exposure of the hydrogen bond donors of the W-C bps to the aqueous solvent environment, in order to permit exchange to occur. Following the original treatment set-up for this problem by the Linderstrom-Lang group,[6] we modeled the exchange properties of the "cooperatively exchanging unit" of the DNA duplex by Eq. (1), which represents a simple two-step exchange process:

$$Closed \underset{k_2}{\overset{k_1}{\rightleftarrows}} Open \xrightarrow{k_3} Exchanged \qquad (1)$$

Here k_1 and k_2 correspond to the opening and closing rate constants, respectively, of the bp(s) that comprise the "cooperative breathing unit" that opens and closes in concert as measured by this probe. The ratio $(k_1/k_2) = K$, and k_3 is the

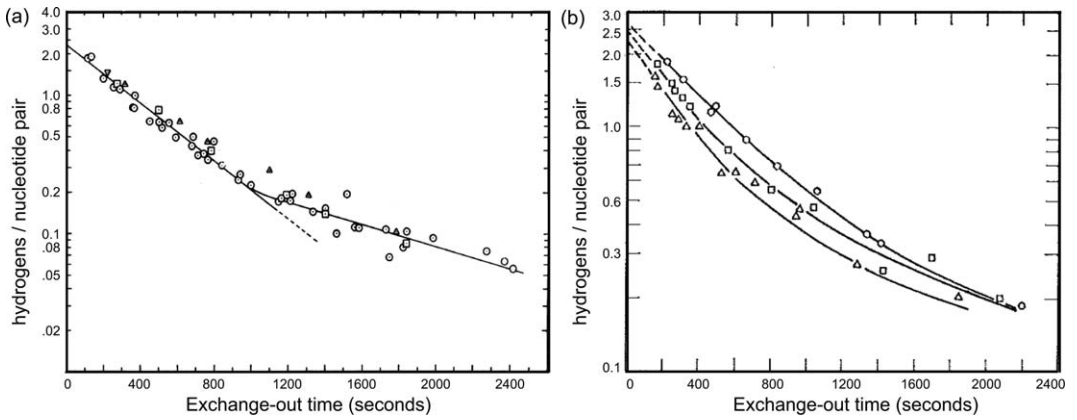


FIGURE 2 (a) Hydrogen-tritium exchange of native calf thymus DNA at 3.5 ± 0.5° in 0.1*M* NaCl; 0.014*M* $(CH_3)_2$ AsOONa (cacodylate), pH 7.6 ± 0.15. (○) DNA, sonicated; (□) DNA, sonicated and EDTA dialyzed; (▽) DNA, unsonicated; (△) DNA (prepared in this laboratory), sonicated.[10] (b) Hydrogen exchange-out curves for native DNAs of different base composition at 0°C (±0.5°C) in 0.1*M* NaCl, 0.01*M* sodium cacodylate, pH 7.6. (○), *Micrococcus lysodeikticus* DNA 72 mole % G+C; (□), calf thymus DNA, 42 mole % G+C; (△) *Clostridium perfringens* DNA, 31 mole % G+C.[11]

exchange rate of the fully open species. [Under our experimental conditions (with the experimental temperature far below T_m), dsDNA is the dominant species present and $k_1 << k_2$. Therefore the fraction of time a given "breathing unit" is open will be $\theta = [k_1/(k_1 + k_2)] \backsimeq k_1/k_2 = K$].

Clearly, based on this two-sequential-step exchange model, there were two possible "limiting-case" scenarios that might be observed. In the first, in which $k_3 >> k_2$, the overall apparent exchange rate should reflect only the rate constant of the opening reaction, k_1, and be independent of changes in the intrinsic chemical rate of hydrogen exchange. In the second limiting-case scenario, in which $k_3 << k_2$, the observed exchange rate should correspond to the product of the chemical exchange rate constant and the fraction of time that the site is open; i.e., Kk_3. If this model applies these scenarios can be easily discriminated, because k_1-dependent exchange should be independent of the rate of the intrinsic chemical exchange process and thus not perturbed by pH changes or the addition of (nonstructure-perturbing) acid-base reaction exchange catalysts (provided these changes do not perturb the opening rate of the "breathing unit" involved). On the other hand, Kk_3-dependent exchange processes should depend on both changes in the pH of the system and on the concentration of exchange catalyst present, because these changes will perturb k_3 and thus the intrinsic exchange rate of the system as a whole.

Duplex DNA H→T Exchange Proceeds by a Kk_3 Reaction Mechanism

Experiments performed at various pH values[9,11,12] showed that the slopes of the exchange-out curves of all the DNAs tested (and thus the overall apparent exchange rate constants) were significantly pH-dependent. Furthermore the exchange-rate minima were observed, as expected, over a narrow pH zone located midway between the two arms of the pH-titration curve of the dsDNA (see Figures 3a and 3b). In these titration curves the acidic arm corresponded to the protonation of the imino N_1 of adenine or the imino N_3 of cytosine (with salt concentration and temperature-dependent pK_a values of ~10), and the basic arm corresponded to the deprotonation of the imino N_3 of thymine or the imino N_1 of guanine (with salt concentration and temperature-dependent pK_a values of ~3).

We note (Figures 3a and 3b) that the apparent tritium exchange-out rate increases as the experimental pH moves closer to the pK_a of either arm of the pH-titration curve, presumably because any of the protonation or deprotonation events listed above will increase the exchange rate and eventually also disrupt the relevant W-C hydrogen bonds (see Figure 1)

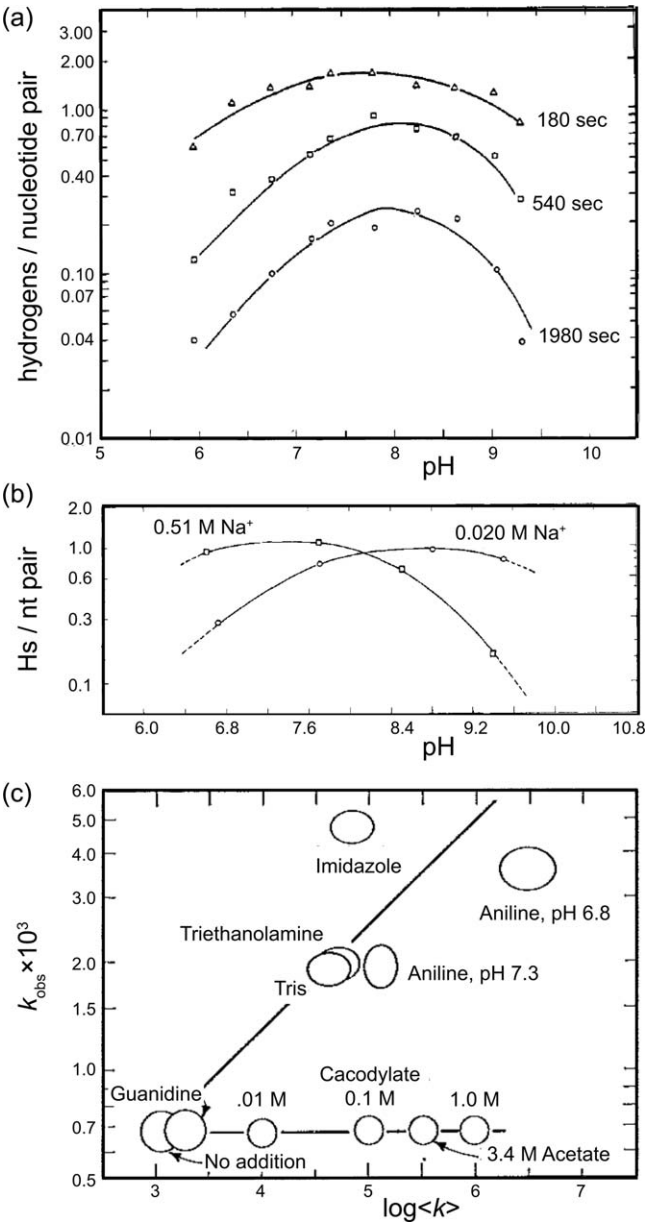


FIGURE 3 (a) Extent of exchange (H/nt) curves at various exchange-out times plotted as a function of pH in 0.11*M* Na^+ after 180, 540, and 1980 s of exchange;[9] (b) 0.51 and 0.020*M* Na^+ after 300 s of exchange.[9] (c) The observed ratio of tritium-hydrogen exchange (plotted on a logarithmic ordinate) versus the logarithm of the expected proton transfer rate $<k>$ calculated as in Ref. 12. Error limits include variation of half-time estimation from tangents as well as experimental error.

and at that point destabilize the local structure of the dsDNA. The apparent rate constants measured at pH values near neutrality were consistent with an acid-base catalyzed Kk_3 hydrogen exchange process at each potential exchange site. Furthermore estimates of k_3 suggested that $K \simeq 10^{-6}$ (the fraction of the total W-C bps of the dsDNA molecule open for H-T exchange at any

instant of time) at the pH-minima of the exchange rates under the various experimental exchange conditions used.

The above view of the exchange of the W-C hydrogens as a standard acid- or base-catalyzed exchange reaction was affirmed by demonstrating the expected rate changes with acid-base catalysts, as shown in Figure 3c for positively charged exchange catalysts (upper line). Melting controls showed that the overall stabilities, melting cooperativities and melting temperatures of the dsDNA molecules tested were unchanged over the range of pH values and exchange catalyst concentrations tested, showing that the effects observed were changing the values of k_3, rather than the dsDNA structure or stability that would indicate changes in K. In contrast negatively charged catalysts were ineffective in catalyzing exchange, presumably because they could not gain access to the exchangeable hydrogens of the W-C base pairs (lower line of Figure 3c) in the face of electrostatic repulsion by the negatively-charged phosphate groups of the sugar-phosphate backbone.

What Could H-T Exchange Measurements Tell us About the Size and Structure of the "Cooperatively Exchanging Unit" of Duplex DNA?

Of course the central motivation of these exchange studies was to attempt to use them to understand the structure of the breathing fluctuations that were transiently exposing the W-C hydrogens (amino and imino) of the bps of dsDNA to the solvent environment, including defining the number of bps that open together; i.e., the size of the "cooperative exchange unit" of Eq. (1). One possibility, of course, was that individual bps were being exposed one-at-a-time, perhaps by some sort of a direct single bp "flipping-out" reaction. This seemed unlikely, because the rate of exchange of A-T and G-C bps at a given pH and temperature would then be expected to differ and thus result in exchange-out curves of different shapes for dsDNA of significantly different base compositions. Thermodynamic principles suggested that because such flipping-out processes must, by definition, require a full unstacking of the bases from those located directly above and below in the duplex DNA, and since base-pair stacking had been shown to be the primary source of stability of duplex DNA, such a flipping process would involve a kinetic mechanism exhibiting a high free energy of activation (This statement is *not* intended to suggest that single bp "flipping out" reactions do not occur; there are several biologically important and enzymatically catalyzed processes (see Discussion) that do facilitate just such reactions. However, our results suggested that DNA breathing fluctuations leading to such single bp flipping should be improbable in the absence of enzymatic catalysis). Furthermore, the complete mechanistic ineffectiveness of the negatively charged acid-base catalysts (Figure 3c) implied that the aqueous solvent must somehow access the net-negatively charged "interior" of the DNA duplex to effect exchange, rather than interacting with individually exposed bases on the dsDNA exterior. All these observations and deductions suggested that the cooperatively breathing unit (or "open" state) of the hydrogen exchange reaction must involve more than one bp.The data of Figures 2a and 2b set an upper limit to the size of this cooperative unit, since the exchange properties of calf thymus dsDNA were not altered by extensive sonication that reduced the average length of the dsDNA fragments to ~300 bps. An additional observation was based on heat denaturation and recooling studies of the rather heterogeneous preparations of calf thymus dsDNA available in the 1970s. These preparations could be melted and recooled to form random networks of short W-C base-paired dsDNA segments separated by noncomplementary ssDNA sequences, and yet showed exchange-out curves with slopes very similar to those of the longer duplexes of Figures 2a and 2b, but in this case extrapolating back to $\sim 1.25 \pm 0.15$ W-C hydrogens per bp. This suggested that ~50% of the denatured DNA had reformed structures with H-T exchange characteristics comparable to those of the native dsDNA preparations. On the other hand these heat denatured and recooled preparations had regained about 75% of their initial absorbance hypochromism and exhibited very broad melting profiles, suggesting that the reformed W-C-paired dsDNA segments of these preparations were, as expected, quite heterogeneous and relatively short. These findings further lowered our estimate of the size of the cooperative exchange unit. Taken together these results suggested that the "cooperatively exchanging" dsDNA segments that correspond to the exchange units of Eq. (1) must be long enough to contain approximately average base-pair compositions and sequences (thus accounting for the roughly constant slopes of the exchange out curves of the different DNA preparations studied), and also be short enough to be significantly reformed within the heat-denatured and recooled calf thymus DNA samples. On this basis, we estimated that the lengths of the cooperatively exchanging dsDNA segments must be on the order of 5 to 10 bps, but of course we were not able to define either the details of the structure or of the open state(s), or—because the reaction mechanism was of the Kk_3 type—the opening-closing (k_1 and k_2) breathing rate constants for the cooperatively exchanging units.

Characterizing H-T Exchange-Monitored dsDNA Breathing by Examining its Temperature Dependence

Study of the hydrogen exchange reaction as a function of $-\Delta T_m$ provided an additional clue to how the hydrogens of

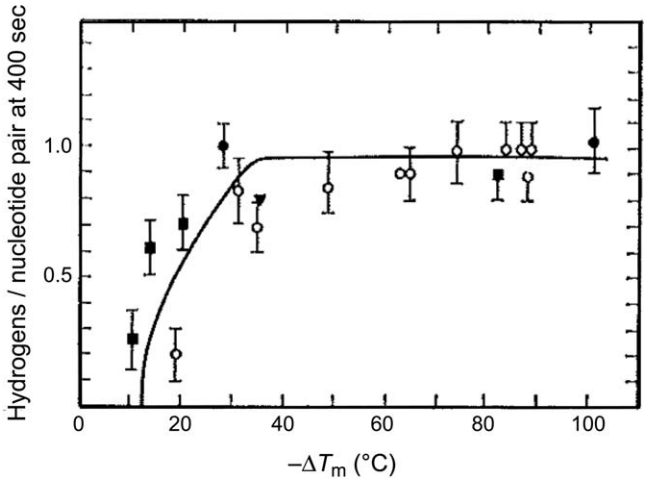


FIGURE 4 Hydrogens per nucleotide pair remaining after 400 s exchange as a function of $-\Delta T_m$ [the difference between the temperature at which the exchange reaction was performed and the melting temperature (T_m) of the DNA duplex] in various concentrations of destabilizing salts (see Ref. 11 (▼) *Escherichia coli* DNA; (○) calf thymus DNA; (•) *Micrococcus lysodeikticus* DNA; (■) *Clostridium perfringens* DNA.)

the W-C bonded interior might be made accessible to the solvent environment. (We note that for experimental consistency the values of $-\Delta T_m$ plotted along the *x*-axis in Figure 4 were not attained by changing the experimental temperature (T), but rather by lowering T_m toward T by adding appropriate types and concentrations of destabilizing salts to the reaction solution.). These data are summarized in Figure 4 for DNAs of four different base compositions and showed that, at temperatures more than ~20°C removed from the melting temperature of the DNA duplex (right side of Figure 4), the exchange rate is essentially independent of temperature. In contrast, the exchange rate plunged steeply as the experimental temperature was moved closer to T_m. This suggested that a low free-energy-of-activation process must dominate exchange at temperatures significantly below T_m, while the dominant exchange reaction appeared to switch to one characterized by a high free-energy-of-activation process at temperatures within ~20°C of T_m.

Defining the Structural Features of Possible "Elemental" Local dsDNA Breathing Mechanisms by Chemical Probe Experiments

It was clear to us early on that if there were to be DNA breathing mechanisms that made the interior W-C hydrogen bonds of the DNA duplex accessible to exchange with the solvent without invoking fluctuation mechanisms resulting in base pair unstacking, the base-pair breathing fluctuations involved must be concerted (or cooperative) in some way. Figure 5 shows an early set of simple schematic cartoons[11] that make this point, and was intended to demonstrate how different types of local breathing fluctuations might explain these different exchange regimes by providing a variety of pathways for access to the dsDNA interior for both tritium (in hydrogen exchange reactions) and for other chemical probes.

Figure 5a was intended to represent the "ground-state" fully stacked and hydrogen-bonded W-C base-paired structure. Figure 5b suggested the structural consequences of a local breathing process in which adjacent bps were "lifted" (and thus unstacked) from one another. We note that adjacent bps in this fluctuation mode unstack without W-C hydrogen bond breakage. Such a fluctuation might be analogized as locally untwisting an initially spiral ladder around its helical axis, thus in effect lifting the rungs off of one another, and might, in principle, be generated by a local "untwisting" fluctuation around the duplex DNA axis, perhaps coupled with a local bending of the dsDNA. Assuming that unstacking is an

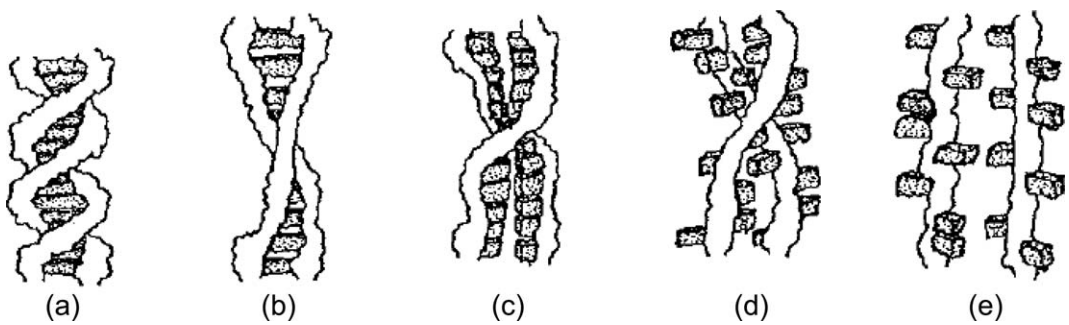


FIGURE 5 Various possible transient local distortions of the native DNA structure. (a) Native structure, stacked, hydrogen bonded and helical. (b) Chains untwisted, bases hydrogen bonded but unstacked. (c) Chains untwisted, bases stacked but not hydrogen bonded. (d) Structure "melted," bases unstacked and not yet hydrogen bonded, partial chain separation. (e) Structure totally "melted," chains separated. From McConnell, B.; von Hippel, P. H., J Mol Biol, 1970, 50, 297-316, © Academic Press, reproduced by permission.

"expensive" process in terms of free energy, Figure 5b might be thought to represent dsDNA breathing fluctuations that would not contribute to hydrogen exchange (the W-C hydrogen bonds remain unbroken), but that could comprise part of a multistep opening process leading to the intercalation of planar ring systems, such as acridine dyes or aromatic amino acid residues, into the dsDNA structure. We elaborate on these possibilities below.

Figure 5c was intended to illustrate a possible local "over-twisting" process that could result in local inter-base-pair hydrogen bond breakage without unstacking, with the duplex strands "backing-away" from one another locally to relieve the strain of over-twisting without incurring the free energy cost of unstacking. Thus here the base-paired duplex is represented as locally "cracked-open" to permit access of the aqueous solvent to the hydrogens of the W-C bps, while preserving base stacking. Since we had speculated that base unstacking should be thermodynamically "expensive," while H-bond breakage should not, such a breathing mechanism (which might also involve a local DNA bending process not shown in Figure 5c) was proposed to account for the low free-energy-of-activation breathing process that resulted in the essentially temperature-independent H-T exchange reaction at temperatures well below T_m (see Figure 4; Ref. 11).

Figure 5d was intended to illustrate a local breathing fluctuation that would result in both local hydrogen-bond breakage and base-unstacking, and likely represented the transition state for interactions of the DNA interior with most chemical probes (e.g., formaldehyde and DNA melting proteins, see below) as well as the hydrogen exchange reactions responsible for the high free-energy-of-activation exchange processes seen at temperatures close to T_m in Figure 4. Finally Figure 5e shows the final chain-separated state that represents the end-point of the complete dsDNA unwinding reaction.

We use this "historical" set of cartoons to illustrate structurally the breathing fluctuations that seemed plausible in 1970. In the last section of these Reflections, we contrast these crude renderings with how such fluctuations might be structurally and dynamically construed via intersecting "free energy landscape" diagrams in the present era, and also how our new spectroscopic approaches can help to define the structures and kinetics of the transition states of the breathing modes actually involved in biologically-relevant dsDNA breathing fluctuations.

Subsequent Hydrogen Exchange Approaches to the Study of DNA Breathing Fluctuations

Obviously the DNA breathing studies monitored by H-T exchange that are summarized above provided fundamental information about some types of dsDNA breathing mechanisms, but also left us with many unanswered questions. After the studies described above, Teitelbaum and Englander[13,14] pushed forward H-T exchange studies of dsDNA and further examined detailed aspects of the acid-base catalysis mechanisms of the exchange process. However, in the late 1970s and early 1980s the fields of both protein and nucleic acid hydrogen exchange were fundamentally altered by a reversion back to the deuterium-exchange methods of Linderstrom-Lang et al., but this time in a "modern" incarnation involving nuclear magnetic resonance (NMR).[15,16] The NMR H-D exchange technique immediately took over from H-T exchange (in Englander's lab and others) as the method of choice for studying issues of globular protein stability and protein folding and unfolding reactions, and has recently (in combination with local mass spectroscopy procedures) led to what looks to be the experimental establishment of folding-unfolding pathways in, at least, some small globular proteins.[17] This approach has also been successfully exploited in measuring some aspects of the stability and breathing dynamics of small folded RNA molecules (especially tRNA; see Ref. 18).

As pointed out above, the central, and very considerable, advantage of the NMR H-D exchange method was (and is) that while previously it had only been possible to speak of "hydrogen-exchange classes" and try to define their molecular origins by numerology and guesswork, it now became possible, in principle, to monitor the rates of exchange of individual hydrogens of defined base pairs in nucleic acids, and of individual hydrogens of the peptide bonds of defined amino acid residues in proteins, which—especially for proteins—has been a real "game-changer." The problem for proteins was easier because small globular proteins (with molecular masses in the 15,000 to 30,000 Da range) were (and are) much more amenable to straightforward NMR study than are the larger protein complexes and fibrous proteins.

Proteins have an additional significant advantage over nucleic acids for this type of study because they contain 20 different types of natural amino acid residues, while nucleic acids contain only four canonical nucleotide residue types. This means that while monitoring the exchange rate of the peptide hydrogen of a particular amino acid residue in a small globular protein is relatively straight forward, monitoring the exchange rate of a particular W-C hydrogen bond within a DNA duplex of length greater than 30 to 40 bps was (and is) virtually impossible unless one uses local isotopic substitution for NMR studies or site-specifically placed base analogues with optical spectroscopic techniques—see below. (This advantage for proteins is partially shared by NMR H-D exchange studies on small structured RNAs, where there are fewer bps of any given type and the

peaks corresponding to the different bases are somewhat spread out by the proximity of a variety of secondary and tertiary structures.) On the other hand, monitoring the exchange rate of specific imino or amino hydrogens of a 10 to 12 bp duplex DNA oligonucleotide was possible, and many such studies have been carried out. Some of these studies have been reviewed elsewhere (see Ref. 19).

As pointed out above, one important advantage of NMR for studying the breathing of small duplex DNA molecules is that the exchanging proton(s) can be assigned to a particular base, and hence dynamics can be probed with single-base resolution. Hydrogen-deuterium exchange times ranging from milliseconds to days have been measured by monitoring imino proton signals using one-dimensional or two-dimensional NMR. The exchange rates of fast-exchanging (on NMR time-scales) imino protons (5 ms to ~10 s) can also be measured by water magnetization transfer methods, in which the magnetization of water protons is selectively perturbed and the transmission of this perturbation to exchangeable imino protons is followed over time. Analysis of hydrogen exchange data by NMR has generally employed a two-state model for base-pair opening similar to that outlined above in Eq. (1) for H-T exchange, with hydrogen exchange only occurring from the open state, except that Eq. (2) presupposes that bps open one at a time. Again a base pair can open multiple times before hydrogen exchange takes place, and equilibrium between the closed and open states of the nucleic acid base pairs is assumed. For example, the hydrogen exchange of the imino proton of a specific thymine base-paired with adenine can be written:[15,16]

$$\mathrm{A{-}{}^{*}T} \underset{k_2}{\overset{k_1}{\rightleftarrows}} \mathrm{A{+}{}^{*}T} \overset{k_2}{\rightleftarrows} \mathrm{A{+}{}^{+}T} \rightleftarrows \mathrm{A{-}{}^{+}T} \qquad (2)$$

where A is adenine T is thymine, the symbols * and $^{+}$ represent an exchangeable proton, and $k_{ex} = K_{diss}\, k_3$ and $K_{diss} = k_1/k_2$.

Just as for the H-T exchange process described above, the overall H-D reaction can be described in terms of three sequential steps: (i) base-pair opening; (ii) hydrogen exchange; and (iii) base-pair closing. Acid–base catalysts such as ammonia can increase the proton exchange rate, and the overall exchange time $\tau_{ex} = 1/k_{ex}$ is proportional to the concentration of catalyst. By extrapolating a plot of exchange time versus 1/catalyst concentration it can be seen that at sufficiently high catalyst concentrations every open bp exchanges at each opening event, the opening step becomes rate limiting and bp lifetimes can be estimated. K_{diss} can also be determined from the slope of such a plot.

Such NMR approaches with short (10–12 bp) dsDNA duplexes have been used to determine the exchange rates of specific imino protons and thus to monitor the opening and closing of specific DNA and RNA bps (for reviews see Refs. 16 and 18). Life times for the opening of bps in short duplex DNA oligonucleotides estimated in this way are typically in the millisecond time range. The fraction of open bps in dsDNA was typically found to be 10^{-5} to 10^{-6} bp^{-1}, in reasonable agreement with the estimates of the fraction of bps open from the early H-T exchange experiments for dsDNA. Finally changes in the free energy of base pair opening (ΔG_{diss}) of an individual bp could be determined from the temperature dependence of K_{diss}. As an example, opening base pairs in the duplex regions within RNA was usually characterized by a large increase in both enthalpy and entropy.[18] Hence the H-D-monitored base-pair opening process presumably reflects the same hydrogen-bond-breakage and base unstacking process that was earlier seen with H-T exchange measurements at temperatures reasonably close to T_m (Figure 4 and schematized in Figure 5d). This process presumably involves a loss of favorable interaction energy with neighboring bases reflecting unstacking, and a gain of favorable entropy that likely arises, at least in part, from the increased conformational flexibility of the DNA backbones associated with the open state.

With regard to investigations of DNA breathing fluctuations, one major source of ambiguity that arises from using NMR signals from defined bps to measure the opening and closing dynamics for bps in short duplexes is that these short dsDNA molecules are likely to open and close as a part of the two-state opening and closing reaction of the whole duplex, rather than as local melting processes within single long dsDNA molecules. This will, of course, result in larger and faster melting fluctuations at the ends of the short duplexes (as observed; see Ref. 15) in contrast with the slower dynamics monitored with the central bps of these short complementary dsDNA oligomers. These long lifetimes for the central bps could reflect the rates of reformation of the entire duplex, and thus may not correspond to the reversible "breathing" of a local cooperative "opening" unit of the type monitored in long DNA molecules by hydrogen-tritium exchange.

Other Chemical Probes to Monitor Local Breathing Fluctuations that Involve Unstacked and Unhydrogen-Bonded DNA Bases

Figure 5d shows a cartoon of a possible elemental (local) dsDNA breathing state in which the inter-base W-C hydrogen bonds are broken and the bases are largely unstacked. This state represents the local, and fully reversible, version of the fully melted state of the DNA duplex at temperatures well above the T_m (Figure 5e). There have been, of course, many experimental demonstrations that the fully melted state exists, and that AT-rich segments melt before GC-rich segments in heterogeneous DNA. [See, for example, the electron microscopy pictures by

Inman[20] of partially melted and fixed dsDNA, which show "melted puddles" in the AT-rich sequences with intervening native dsDNA structures in the GC-rich regions.] It is likely that these states correspond to the high free-energy-of-activation breathing process that manifests itself in H-T exchange experiments at experimental temperatures within 20°C of T_m (Figure 4). Can this locally melted state be observed using chemical probes other than hydrogen exchange?

We began to ask such questions when Alberts and his colleagues isolated gene 32 protein (gp32), the T4-coded single-stranded DNA binding protein of the bacteriophage T4 DNA replication system. The early studies of Alberts and Nossal and their co-workers[21–23] showed clearly that gp32 bound much more strongly to ssDNA than dsDNA. Thus in principle (based on the laws of thermodynamics) gp32 should act as a "DNA melting protein," lowering the T_m of the dsDNA as the concentration of gp32 was increased and eventually pushing the apparent T_m down past the experimental temperature and thus stably coating the resulting ssDNA strands with protein. However, despite these well-founded thermodynamic expectations, such concentration-dependent gp32-driven dsDNA melting processes were not observed with most duplex DNA molecules, suggesting that the expected melting must be "kinetically blocked."

This suggested that the dsDNA breathing fluctuations available under the experimental conditions were not adequate to provide access for these melting proteins to the dsDNA "interior," and in kinetic terms that the transition state for the gp32 binding reaction was too high to be surmounted on reasonable time scales. We will return to this point in a subsequent section, but this observation led us to consider "simpler" melting protein models that might successfully probe breathing states such as that modeled in Figure 5d. As a consequence Jim McGhee in our laboratory undertook a detailed examination of the binding and melting properties of formaldehyde (HCHO) as a simple ssDNA binding ligand, leading eventually to a complete characterization of what might be termed the "world's simplest dsDNA melting protein model."

Formaldehyde as a Quantitative Chemical Probe of Duplex DNA Breathing Fluctuations

In a series of detailed articles McGhee and von Hippel[24–27] used UV absorbance spectroscopy and other types of analysis to determine the thermodynamic and kinetic parameters for HCHO-adduct formation with the amino and imino groups of the canonical DNA and RNA bases, first free in solution as monomers and then as components of ss- and dsDNA and polynucleotides. They used NMR to confirm that the adducts formed with these functional groups of the DNA and RNA bases were indeed hydroxymethyl (or methylol) derivatives, and showed that the equilibrium constants for formation of the monomethylol adduct with both the imino and amino groups of the nucleic acid bases ranged from ~2 to $12M^{-1}$ (K_1), while small amounts of the dimethylol derivative ($K_2 = 0.4M^{-1}$) formed at high formaldehyde concentrations with the amino groups as well. The overall reaction (for amino groups) is shown in Eq. (3).

$$\begin{aligned} \mathrm{R-N-H_2 + CH_2O} &\overset{K_1}{\rightleftarrows} \mathrm{R-N(H)} \\ \mathrm{-CH_2OH} &\overset{K_2}{\rightleftarrows} \mathrm{R-N-(CH_2OH)_2} \end{aligned} \tag{3}$$

It turned out that the equilibrium constants for the formation of both the imino and the amino formaldehyde adducts were of comparable magnitude, meaning that mono-adduct formation was significantly favored for both types of groups at equilibrium. On the other hand the rate constants for the two types of groups were very different, with the amino adducts forming and dissociating slowly ($k_{diss} = 10^{-5}$ to 10^{-6} s^{-1}) while the imino adducts formed and dissociated about 10^5-fold faster. Thus on reasonable time scales (minutes to hours) the imino reactions were easily reversible, while the reactions with amino groups could be considered to be essentially irreversible.

Formaldehyde turned out to be a useful chemical probe of local dsDNA conformation as well. While the rate constants and temperature dependences for the formation of methylol adducts for ssDNA or ssRNA were comparable to the rates with the free nucleotides, the corresponding rates of reaction with dsDNA in the W-C B-form conformation turned out to be very slow. Moreover, these rates became much faster as the temperature was increased towards the T_m of the nucleic acid duplex. The binding assays monitored absorbance changes in the DNA spectra to track HCHO binding. Typical UV absorbance and difference spectra, which were used to determine equilibrium and rate constants for the reactions are shown for the free mononucleotides 5′-dAMP and 5′-dCMP in Figure 6a. The analyses of these data to measure equilibrium binding as a function of formaldehyde concentration ([HCHO]) according to Eq. (4), are shown in Figure 6b, and the rate constants for mono-adduct formation with 5′-dAMP, 5′-dCMP, and 5′-dGMP as a function of [HCHO] are shown in Figure 6c.

$$\frac{\Delta A_\lambda}{[\mathrm{HCHO}]} = \frac{K_1 C\Delta\varepsilon_1 + K_1K_2C\Delta\varepsilon_2[\mathrm{HCHO}]}{1 + \mathrm{K_1[HCHO]} + K_1K_2[\mathrm{HCHO}]^2} \tag{4}$$

Here $\Delta\varepsilon_1$ and $\Delta\varepsilon_2$ represent, respectively, the difference in extinction coefficient (at wavelength λ) between the mono-adducts- (and the di-adduct) and the starting material. To estimate the four parameters (K_1, K_2, $\Delta\varepsilon_1$, and $\Delta\varepsilon_2$), we

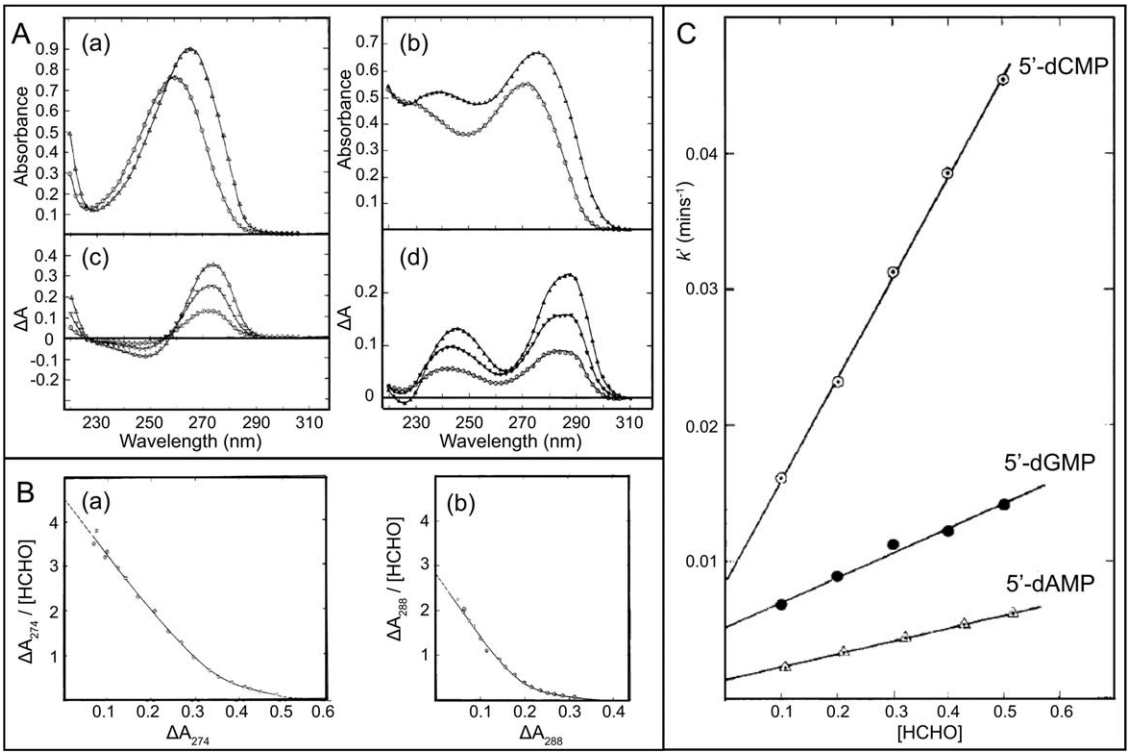


FIGURE 6 (A) (Left panels) (a) Spectrum of 5′-dAMP: (⊙) control; (△) with 1.05*M* formaldehyde. (c) Difference spectra of 5′-dAMP in the presence of added formaldehyde: (⊙) 0.05*M*; (▽) 0.21*M*; (△) 1.*05M*. (right panels) (b) Spectrum of 5′-dCMP: (⊙) control; (△) with 1.07*M* formaldehyde. (d) Difference spectra of 5′-dCMP in the presence of added formaldehyde: (⊙) 0.05*M*; (▽) 0.21*M*; (△) 1.07*M*. (B) (a) Equilibrium absorbance changes (274 nm) of 5′-dAMP with added formaldehyde plotted according to Ref. 25. (b) Equilibrium absorbance changes (288 nm) of 5′-dCMP with added formaldehyde. (C) Plot of pseudo-first-order rate constant, k', vs. formaldehyde concentration: (⊙) 5′-dCMP; (•) 5′-dGMP; (△) 5′-dAMP. From McGhee, J. D.; von Hippel, P. H., Biochemistry, 1975, 14, 1281-1296, © American Chemical Society, reproduced by permission.

extrapolated Eq. (4) to [HCHO] concentrations of either zero or infinity.

In contrast to these relatively "well-behaved" formaldehyde binding results obtained with the various free nucleotides and the very similar results obtained with single-stranded homopolynucleotides and natural ssDNA and RNA sequences,[26] the rate constants for adduct formation with duplex calf thymus and T7 DNAs as a function of temperature behaved quite differently. As shown in Figures 7a and 7b, the rates of adduct formation with DNA duplexes, monitored as the rates of formaldehyde-driven denaturation of duplex calf thymus and T7 DNA, respectively, proceeded both much more slowly than the rates of adduct formation with free nucleotides, ssDNA and polynucleotides, and were strongly dependent on the proximity of T_m. The latter finding, of course is reminiscent of the behavior of the H-T exchange rates at temperatures close to T_m (see Figure 4).

Based on many experiments of the types shown in Figures 7a and 7b, McGhee and von Hippel[24–27] were able to calculate, for DNAs of base compositions ranging from 0 to 100% AT, the relationship between the T_m values of the various DNAs and formaldehyde concentration, thus providing the information needed to use this reagent as a probe of breathing (stability) fluctuations. The results are shown as a phase diagram in Figure 8 for a solvent environment containing ∼0.01*M* NaCl, where the T_m value for pure AT-containing dsDNA is ∼50°C, and ∼90°C for pure GC-containing dsDNA.[26]

These results served to make formaldehyde an excellent and easily quantifiable chemical probe to study the breathing fluctuations of nucleic acid duplexes, because it had been shown that the stereochemistry of the reaction for forming the hydroxymethyl adducts with the imino and amino groups of DNA and RNA bases require an orthogonal approach of the

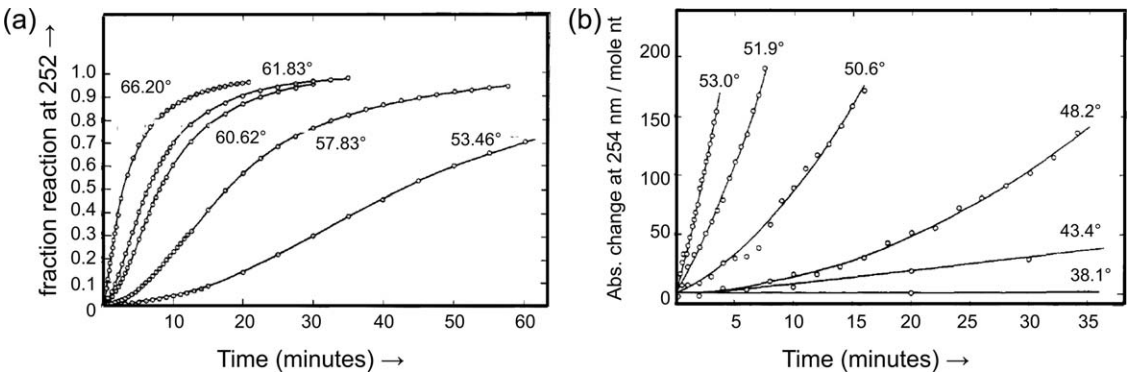
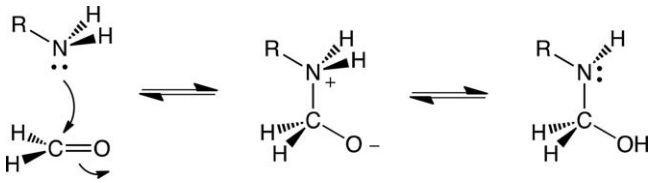


FIGURE 7 (a) Kinetics of denaturation of calf thymus DNA by 1.06*M* HCHO at various temperatures from 53 to 66°C, as marked. Reaction observed at A_{252} and normalized to overall absorbance change; as described in Ref. 27 this wavelength measures (for calf thymus DNA) an average degree of helix denaturation independent of the chemical reaction. Buffer was 0.02*M* phosphate pH 6.96. (b) Initial phase of denaturation of T7 DNA by 1.0*M* HCHO at temperatures ranging from 38 to 53°C. Reactions (in 0.002*M* phosphate, pH 7.1) observed at 254 nm and expressed as change in absorbance per mole of DNA nucleotide.

reacting HCHO to the surface of the base, as shown in the following schematic of the overall reaction mechanism.



This mechanism requires that the local bases be unstacked to permit the reaction to proceed. While the rate constants for adduct formation with bases in ssDNA and single-stranded polynucleotides were only slightly slower than those obtained with free nucleotides, presumably because the base stacking in these polymers is noncooperative and labile, the rigid and "cooperatively stacked" structure of DNA duplexes in the W-C conformation is only "breached" with difficulty at temperatures significantly below T_m (Figure 7b). In DNA duplexes, local breathing fluctuations resembling those shown in Figure 5d are presumably required for adduct formation to go forward.

We note that while formaldehyde adduct formation at moderate HCHO concentrations and at temperatures well below T_m proceeds very slowly, or is undetectable with B-form W-C base-paired dsDNA, this may not be true for base paired segments within dsDNA that exist in, or can "flip" into, alternate conformations. For example, HCHO attacks local dsDNA regions containing Hoogsteen base pairs very rapidly,[28] presumably because the interfaces in the DNA between W-C and Hoogsteen bp segments represent either actual openings or nucleate frequent breathing to afford ready access to the target amino and imino groups on the bases. This suggests that local structural discontinuities along the dsDNA contour may represent sites of altered or increased local "breathing" that could be exploited by regulatory proteins showing specificity for such special local structures or dynamics (for example, see Refs. 29 and 30).

These findings show that formaldehyde can serve as a generally useful, quantitative and simple chemical probe to map the stability (and thus the breathing fluctuations) of sequences of varying base composition in "natural" dsDNA molecules of heterogeneous sequence. The details are fully spelled out elsewhere,[26,27] but the following summarizes the outlines of the approach. At 25°C the destabilizing effect of formaldehyde is

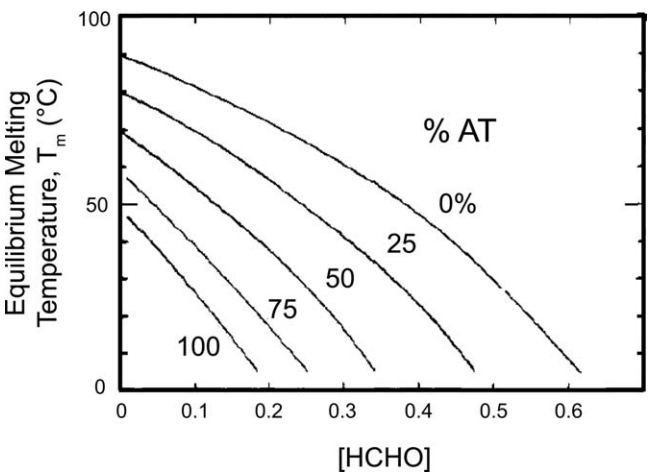


FIGURE 8 Calculated relation between DNA T_m and formaldehyde concentration, for DNAs of base composition ranging from 0 to 100% AT as marked on each curve. Calculations are described in Ref. 26 and correspond to a solvent containing approximately 0.01*M* Na^+.

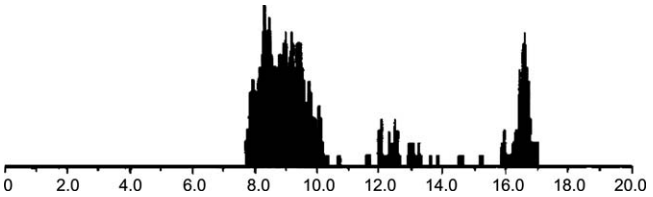


FIGURE 9 Histogram of 26 molecules of λDNA partially denatured by 1*M* HCHO at 58°C, in 0.02*M* phosphate, pH 6.95.[27]

roughly the same for a G-C pair as an A-T pair, which at 0.1*M* HCHO corresponds to an unfavorable (destabilizing) free energy contribution of ∼+0.57 kcal per A-T pair and ∼+0.59 kcal per G-C pair. This equivalence of A-T and G-C bps had long been assumed in experiments using formaldehyde as a denaturant to map DNA regions of differing base composition. However, because the equilibrium enthalpy change for the adduct formation with cytosine is somewhat greater than for the reaction with adenine,[25] formaldehyde somewhat preferentially destabilizes A-T regions of dsDNA at the higher temperatures generally used for "denaturation mapping."

However this is a fairly minor effect, and formaldehyde shows remarkably little base specificity, meaning that to a first approximation HCHO denaturation maps (an example is shown in Figure 9) can be used to quantitatively demonstrate the increased breathing (decreased stability) of AT-rich regions of duplex DNA under physiological solution conditions. Furthermore, because the formaldehyde reaction itself is very insensitive to salt concentration and only weakly dependent on temperature, the curves shown in Figure 8 can be easily applied to other salt concentrations and temperature conditions by simply shifting the T_m-axis of the graph. Chemical probing with formaldehyde is discussed further in subsequent sections of this article.

Breathing Processes that Could Lead to Intercalation of Planar Dyes and Aromatic Amino Acid Side-chains into Duplex DNA

Finally, in this survey of possible elemental breathing processes in DNA that can, in principle, be revealed by simple chemical probes, we turn to the cartoon structure shown in Figure 5b, which models a possible breathing process involving the local untwisting and unstacking of dsDNA without breaking the inter-base W-C hydrogen bonds. Of course this is a purely hypothetical and doubtless over-simplified model, but it makes some predictions and may form at least part of the transition state for the types of opening required for the intercalation of dyes and aromatic amino acid side-chains of proteins into the dsDNA interior.

In the mid-1960s Lerman and others[31–33] discovered that planar fluorescent dyes (e.g., acridine orange or ethidium bromide) could insert themselves between the base-pairs of dsDNA, resulting in a significantly extended duplex structure and greatly increased fluorescent intensity of the dyes, presumably as a consequence of removing the "surfaces of the dyes" from exposure to the aqueous solvent and thus protecting them from solvent-induced collisional quenching. Such intercalation processes have been much studied, and might well serve as models for how tryptophan and tyrosine residues of DNA-binding proteins, such as the TATA-binding protein that inserts tyr residues into duplex DNA and thus bends it at the promoter in transcription initiation, or polymerases that use tyr residues to block further extension of the RNA-DNA hybrid duplex in transcript elongation, might insert themselves into the relevant sites in nucleic acid duplexes.

Given that this type of opening involves base unstacking, one might predict that such intercalation processes—if such a breathing event comprises a rate-limiting step in what is likely to be a multistep binding and insertion process—should be characterized by significant temperature dependencies and perhaps also a dependence of the insertion rate on the proximity of the reaction temperature to the melting temperature of the local DNA. Some recent molecular dynamics simulations of such dye intercalation processes have suggested that at least local bp separation processes, perhaps in conjunction with local bending, may comprise portions of the transition state for dye (or planar aromatic antibiotic) insertion reactions into DNA.[34] In addition recent crystallographic studies of putative transition states for DNA repair proteins that must "flip-out" bases from B-form in order to check for misincorporated or damaged bases suggest that part of the mechanism of these enzymes involves stabilizing local configurations with unstacked bps on either side of the target bp prior to enzymatic removal of the damaged nucleotide.[35] It is tempting to speculate that local breathing of the type schematized in Figure 5b may provide at least the initial fluctuation that affords access to their DNA targets for DNA repair enzymes of this general type.

HOW MIGHT dsDNA "BREATHING" FLUCTUATIONS FACILITATE ACCESS OF PROTEINS INVOLVED IN GENE EXPRESSION TO THE DUPLEX DNA "INTERIOR"?

A little thought makes it clear that if "DNA melting proteins" (defined as proteins that bind more tightly to ss- than to dsDNA sequences) could reach the interior of DNA duplexes, then simply on a thermodynamic basis—that is, by increasing the protein concentration sufficiently so that dsDNA molecules would simply melt (at least locally) and end up as "melting-protein-coated" ssDNA strands—many of the regulatory

proteins that control the transactions of gene expression that we know about today would be unnecessary. That would include, of course, the carefully environmentally balanced NTP-dependent helicase (or DNA unwinding) activities of DNA replication complexes and RNA-polymerase-containing transcription complexes, as well as the helicase-containing complexes of DNA recombination and repair.

Access of formaldehyde to the amino and imino groups of bps involved in W-C base-pairing in the duplex DNA "interior," presumably via local breathing fluctuations such as those schematized in Figure 5d, permits this ligand to function as a model thermodynamic "melting" (T_m-depressing) protein as described in the previous section. Thus, as the HCHO concentration in solution is raised, the AT-rich regions of dsDNA molecules "melt" and form HCHO-adduct coated "puddles"—and if one goes still higher in formaldehyde concentration eventually the entire DNA duplex melts and forms HCHO-coated ssDNA strands. Why then cannot gp32, which coats ssDNA cooperatively with a ssDNA binding site size of seven nucleotide residues (nts) per gp32 subunit, act as a seven-nucleotide-residue-site "super-formaldehyde" and—at sufficiently high gp32 concentrations—melt dsDNA to equilibrium in the same way?

The answer—as might be expected from the theme of this Reflections article—seems to lie in the nature and magnitude of the breathing fluctuations of dsDNA. We have developed the argument on this point in detail elsewhere,[36] but in essence the small fluctuations that permit access to formaldehyde—which attacks one DNA base at a time—seem to be sufficiently available, at high enough HCHO concentrations and temperatures reasonably close to T_m, to permit formaldehyde binding, and thus local dsDNA melting, to proceed to an equilibrium that depends on the input HCHO concentration (see Figure 7). As shown in the denaturation map of the DNA genome of phage λ (Figure 9), at appropriate HCHO concentrations and temperatures and salt concentrations this results in the stable formaldehyde-stabilized melting of AT-rich regions—presumably reflecting the presence of significant local "opening" fluctuations in these regions, while leaving GC-rich regions largely intact. In contrast, seven bp fluctuations with appropriate geometry to provide access to an entire gp32 subunit are clearly not available at temperatures more than a few degrees below the T_m of natural DNA molecules.

Thus formaldehyde can melt natural dsDNA to equilibrium by "trapping" interior bases one-at-a-time as bps "breathe" in small cooperative units, while gp32 and other DNA "melting proteins," even at very high protein concentrations and at temperatures just below T_m, appear to be kinetically blocked from trapping such fluctuations. Thus we can speculate that it is the existence of these kinetic blocks, which depend on the limited sizes and shapes of the dsDNA fluctuations that occur as a function of bp-sequence and environmental conditions, that opened the door for the evolutionary development of NTP-driven nucleic acid helicases and their multiple, complex and highly regulated roles in opening the DNA genome and controlling its interactions with the cellular environment.

It is interesting to note (as first shown by Alberts and Frey[37]) that, in contrast to duplex sequences of "natural" dsDNA, gp32 is not kinetically blocked from binding to equilibrium the special duplex structure presented by long sequences of alternating poly(dA•dT). This presumably reflects the fact that these poly(dA•dT) sequences, because of the special transiently looped-out palindromic structures that they can form at temperatures near T_m, can—unlike natural dsDNA of heterogeneous sequence—expose ssDNA binding sites that are long enough to be stably trapped by gp32 subunits. These differences are shown schematically in Figure 10 and provide alterative fluctuation opportunities that we are currently trying to exploit and interpret by the kinetic spectroscopic methods described in the last section of this article.

Helicase Initiation and Unwinding Mechanisms, Duplex DNA Fluctuations, and Coupled Translocation and NTP Binding and Hydrolysis Cycles

Nucleic acid helicases tend to unwind dsDNA from the ends of dsDNA molecules or sequences or from a previously formed (Y-shaped) ss-dsDNA fork junction, and then move processively through the dsDNA sequence, opening the duplex with a repeating step size of one or more bps on the basis of a chemical reaction cycle that corresponds to the sequential binding and hydrolysis of single NTP molecules. Clearly the process of loading a helicase onto—and then into —the dsDNA at a specific initiation site is likely to involve very different kinds of structural breathing fluctuations than the processive NTP-driven advance into and through the duplex DNA that follows. Thus, whether it is the hexameric helicase-primase complex of the DNA replication system binding at and then "opening" a replication origin, or the transcription complex that must recognize a promoter and then facilitate the opening of the dsDNA at that site, the events of helicase initiation must be reasonably complex.

We speculate that the helicase "loading" process must require one or more opening fluctuations at the replication origin or the transcription promoter that may well be perturbed (or "focused") by the exterior binding of the replication or transcription complex, likely followed by one or more "directed" fluctuations that provide the initial access to the duplex interior. The processive elongation process that follows the initial loading likely involves a localized breathing event "ahead" of the helicase (or polymerase), which sequentially opens the dsDNA before

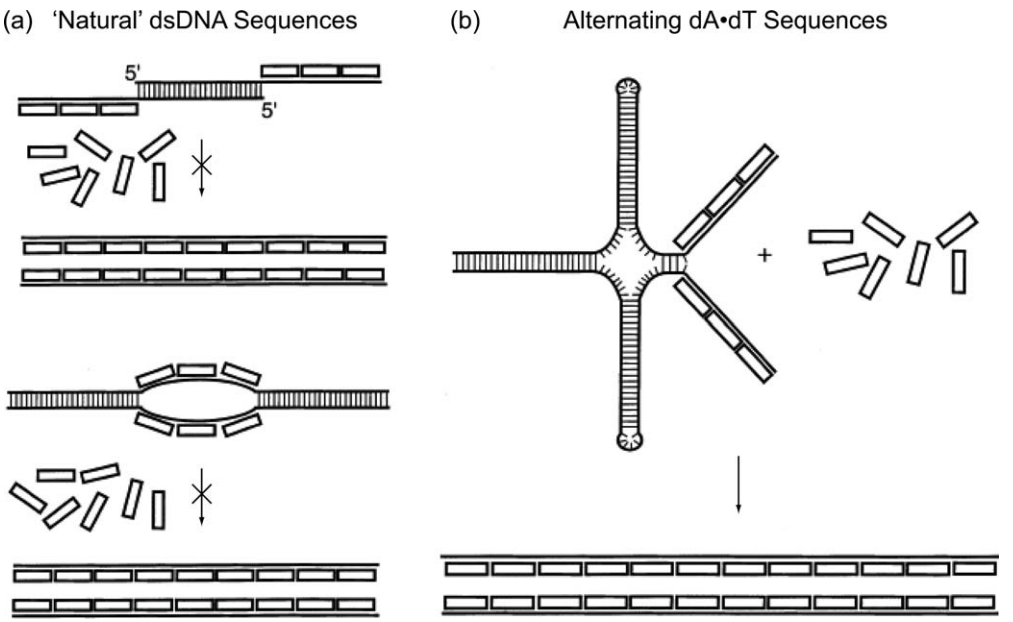


FIGURE 10 Melting activity of an excess of single-stranded DNA binding protein on dsDNAs of varying sequence. (a) "Natural" dsDNA sequences are kinetically blocked from binding single-stranded binding protein (here T4-coded gp32 with a binding site size [*n*] of 7 nt), even when free ssDNA ends (above) or internal (dA-dT-rich) internal loops (below) are precoated with the protein to nucleate the melting process under protein concentrations and temperature conditions at which melting process is thermodynamically favored. (b) In contrast, dsDNA consisting of alternating dA-dT sequences does melt to equilibrium under these conditions, presumably because this DNA can form transiently palindromic looped-out structures that expose ssDNA binding sites that are long enough to be trapped by gp32 at temperatures close to T_m (see Ref. 36).

each helicase translocation step as it moves the replication or transcription fork junction through the DNA duplex. The sequence of binding and destabilization events corresponding to the initial opening process at both the replication origins and at transcription promoters are complex and not well understood in terms of the details of either the conformational changes or the dynamics involved. The sequential elongation or translocation steps through the DNA duplex that follow are easier to think about and potentially describe, and will comprise the focus of the discussion of the coupling of fluctuations and cyclical NTP binding and hydrolysis events that form the main topics of the remaining sections of this article.

Processive NTP-Driven Helicase Unwinding Reactions can Circumvent the "Kinetic-Block" Problem

As pointed out above, NTP-dependent helicase-driven unwinding processes also are likely to take advantage of dsDNA fluctuations, but these fluctuations are quite different from those needed to insert entire gp32 subunits, with a binding site size of seven nts, into the DNA interior. As discussed in detail elsewhere[36] after the initiation process, which generally requires the participation of regulatory subunits in addition to those of the helicase "core," a helicase translocates into and through the duplex DNA by processively advancing a fork junction (or primer/template – p/t – junction) by means of alternating cycles of DNA binding and release that are driven by concomitant NTP binding and hydrolysis cycles.

Of course if the helicase is functioning in isolation it moves through and leaves behind an intact duplex (Figure 11a), marking its passage only by release of NTP hydrolysis products (NDP and P_i). In order to retain the open state after the helicase has passed through its binding and release cycle this process must be "coupled" with a concomitant binding reaction that leaves the ssDNA coated (and thus prevented from reannealing) with, for example, a single-stranded DNA binding (or melting) protein such as gp32 (see Figure 11b). Other coupling factors that also prevent reannealing of the parent DNA strands include the RNA synthesized by the elongating RNA polymerase "helicase" as it advances. In general helicases translocate along the DNA by one (the T4-coded primosome helicase or RNA polymerase) or more nts per NTP bound and hydrolyzed (Figure 11c). In addition we presume (and have shown elsewhere[38]) that the helicase movement is dependent on

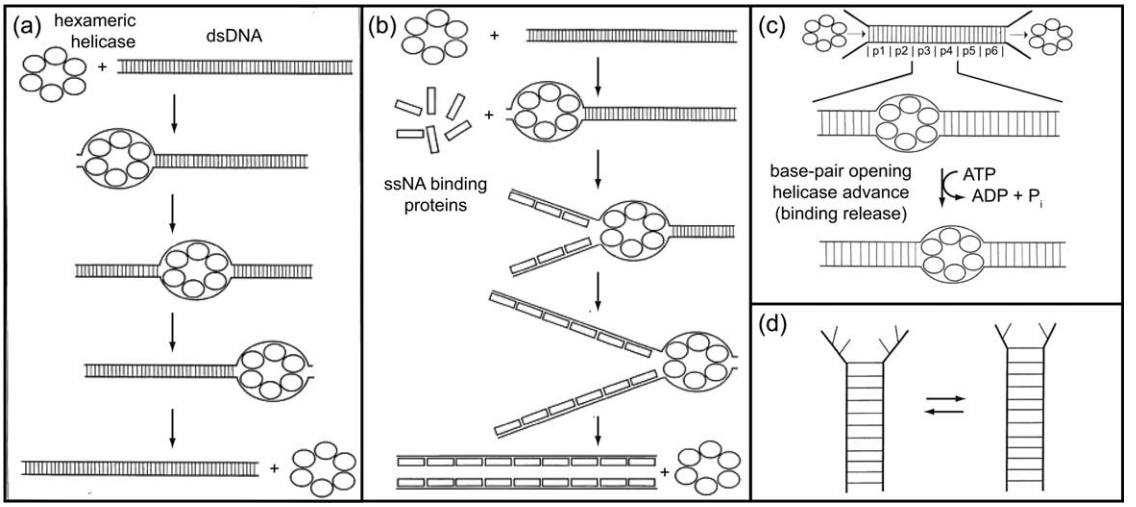


FIGURE 11 Uncoupled and coupled overall helicase reactions, the single-step helicase reaction cycle and the base pair opening process. (a) The advance of a typical hexameric helicase through a dsDNA segment in an uncoupled reaction. (b) The advance of a hexameric helicase through a dsDNA segment in a reaction coupled to the binding of single-stranded DNA binding protein. (c) Depiction of the "single-step" helicase reaction cycle resulting from the binding and hydrolysis of a single molecule of ATP. In this hypothetical example the helicase unwinds this 30 bp duplex in six single-step cycles. The positions reached at the end of each cycle are numbered p1 to p6. (d) The thermally driven sequential (one bp at a time) DNA unpairing ("fraying") process that is required for helicase advance.[36]

sequential " fraying" fluctuations of the dsDNA at the fork (Figure 11d).

The reason all this works, of course, is that by means of this coupled and NTP-driven process the helicase is able to advance by single steps preceded by fraying events that have activation barriers that are of the order of kT in height, which—unlike the barriers that must be circumvented to bind whole single-stranded binding protein subunits such as gp32—are easily accessible at room temperature.

Finally, it is important to comment on the role of NTP binding and hydrolysis in this process. It has often been conjectured that somehow the chemical free energy of hydrolysis of the NTP molecules involved in the helicase translocation process is used to destabilize the base pairs ahead of the helicase by compensating for the base-pairing free energy that is lost by the duplex DNA during the helicase-catalyzed DNA unwinding process. However this is clearly not necessary, and probably not true, because the favorable free energy gained by the system as a consequence of the binding of the helicase to the newly opened ssDNA sequences ahead of the fork is likely to significantly exceed the free energy lost by opening the next bp (or bps).

Therefore, in our view it is much more realistic to think of the helicase translocation and binding step within each reaction cycle as being thermodynamically "down-hill," with the only up-hill step being the release of the bound helicase from the sugar-phosphate backbone that initiates the next round of helicase translocation. As a consequence such processive helicases can be thought of as simple molecular motors that move along their binding strand and displace the complementary DNA strand by taking advantage of spontaneous thermal fluctuations that form "ahead" of them as they bind and release from the DNA backbone along which they move. This point is illustrated in Figure 12 for a "generic" helicase of the hexamer type that is central to the helicase complexes of DNA replication.

USING BASE ANALOGUE PROBES TO MONITOR DNA FLUCTUATIONS AT FORK JUNCTIONS IN THE ABSENCE AND PRESENCE OF THE T4-CODED REPLICATION HELICASE COMPLEX

We turn now to some recent experiments using spectroscopic methods with DNA replication fork constructs containing site-specifically placed fluorescent base analogues to actually monitor such single-step DNA breathing and helicase reaction cycles under steady state or equilibrium conditions. These base

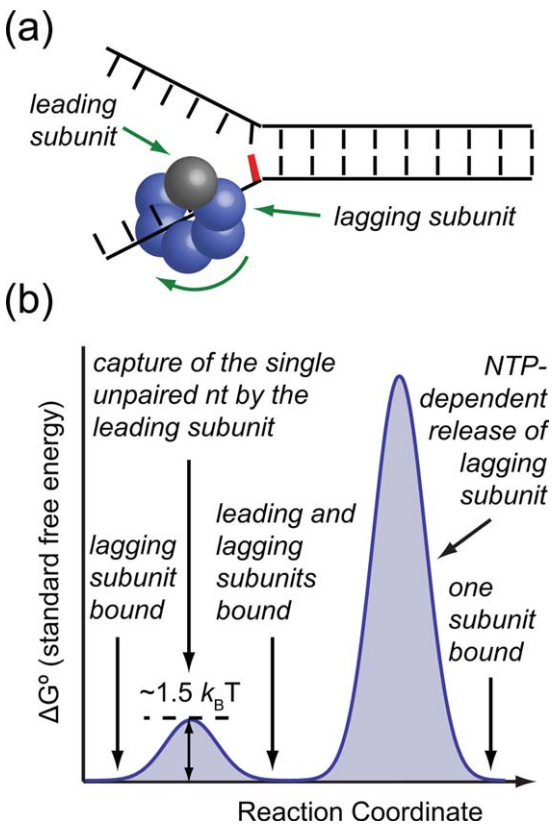


FIGURE 12 A model of the dsDNA unwinding process catalyzed by a hexameric helicase using a sequential single base pair opening and trapping process. (a) Here the unpairing reaction is driven by thermal fluctuations and the hexameric helicase uses the ssDNA binding site of its leading subunit (shaded gray) to capture and accumulate a single unpaired nt. Once a full binding site size of DNA has been accumulated (here in a tightly coupled helicase step devoid of slippage), the lagging helicase subunit (blue sphere) is released and relocated to participate in the next single-step helicase reaction cycle. The lagging helicase domain release and reorientation phase is driven by ATP hydrolysis. (b) Activation free energy diagram for the above single-step process. Note that the entire reaction coordinate shown corresponds to the reaction of only one helicase-binding domain.

analogues, unlike the canonical bases, display absorbance (and circular dichroism) spectra in the near UV ($\lambda >$ 300 nm), a spectral region in which the other protein and nucleic acid components of the "molecular machines" of DNA expression are transparent. We have used these approaches to demonstrate that local breathing fluctuations actually occur, and can be perturbed and utilized by a bacteriophage T4-coded replication helicase bound at the replication fork to translocate through dsDNA. In the next section we extend these studies by using more advanced spectroscopic techniques to make structural interpretations and to monitor these processes in real time as well. Variations of these approaches should make it possible for our labs and others to study and at least partially characterize the more complicated (and likely "protein-binding-perturbed") fluctuations that are doubtless involved in processes such in transcription and replication initiation and the other more complex mechanisms of genome expression.

Structures and Spectra of Base Analogue Probes

About 10 years ago we began to try to investigate conformational changes that might occur within functioning macromolecular machines by substituting specific DNA or RNA bases located at interesting positions within the nucleic acid framework of such complexes with base analogue probes. The base analogues we used for these purposes, and the chemical changes that they manifest in bps involved in W-C base-pairing equations, are shown in Figure 13, and their fluorescence and CD spectra within different single- and double-stranded nucleic acid structures are shown in Figures 14a and 14b.

Of course base analogue probes, especially single 2-AP base substitutions in DNA sequences, had been frequently used before as phenomenological monitors of local ds- to ssDNA conformational conversions, because it had long been recognized that such (often helicase or polymerase-driven) conversions seemed to result in significant enhancement of the 2-AP fluorescent intensity signal. However, our purpose was to attempt to develop these probes further as ways to monitor and characterize local conformational changes within functioning macromolecular machines in quantitative and perhaps even in structural terms. While it was clear that the fluorescence spectra of these systems varied significantly in intensity, though not in shape, in going from a dsDNA to an ssDNA environment (Figure 14a), we thought that there might also be useful—and perhaps more structurally detailed—information in the circular dichroism spectra of these probes in the near UV, and of course a quick inspection confirmed this to be the case (Figure 14b).[39–47]

Thus we set out to build a database of the spectral signatures of these probes as site-specific base substitutions within defined DNA (and RNA) structures as a first step in attempting to make some (initially largely phenomenological) interpretations of these spectra in structural terms.[39–41,44,47] One complication that quickly became apparent was that the details of the CD spectra of these base analogue probes appeared to depend, at least in part, on the sequence within which they were placed in heterogeneous ss- and dsDNA segments, and thus differed from one another even when placed as inserts within a common structure (for example, the B-form helix of dsDNA). However, because these sequence-dependent complications turned out to be largely nearest neighbor effects and seemed to depend in large part on how well the bases could "couple" electronically with the adjacent bases on the same strand, we found that we could largely avoid such nearest neighbor spectral heterogeneity by inserting the base analogues as "homo-dimer" pairs.

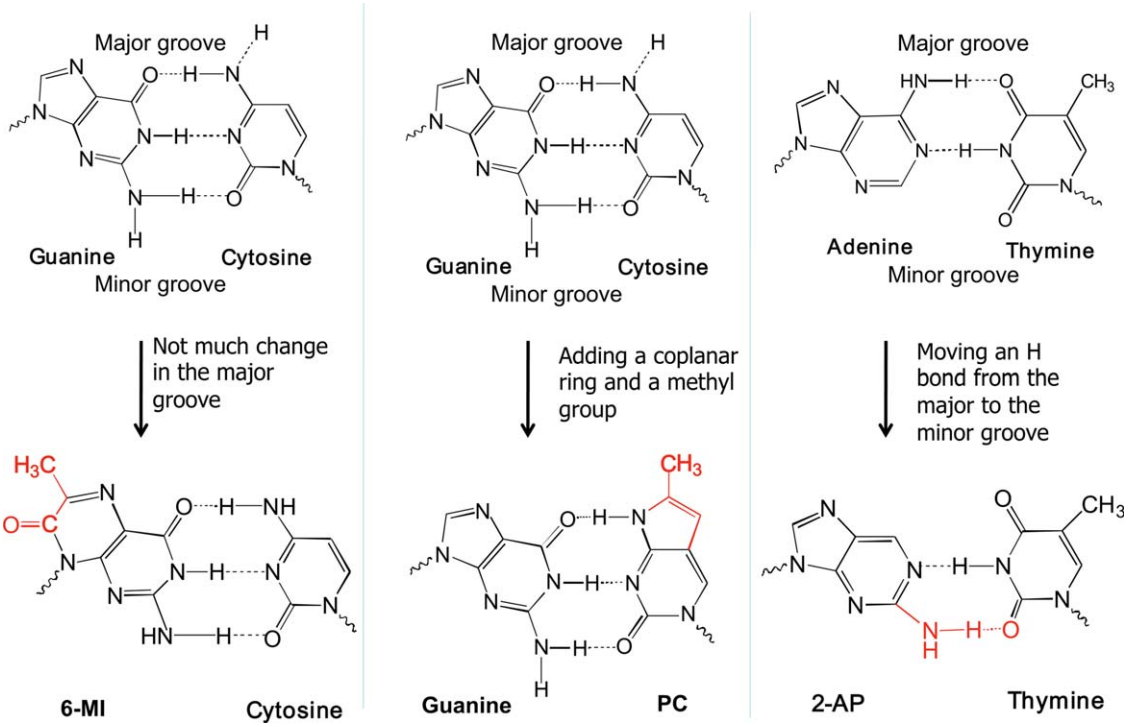


FIGURE 13 Structures of the base analogues 6-MI-C, PC-G, 2-AP-T, and canonical base pairs. Modifications are in red.

Although we continued to use monomer base analogues as phenomenological probes in many applications, we also found that dimer pairs of all three of the probes shown in Figure 13 resulted in interpretable exciton-coupled CD spectra—especially in duplex B-form DNA (see Figure 14C)—while in other conformations (and in DNA-RNA hybrids)—the spectra were clearly and reproducibly different. Furthermore it turned out that monitoring both fluorescent and CD spectra produced powerfully complementary information, with the intensity of the fluorescence being significantly "self-quenched" in stacked base analogue pairs, while in contrast the CD signals were significantly augmented in the ordered and stacked "dimer-probe" conformations.

An important feature of these approaches—seen particularly in Figure 14b—is that while the base analogues fit well into the cooperative conformational structures of the DNA or RNA (or RNA-DNA hybrid) duplexes, as manifested by the unchanged nature of the canonical CD spectra at wavelengths $<$300 nm (left panel of Figure 14b), these protein-nucleic acid complexes were totally transparent at wavelengths $>$300 nm (right panel of Fig. 14b) and so one could focus on exactly what was happening, in a conformational sense, as the protein components of the "macromolecular machine" of interest moved through a defined (and presumably biologically significant) position in the nucleic acid framework of the nucleic acid-protein complex. For example, we were able to use limited nucleotide sets to "walk" DNA or RNA polymerases from one defined position in a nucleic acid construct to another, and thus determine the spectral changes that occurred as these proteins moved into and through known DNA or RNA terminators, pause sites, hairpins, etc.[42,43,46] By using, in particular, the two-dimensional fluorescence spectra (2DFS) approach described in the following section, we have recently been able to start to realize the "spectroscopist's dream" of making actual structural interpretations of such spectra. In contrast, in our earlier work we were largely restricted to using the database of spectral shapes as indicators of the presence or absence of specific local conformations.

DNA Breathing Fluctuations at Replication Fork or Primer/Template (p/t) Junctions

While these results cast considerable light on mechanistic issues of interest in DNA-dependent DNA replication and RNA transcription complexes, it did not escape our attention that we should also be able to learn something from these approaches about DNA breathing fluctuations, and perhaps how these fluctuations might be perturbed (or utilized) by nucleic acid helicases. The details, at the equilibrium and steady state "breathing" level, are presented elsewhere.[38,45,48] Here we briefly summarize our findings and conclusions.

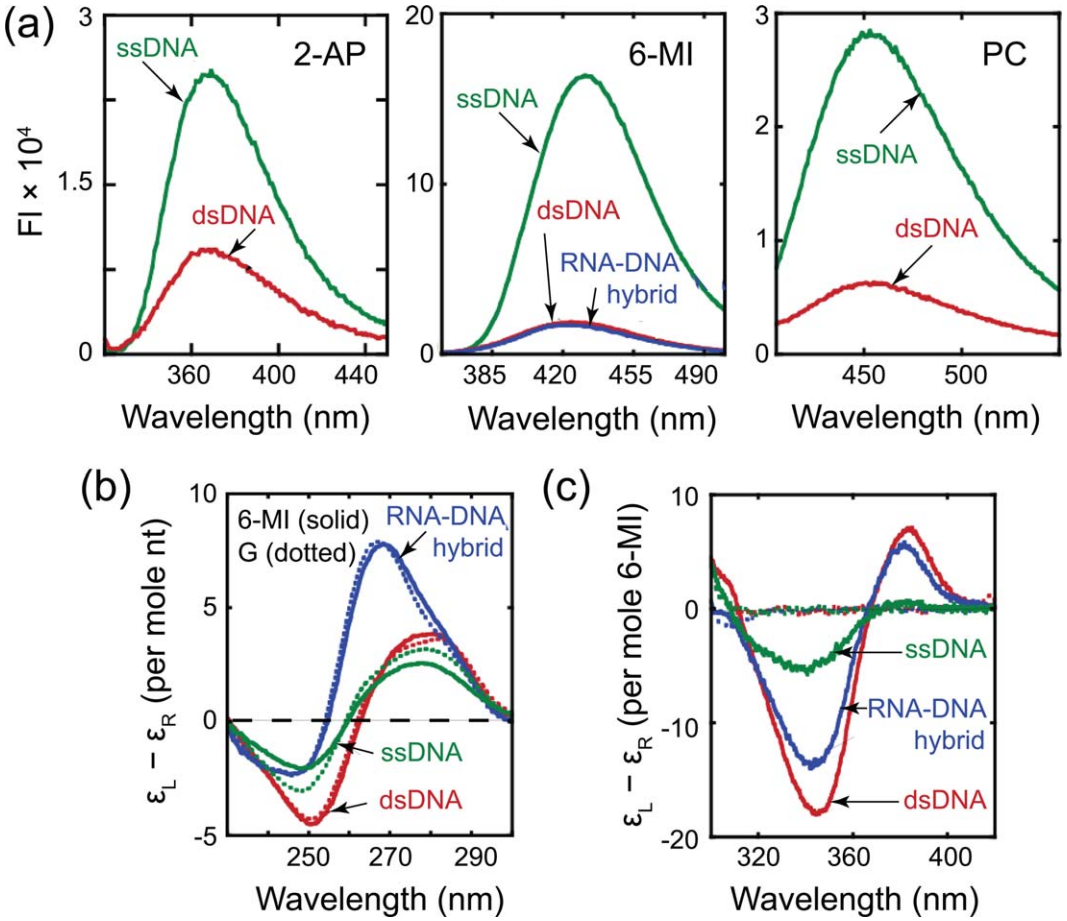


FIGURE 14 Spectroscopic properties of DNA containing different base analog dimer probes. (a) Quenching of dimer probes in dsDNA relative to ssDNA. CD spectra of 6-MI dimer probes (b) below 300 nm and (c) above 300 nm.

We started by using our base-analogue probe approach with 2-AP replacing adenine nucleotides at a series of defined positions in a DNA replication fork model of defined sequence, such as shown schematically in Figure 15b. By monitoring either the relative intensities at the fluorescence peak (Figure 15a—entire fluorescent spectra are shown in Figure 14a) or the CD spectra at $\lambda > 300$ nm (Figure 15b) at defined positions on both the dsDNA and the ssDNA sides of the fork junction, we were able to come to a number of conclusions about the steady state breathing properties of such fork constructs. The detailed results and interpretations are presented elsewhere.[38,45]

First, we were able to conclude from numerous examples of the data shown in Figure 15 that the first bp—position (+1)—in the dsDNA section of the DNA replication fork constructs used was ~50% "breathed open" at 25°C, even at 25 to 30°C below T_m. This seemed to be true whether the first duplex bp was A-T [containing an adenine (2-AP) analogue] or G-C [containing a guanine (6-MI) analogue].[47] Given that G-C pairs are significantly more stable at interior positions of DNA duplexes, this argued that the stability of the first bp at the fork is determined more by its structural location than by its bp type. The bp at position (+2) in our fork constructs was also measurably breathed open (~20%) at 25°C, while at position (+3) the bp appeared to be only ~1 to 2% open. These percentages also appear to shift towards further "openness" as the DNA fork melts (see Table I; DNA fork construct position numbers are defined in Fig. 15). At positions (+4) and beyond the bps appeared, in spectroscopic terms, to be fully closed at 25°C. In contrast, the bases on the non-complementary (ssDNA) side of the fork junction were largely, although not fully unstacked as measured by the fluorescence and CD spectral criteria, presumably reflecting the partial and noncooperative stacking seen, particularly for purine residues, in ssDNA.

Such "fraying" at the fork junction was expected, of course, both from "end-effect" considerations and from NMR-detected H-D exchange rates on short DNA duplexes, but an unexpected observation that may be seen in both the fluorescence and the CD spectra of Figure 15 is that the first base in

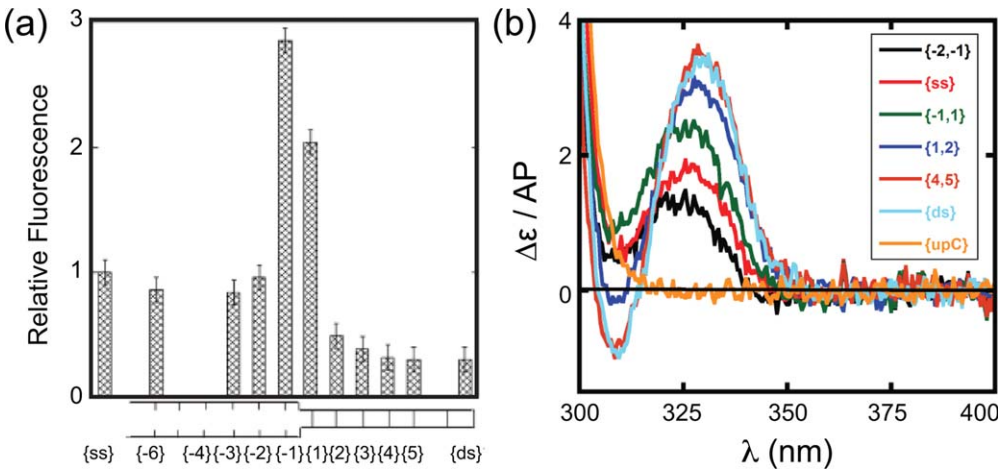


FIGURE 15 Spectroscopic properties of forked DNA constructs as a function of position relative to the ssDNA–dsDNA junction. (a) Fluorescence intensity changes at 370 nm for constructs containing single 2-AP probes. (b) CD spectra for constructs containing 2-AP dimer probes at the indicated positions. (Spectrum labeled upC corresponds to dsDNA control construct containing no 2-AP probes).[45]

the ssDNA sequences on both strands of the fork (position −1) appears to be much less stacked with its neighbors than an average ssDNA base. This seems to be general and may be significant because it could offer an explanation for why helicases, which bind preferentially to ssDNA sequences relative to dsDNA sequences, might preferentially bind at the fork junction rather than elsewhere on the ssDNA of the fork, presumably because the favorable free energy of binding directly at the fork is increased by the fact that the helicase need not expend as much free energy of unstacking of the bases at this position as it presumably has to expend elsewhere on the ssDNA strands.

The above results apply to various free DNA fork constructs for which we studied the degree to which single and paired base analogue probes were "open" (defined as the extent to which they were unstacked from their neighbors on the same DNA strand). Of course the purpose of these measurements in the present context of DNA fluctuations and how they might be utilized (and also perturbed) by DNA replication helicases moving into the duplex part of the forked DNA constructs was to attempt to measure (here under equilibrium or steady state conditions) the effects of the presence of the helicase at the fork on these fluctuations. In a separate study[48] we had shown that the bacteriophage T4-coded replication primosome helicase is formed from six gp41 helicase subunits complexed with six intersubunit NTPs, which assemble into a stable hexagonal helicase structure that binds weakly to, and can slowly and non-processively unwind, a DNA model replication fork construct. [We note that it is this initial (equilibrating) helicase complex that we were able to use to determine time correlation functions for the effects of transient helicase binding on the fluctuations at the replication fork, as described in the next section.] We were also able to show, in that same study, that this weakly binding helicase complex can be further assembled into the stable and processively unwinding helicase complex that functions in vivo by adding a single gp61 primase subunit to form the fully active T4 primosome complex.

In order to use these base analogue probe methods to determine the effects of primosome helicase binding on the structure and steady state breathing fluctuations at the replication fork, we needed to find a way to "freeze" the helicase-fork construct after the helicase had moved into the duplex region of the fork by one step in order to afford time to make the necessary fluorescence and circular dichroism measurements. It turned out—because of the arguments in the previous section that showed a single step of the helicase unwinding process (up to the free energy-requiring subunit release step shown in

Table I Percentage of Double-Stranded Character at Various Positions on Forked DNA Constructs at Different Temperatures from CD Melting Curves

Temperature (°C)	{−1,1}, %	{1,2}, %	{2,3}, %	{4,5}, %
25	48 ± 3	80 ± 3	100	100
37	46 ± 3	73 ± 3	92 ± 2	92 ± 2
55	32 ± 4	42 ± 3	51 ± 2	51 ± 3
65	14 ± 3	14 ± 4	15 ± 4	14 ± 3

Figure 12) could be taken by the primosome without NTP hydrolysis—that we could assemble the primosome helicase complex using nonhydrolyzable GTPγS as the bound ligand instead of a standard NTP moiety, and thus arrest the primsome helicase complex after it had, effectively, taken one step into the fork (for details see Ref. [38]). We were able to show that this single step was on the helicase unwinding pathway by completing the duplex unwinding process (after spectroscopic characterization of the frozen "initiation-intermediate" complex) by swamping the system with GTP to exchange out the nonhydrolyzable GTPγS.[38,49]

Analysis of this "frozen intermediate" showed that the replication fork had, indeed, unwound by one step, that the steady-state breathing pattern described above for the fork in the absence of helicase had "moved in" by one step, and that essentially the first bp (in fork postion +1) was now fully "locked open," the second bp (position +2) was now ~50% open, the third bp (+3) was now ~20% open and the 4th bp (+4) was essentially unperturbed. Thus, in effect, the breathing pattern seen with the unperturbed DNA fork had essentially moved by one bp into the fork duplex. This is shown schematically in the context of a proposed primosome helicase mechanism in Figure 16. Additional experiments[38] showed that the binding of the helicase to the construct at each step must involve binding interactions of the sugar-phosphate backbones of both strands of the fork with at least two (gp41) subunits of the helicase hexamer and the "mobile" primase subunit. Collisional acrylamide quenching measurements further showed that the backbones of the two strands were bound asymmetrically (the lagging strand presumably passing through the hole in the central of the hexagon to hold the fork open and the leading strand binding to the outer surfaces of 2 or 3 of the gp41 subunits of the hexamer—see Figure 16). Finally fluorescence and CD measurements suggested that the "breathing bases" located at the fork were essentially unperturbed (in terms of steady-state stacking) by the transient helicase binding.

In the next section we report further studies in which we have been able to use kinetic single molecule methods to study the breathing dynamics of the DNA replication fork in the absence and in the presence of both the weak-binding $(gp41)_6$ helicase complex and the tight-binding primosome helicase complex in real time.

NEW APPROACHES TO THE STUDY OF DNA LOCAL STRUCTURE AND BREATHING KINETICS

In the last section of this retrospective and prospective overview of DNA breathing we return to the issue of the approaches that might now be available to observe (conceptually, structurally and dynamically) breathing fluctuations in dsDNA and how these fluctuations might be used and manipulated by regulatory proteins to access and interact with the coding elements of the interior of the duplex DNA genome. As shown in Figure 5, above, in order to attempt to provide a first-order description of the breathing structures in DNA that might be accessed by the simple chemical probes that were initially used in raising these questions in the 1960s and 70s (hydrogen exchange, formaldehyde binding, dye intercalation) we resorted to simple conceptual cartoons. To now go further with the powerful optical and spectroscopic tools available in this "modern" era, more sophisticated conceptual descriptions and diagrams of the possible structures of breathing fluctuations are clearly needed, and so we turn here to some preliminary descriptions in terms of intersecting free energy surfaces (FES) and landscapes. These lead us in turn, in the final portion of this discussion of local dsDNA fluctuations and their interactions with regulatory and functional proteins and protein complexes, into a description of the type of experiments we are now attempting to use to "flesh out" these diagrams with real molecular structures visualized both in real time and in real space.

Free Energy Surfaces and Landscapes

An important conceptual tool often used to visualize fluctuations in proteins, nucleic acids and other macromolecular systems is the free energy surface (FES) or landscape diagram. The FES provides a Boltzmann distribution of conformational states that depends on the relative thermodynamic stabilities (ΔG°) of the conformations involved. Taken in this context, we may view breathing as a distortion of the local nucleic acid conformation at a particular site, with the likelihood of occurrence depending on the curvature of the relevant FES that is plotted as a function of free energy (Gibbs free energy for biological systems) versus a relevant configurational coordinate.

We consider the free energy surfaces experienced by a locally labeled site within a DNA construct, and the relationships between these surfaces (or landscapes) and DNA breathing. A hypothetical FES for a labeled base at a position within the duplex region of a DNA construct is shown in Figure 17a. The FES is a manifold of conformational states whose energies are plotted versus the generalized local coordinate Q, which may be, for example, the separation or orientation of the labeled base relative to a neighboring base on the same strand. The free energy minimum corresponds to the most stable conformation of the labeled site, which for dsDNA in aqueous solution is the B-form helix. As the local environment of the labeled site undergoes breathing fluctuations, such as base unstacking or Watson-Crick hydrogen bond breaking, the local

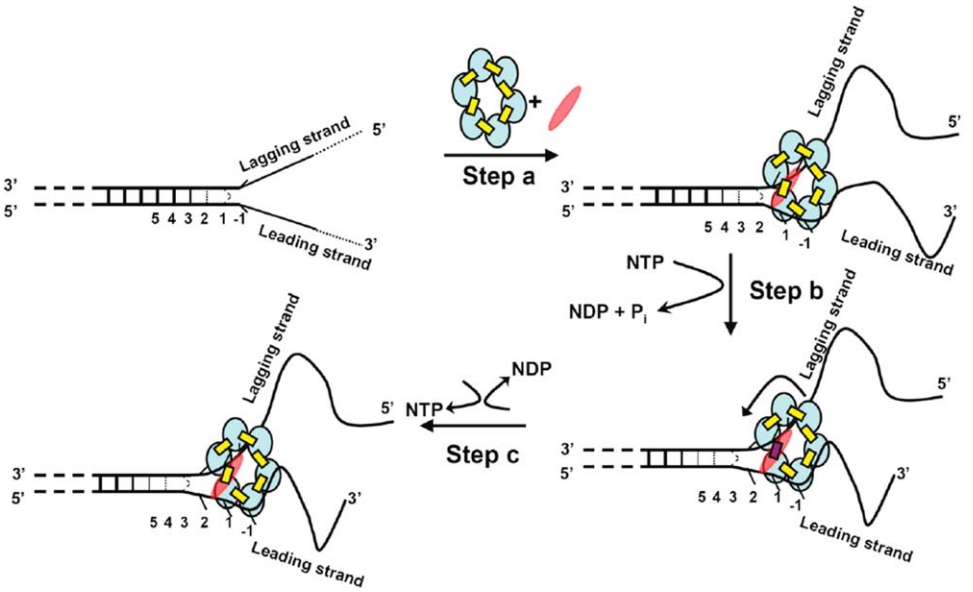


FIGURE 16 Proposed unwinding mechanism of the T4 primosome helicase. The constituents of the primosome are gp41 subunits (blue ellipses), gp61 subunit (red ellipses), GTP (yellow rectangles), and GDP (red rectangle). The "degree of openness" of the base pairs adjacent to the ss–dsDNA junction is indicated in each panel, and the numbers below each DNA construct represent the numbering of the various base pairs prior to initial helicase binding. Step a: The GTP-bound gp41 hexameric helicase loads onto the free DNA fork construct and the gp61 primase subunit binds and stabilizes the complex at the fork junction. Positioning is facilitated by the uniquely unstacked conformation of the −1 bases. As a result of this initial binding the first duplex (breathing) base pair at position 1 is fully unwound and the breathing of the base pairs at initial positions 2, 3, and 4 are all enhanced. Step b: GTP hydrolysis occurs at the gp41–gp41 interface positioned adjacent to the bound gp61 subunit, destabilizing that subunit interface and permitting the primosome to "capture" the now unwound first base pair of the original duplex. This base pair becomes the new −1 position, thereby moving the breathing properties of each base pair at the fork one position further into the duplex sequence. The gp41 hexamer rotates by one subunit (approximately 60°) and the primase translocates to the next gp41–gp41 interface. Step c: The GDP (and Pi) hydrolysis products formed in step B dissociate, a new GTP binds and stabilizes the previously destabilized gp41–gp41 subunit interface, and the primosome helicase is ready to begin a new unwinding-rotation-hydrolysis cycle.

conformation will deviate from the minimum of the basin, corresponding to an increase in the free energy. If a sufficiently large distortion of the local base conformation occurs, such an event can serve as a bridge (or transition) to a different FES with a minimum energy state corresponding to a locally disordered conformation, or "breathing bubble." The intersection between the two FESs defines the "open conformation" transition state, as shown in Figure 17a. Because a free energy barrier separates the two basins of stability, the breathing bubble conformation may persist for a relatively long time (e.g., microseconds) before reverting back to the more stable B-form conformation.

The FES is a useful concept to describe DNA melting, which can be viewed in the context of a first-order phase transition.[50] At room temperature the activation energy required to nucleate a bubble may be as small as $\sim 1 k_B T$. The effect of increasing temperature is to stabilize the bubble conformation relative to that of the stacked and hydrogen-bonded B-form, as shown in Figure 17b. When the experimental temperature is close to the overall melting temperature of the DNA duplex, multiple bubbles can nucleate cooperatively at AT-rich regions in natural (heterogeneous sequence) DNA, and their relative population is increased. For a first-order transition in a short homogeneous dsDNA oligomer, which melts primarily from the ends in a two-state transition, there should be coexistence between fully separated single strands and duplex DNA at the melting temperature because both the disordered and B-form base-paired conformations have equal stabilities. Above the melting transition, the bubble (or eventually the strand-separated conformation) is thermodynamically favored.

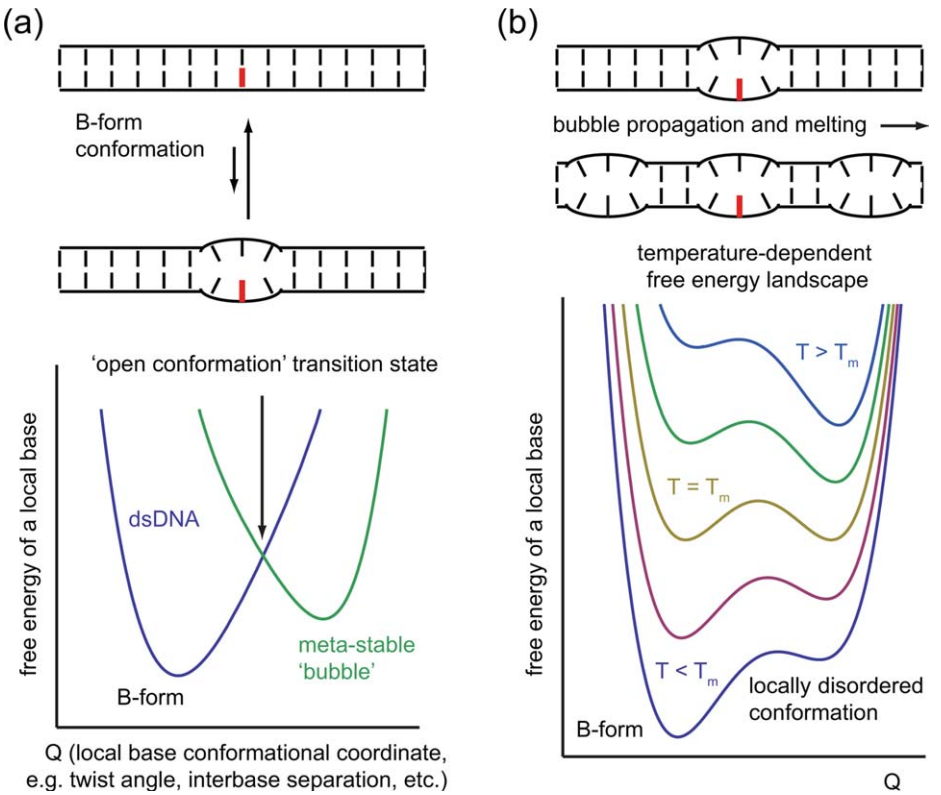


FIGURE 17 Free energy surface (FES) depictions of DNA breathing fluctuations and the equilibria between open and closed states. (a) A labeled base site (shown in red) in duplex DNA experiences distortion of its local environment, which corresponds to an increase of the free energy. Nucleation of a locally disordered "bubble" conformation can be described in terms of the intersection of two distinct free energy surfaces. (b) At temperatures below the melting transition, the ordered B-form conformation is favored, while disordered conformations become progressively more stable as the temperature is raised.

Similar diagrams can be drawn at temperatures near T_m for longer dsDNAs of heterogeneous sequence, but here the situation is different because at T_m the duplex will be significantly melted (into local bubbles of varying size, depending on temperature and base composition and sequence) in AT-rich regions, while still being largely structured in GC-rich regions. Thus long duplex DNA molecules undergo multi-state melting transitions, and these transitions will be reversible, even at temperatures significantly above T_m, because the strands are held together by GC-rich sequences. As a consequence, the strands of the duplex will not separate and the base pairs within local bubbles remain largely "in register." Clearly under these conditions fluctuations involving bubbles of local melting will be favored in AT-rich sequences, even at temperatures significantly below T_m, and these heterogeneous bubbles will form and collapse locally as dictated by the relevant landscape surfaces while being stabilized against total melting by the vicinal GC-rich segments. This situation is shown schematically in Figure 18.

Using Intersecting Free Energy Surfaces to Understand DNA-Ligand and DNA-Protein Interactions

We next consider how such spontaneous, thermally activated breathing fluctuations might facilitate DNA-ligand interactions, DNA-protein complex formation, and higher order macromolecular assembly. Similar to bubble nucleation, a thermally activated fluctuation that distorts the local base environment could represent a transition state to a weakly associated DNA-ligand complex, as illustrated in Figure 19a. The relative stabilities of the free B-form duplex and that of the DNA-ligand complex in the figure indicate that in this example the equilibrium favors the unbound form.

A similar approach can be applied to a local site in proximity to a ss-dsDNA fork or a primer-template (p/t) junction (see Figure 19b). DNA forks and junctions are the primary targets for replication, transcription and repair proteins that enact and regulate gene expression. Thermal activation of a local site in

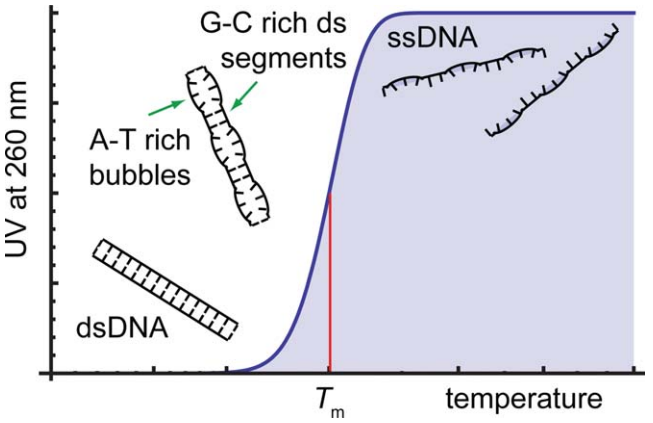


FIGURE 18 Schematic representation for the melting curve of long heterogeneous duplex DNA, as monitored by UV absorbance. The cartoons show the initial formation of open AT-rich "bubbles" as the temperature approaches T_m.

the duplex DNA close to a DNA replication fork can lead either to the nucleation of a metastable bubble, or else merely to a further (transient) melting from the fork into the duplex region. The relatively long lifetimes (~microseconds) of such locally disordered regions increases the likelihood that a nearby protein will encounter these transient structures in an open state. Upon formation of a DNA-protein complex, the labeled site will experience a new FES (shown in blue) with a minimum that is altogether different, and likely more stable, than the most favored conformation in the absence of the protein. In this way, the DNA-protein association reaction can be viewed as a "trapping" event, in which a thermally activated metastable and local DNA conformation can serve as a bridge (or transition state) to the formation of a more stable protein-nucleic acid complex.

The significance of the labeled site position in relation to the DNA fork or p/t junction is similar to that for a fixed site in the duplex as the temperature is raised and the local "melting" (order-disorder) transition is approached from below. For a labeled site within the duplex region, the FES is expected to appear as shown at the bottom of Figure 19c (dark blue trace). The site may undergo a transition from its most stable B-form conformation to a metastable "open" conformation. As the labeled site position is varied across the fork junction, its local environment gradually transitions from that of the duplex to that of the single-strand. For site positions very close to the fork junction, the relative stabilities of the B-form and the disordered conformation are nearly equal. For site positions beyond the junction, within the single-stranded region, the disordered conformation becomes more stable.

An interesting question pertains to the nature of the FES for site positions at the ss-dsDNA junction. In Figure 19 we have depicted this situation in terms of iso-energetically ordered and disordered conformational states separated by a barrier, similar to that described by a first-order phase transition at T_m. The height of this barrier likely depends on several variables, such as salt concentration, pH, and local base composition and sequence. Nevertheless, there is currently little information available about the structural and kinetic properties of these transition states, and the mechanisms by which they might facilitate protein and ligand binding interactions.

Development of New Methods for Studying Free Energy Surfaces on Reaction Pathways for DNA-Protein Interactions

DNA breathing fluctuations, and their involvement in complex formation of defined dsDNA sites with ligands and proteins attempting to access the duplex DNA interior, are characterized by the FESs that govern the time evolution of the system as a whole. Reaching a deep mechanistic understanding of these processes will require not only information about the structures of end states and transition states and their relative stabilities and barriers heights, but also information about the distribution of states and inter-state couplings. What new methods exist that can help us to attain such information, and how can these methods be deployed?

As described in the first major section of this article, most of what we have learned to date about the breathing fluctuations of duplex DNA has come from chemical probe studies, with probes ranging from hydrogen isotope exchange techniques to measure the rates of opening of peptide hydrogen bonds within proteins and inter-base (W-C) hydrogen bonds in nucleic acid duplexes under a variety of conditions, to studies of the rates and extents of base-pair unstacking within nucleic acid duplexes by measuring reaction rates with intercalating dyes and/or adduct formation with ligands (such as formaldehyde) that require unstacked intermediate states to achieve binding. Although these approaches—especially with NMR to identify the peptide hydrogen bonds being observed—have led to rather sophisticated insights into the folding and unfolding landscapes of globular proteins, our understanding using these techniques has led only to rather simplistic conclusions (not much advanced from the ideas of the 1960s summarized in Figure 5) for dsDNA and its breathing modes.

Some additional insight—at least conceptually (see earlier section on the role of DNA breathing in protein binding) and experimentally for specific DNA sequences of biological interest (see previous section)—has come from using the proteins and protein complexes that drive genome expression themselves as probes of DNA structure. Most of these approaches have been optical in nature, and have used site-specifically

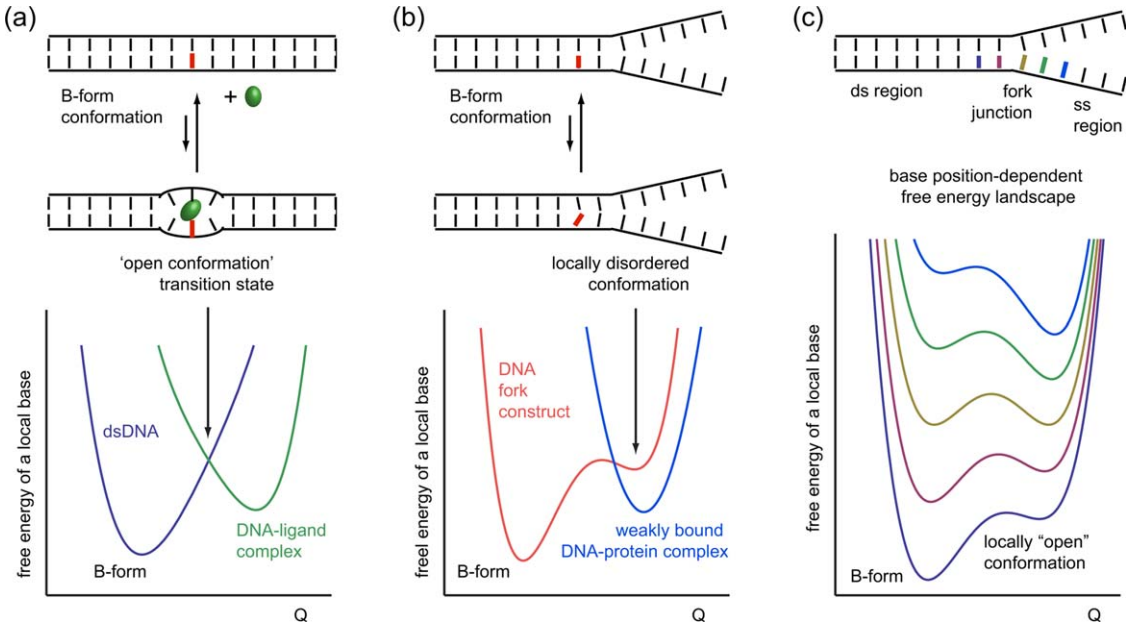


FIGURE 19 FES diagrams for DNA "breathing" and DNA-ligand (DNA-protein) complex formation. (a) Ligand association can be described as a process facilitated by a local conformational fluctuation experienced by a labeled base site (shown in red). A thermally activated "open conformation" can form a portion of the transition state connecting the stable duplex state and the DNA-ligand complex. (b) A conformational transition from B-form to a disordered state at a local site close to a ss-dsDNA fork junction can be a bridge to the formation of a DNA-protein complex. The presence of the bound protein can stabilize or "trap" the disordered state, as indicated by the locations of the minima of the corresponding FESs. (c) As the labeled site is varied across the boundary separating the duplex from the single-stranded region, the relative stabilities of the native B-form conformation and the disordered "open" conformation interchange, similar to the melting phase-boundary behavior depicted in Figure 17c.

placed DNA and RNA base analogues in bulk solution experiments to obtain some steady state and equilibrium information about nucleic acid breathing and to determine how such breathing might be exploited by specific proteins. Single molecule (sm) force experiments (e.g. magnetic and optical tweezers),[51,52] sm fluorescence imaging,[53,54] dynamic light scattering, and fluorescence correlation spectroscopy experiments[55] have also provided insight, as has theory and computer simulation[56,57]. NMR has also been useful in measuring millisecond relaxations at specific base sites within well-defined duplex DNA oligomer sequences at ~1 m*M* specific base-pair concentrations.[16,18]

The relatively high sensitivity of fluorescence-based methods opens the possibility of examining fluorophore-labeled macromolecular complexes at physiological concentrations (sub-μM) and provides opportunities to probe breathing in these systems on the tens-of-microsecond time scales—a difficult experimental regime to access. Here we outline two fluorescence-based approaches that our groups have co-developed to study structural and kinetic aspects of dsDNA breathing: (i) two-dimensional fluorescence spectroscopy (2DFS); and (ii) simultaneous single molecule Förster resonance energy transfer (smFRET)/fluorescence-detected linear dichroism (smFLD). Both methods make use of fluorescent chromophores that label pre-selected sites within model DNA constructs. Such approaches can use either nucleic acid base analogues to site-specifically label a local DNA (or RNA) base position, or internally (rigidly) incorporated cyanine dyes to label specific positions along the sugar-phosphate backbones.

Determining Local Nucleic Acid Conformations Using Two-Dimensional Fluorescence Spectroscopy (2DFS)

The local conformations of two adjacently substituted fluorescent nucleic acid base analogues can be studied using 2DFS. The 2DFS is a spectroscopic approach that can measure the interactions between electronic transitions in closely spaced molecular chromophores in much the same way that 2D-NMR can determine the spatial relationships between nuclear

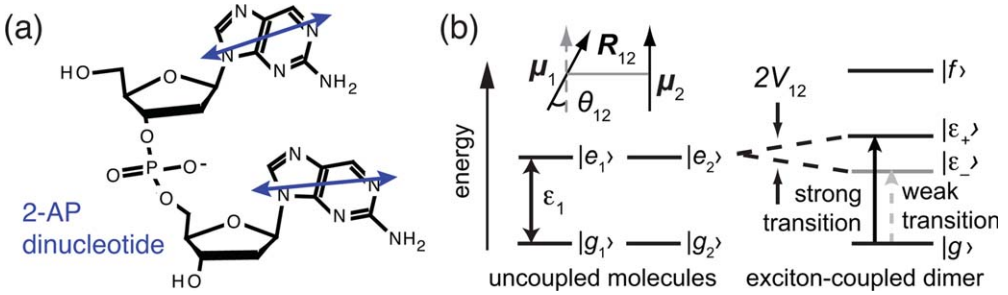


FIGURE 20 Electronic coupling within base analogue dimer probes. (a) 2-aminopurine (2-AP) is a fluorescent analogue of adenine. The electronic transition dipole moments of the 2-AP base residues (shown as blue double-headed arrows) in the dinucleotide can couple producing the shared exciton levels shown in panel (b). The inset in panel (b) shows as an example the configuration of a "side-by-side" dipole configuration with twist angle θ_{12} and base separation R_{12}. Such "base-stacked" configurations lead to blue-shifted absorbance and quenched fluorescence peaks. From Widom, J. R., N P. Johnson, P. H. von Hippel, A. H. Marcus, New J Phys, 2013, 15, 025029-025016, © Institute of Physics Pub, reproduced by permission.

spins in molecules.[58] When two base analogues are substituted for pairs of adjacent native bases on a single strand within a dsDNA construct, the resulting exciton-interactions depend sensitively on the relative separation and orientation of the probes. The electronic coupling between probe residues creates shared electronic states of these "dimer probes" and alters the absorption, circular dichroism and fluorescence spectra displayed by these chromophores in their uncoupled states.

An electronic energy level diagram of the 2-aminopurine (2-AP) monomer states, and how they are split due to electronic coupling in the 2-AP dinucleotide, is shown schematically in Figure 20. Four delocalized electronic states (excitons) are supported by the dimer: a ground state (labeled $|g\rangle$), two nondegenerate singly excited states (labeled $|\varepsilon_{\pm}\rangle$), and a doubly version excited state (labeled $|f\rangle$). A side-by-side arrangement of transition dipole moments, as expected for a fully stacked duplex, leads primarily to a blue-shifted absorption spectrum, self-quenching of fluorescence, and a non-zero CD signal.[39,59] While absorption, CD and fluorescence quenching techniques can monitor changes in local base stacking, they cannot by themselves determine the precise spatial arrangement of the probe bases. The structural elucidation of electronically coupled molecular dimers can be accomplished using the 2DFS method.[59–63]

The 2DFS method uses a sequence of four femtosecond optical pulses to resonantly excite the electronic transitions of the coupled probe chromophores (see Figure 21a). This pulse sequence excites four successive electric dipole-induced transitions among the four electronic states, and ultimately produces nonlinear excited state populations that can be detected via the ensuing fluorescence. A series of measurements are carried out for systematically varied inter-pulse delays and optical phases, producing a modulated signal that oscillates at the various resonance frequencies of the system. Fourier transformation of the time-domain spectra yields a two-dimensional optical spectrum that contains detailed structural information that can uniquely specify the spatial arrangement of the probe bases.[59–63]

Figure 21b shows experimental 2DFS data obtained from the 2-AP dinucleotide in solution.[59] Previous experiments on this system using CD had suggested that the 2-AP probe residues were at least partially unstacked.[39] However, the results of a conformational analysis of the 2DFS data, which are also consistent with the earlier CD results, show that the 2-AP dinucleotide conformation (with a relative twist angle of ~5° and an inter-base separation of ~3.5Å) is significantly more stacked than such a dimer-probe would be in the B-form conformation, where the twist angle would be 36° and the inter-base separation ~3.4Å. These results suggest that the base-base version of the hydrophobic bonding forces that favor base stacking in this dinucleotide exceed the local "strain energy" that this fully stacked conformation induces in the connecting sugar-phosphate linkage. On the other hand, in longer and cooperatively structured B-form DNA duplexes, this stacking is presumably reduced to still permit some stacking while avoiding longer-range sugar-phosphate backbone distortion and reducing intramolecular electrostatic repulsion by separating the partially charged backbone phosphates as much as possible, thus leading to the formation of the long B-form duplexes characteristic of the W-C structure. Obtaining similar information from 2DFS data for fluorescent base analogue dinucleotides substituted into various positions within double-stranded and fork DNA constructs of biological interest should provide further insights into the landscape surfaces that govern DNA-

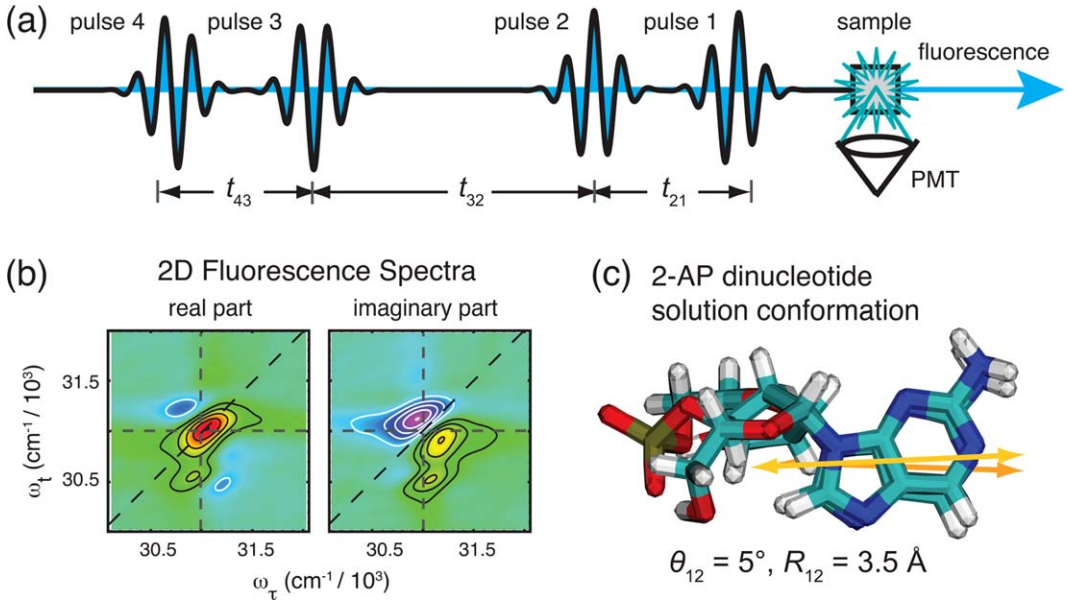


FIGURE 21 The principles of two-dimensional fluorescence spectroscopy (2DFS) measurements on fluorescent dinucleotide base analogues. (a) A pulse sequence used to determine the conformation of an exciton-coupled molecular dimer in a 2DFS experiment. The experiment detects fluorescence from electronic populations excited by the four-pulse sequence. (b) An experimental two-dimensional fluorescence spectrum (both real and imaginary parts) is shown, which was obtained from a 5 μM solution of 2-AP dinucleotide in aqueous buffer. The two-dimensional spectra exhibit diagonal peaks and off-diagonal cross-peaks that indicate the exciton-split states of the dimer. (c) Analysis of the 2DFS data revealed that the 2-AP dinucleotide is fully stacked in solution. Adapted, with permission, from Ref. 59.

protein interactions and the assembly pathways of DNA-protein complexes within the extended DNA genome.

Measuring Site-Specific Dynamics and Free Energy Landscapes of DNA-Protein Complexes by Simultaneous Single Molecule Förster Resonance Energy Transfer (smFRET) and Fluorescence-Detected Linear Dichroism (smFLD) Measurements

Structural methods such as 2DFS can sensitively determine the local conformations at site-specifically labeled positions within nucleic acid-protein complexes, and by careful manipulation of experimental conditions they can reveal information about the structural nature of end-states, and possibly transition-states, in protein-nucleic acid binding interactions. Experiments that measure the proportions of different conformational states in a mixture can determine their relative energies.[62] In any case experiments that probe the kinetics of interconversion between end-states are needed to understand the nature of the transition states and their associated barrier heights. Single molecule fluorescence experiments are especially well suited to probe the real-time kinetics of DNA breathing because such studies can, at least in principle, directly sample the conformational fluctuations governed by the local free energy landscapes.

In order to probe microsecond-breathing fluctuations of a site-specifically labeled nucleic acid construct, we developed single molecule (sm) experiments on DNA replication fork constructs that had been "internally" labeled with the Förster resonance energy transfer (FRET) donor-acceptor iCy3 and iCy5 chromophore pairs.[49] These cyanine dyes can be rigidly incorporated into the sugar-phosphate backbone using phosphoramidite chemistry, such that the fluorophores replace an opposed DNA base in each strand at the labeled positions (see Figure 22). The fluorophores can be positioned close to a replication fork junction or deep within the duplex DNA region. In these experiments we employ a polarized-excitation scheme that simultaneously monitors, on sub-millisecond time scales, the smFRET between an iCy3 (donor)—iCy5 (acceptor) chromophore pair and the fluorescence-detected linear dichroism (smFLD) of the donor chromophore.[64]

These measurements are based on the idea that a single oriented DNA replication fork construct, which is chemically tethered to a microscope slide surface, will absorb light with different probabilities depending on the polarization of the exciting laser field. By rapidly modulating the laser polarization

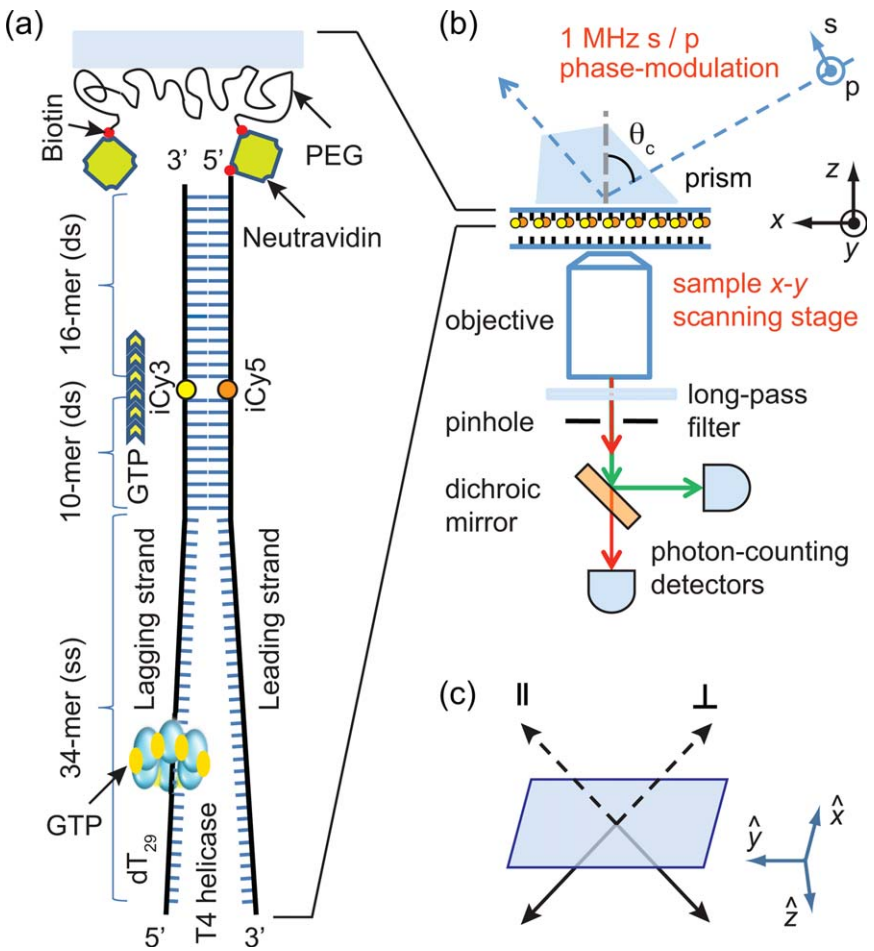


FIGURE 22 Molecular and instrumental setups for monitoring simultaneous smFRET and smFLD signals. (a) The T4 gp41 helicase binds to the $d(T)_{29}$ loading sequence on the lagging strand of a model DNA replication fork construct. An assembled $(gp41)_6$·gp61·DNA primosome complex can unwind the duplex region of the DNA in the presence of GTP. The strands within the dsDNA region are internally labeled with the FRET donor-acceptor iCy3 and iCy5 chromophores, respectively. (b) TIRF excitation scheme and detection method. The polarization of the excitation beam is modulated at 1 MHz. The p-polarization component points in the direction of the *y*-axis, and the s-polarization component is contained within the *x*-*z* plane. (c) Orthogonally polarized directions used to measure the FLD signal, from the perspective of the incident beam. Adapted, with permission, from Ref. 64.

and recording the phase of the modulation for every detected signal photon, we can monitor the real-time projection of the absorption transition dipole moment onto arbitrarily defined laboratory axes. Thus, while the smFRET signal is sensitive to the relative separation and orientation of the donor-acceptor chromophore labels, the smFLD signal is sensitive to the orientation of the donor chromophore within the laboratory frame. As a consequence the two signals can, in principle, provide different and complementary structural and dynamic information.

We have recently begun to apply this approach to the study of DNA breathing and its role in the binding and unwinding mechanisms of the bacteriophage T4 helicase, the equilibrium and steady-state binding and assembly properties of which have been described in the previous section. As stated in that discussion, the bacteriophage T4 is the simplest organism that uses essentially the same sub-assembly complexes to achieve DNA replication as do higher organisms. These sub-assembly components include a pair of DNA replication polymerases with processivities regulated by sliding clamps that are loaded (and unloaded) onto the DNA-bound polymerases by a five subunit ATP-driven clamp-loading complex. The sub-assemblies also include a hexameric helicase-primase (primosome) sub-assembly that unwinds the DNA base pairs ahead

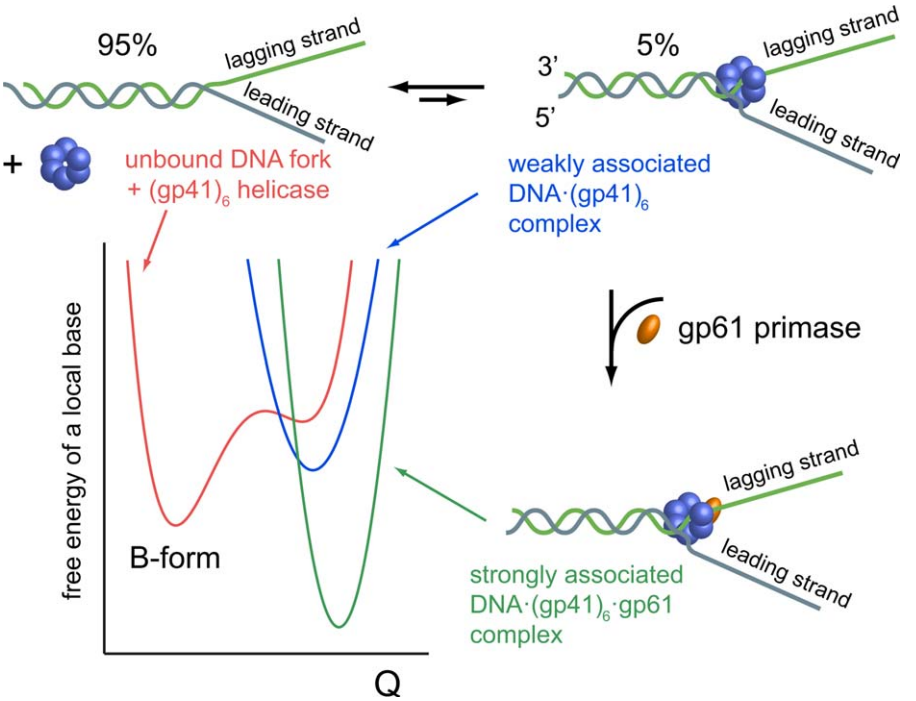


FIGURE 23 Free energy landscapes that describe the weak and strong binding of the T4-coded helicase complexes to the replication fork junction. Hexameric gp41 (blue spheres), is bound together by inter-subunit NTP ligands (not shown), and binds weakly to the ss-ds junction of a replication fork, favoring disassociation from the DNA. Introduction of gp61 primase, in a 6:1 gp41 to gp61 subunit ratio, causes the T4 primosome helicase to bind strongly to a replication fork. A hypothetical free energy landscape describing the sugar-phosphate backbone fluctuations near the fork junction is shown representing each of the three conformational states: unbound DNA fork substrate (red), weakly bound DNA·$(gp41)_6$ helicase complex (blue), and strongly-bound DNA·$(gp41)_6$·gp61 primosome complex.

of the polymerases at the replication fork, and also uses a separate subunit to catalyze the synthesis of the RNA primers required for re-initiation of lagging strand synthesis at the 3′-ends of newly formed Okazaki fragments.[65–67] In order for the gp41 helicase to recognize and to interact with the replication fork, it is necessary that the coding bases near the fork junction be transiently exposed from behind the sugar-phosphate backbones that shield them from the solvent environment.

As illustrated in Figure 23, the T4 gp41 hexameric helicase binds weakly to the replication fork junction, and is more likely to dissociate from the fork rather than to engage in processive dsDNA unwinding. In contrast, the T4 primosome helicase, which is composed of a gp41 hexamer and a gp61 monomer, binds tightly at the replication fork junction and manifests very processive unwinding activity in the presence of NTP. The breathing of bps located in the immediate vicinity of a replication fork or a p/t junction is much more extensive than that of bps located deep within a dsDNA region. A hypothetical free energy landscape describing bp dynamics near the fork junction is portrayed in Figure 23. In this model, bps can thermally fluctuate into the disordered state, which is separated from the B-form DNA structure by a potential barrier. Helicase association at the fork junction, and subsequent assembly of the fully functional primosome complex, is represented as a series of successive crossover events between distinct FESs.

By observing the fluctuations of dsDNA constructs within regions that have been labeled with fluorescent dyes, it should be possible to obtain information about the free energy landscapes to which the various DNA bases are subject, and also to determine how those landscapes are affected by the presence of the T4 helicase. In Figure 24, we show histograms of the relaxation times (τ_c) obtained from smFRET data using DNA fork construct labeled either deep in the duplex region (blue) or at the replication fork junction (red). Both types of DNA constructs exhibit backbone fluctuations on time scales ranging over hundreds-of-microseconds. These data indicate that in the absence of unwinding proteins, the closing rate of a spontaneously formed "bubble" in the duplex region is faster than

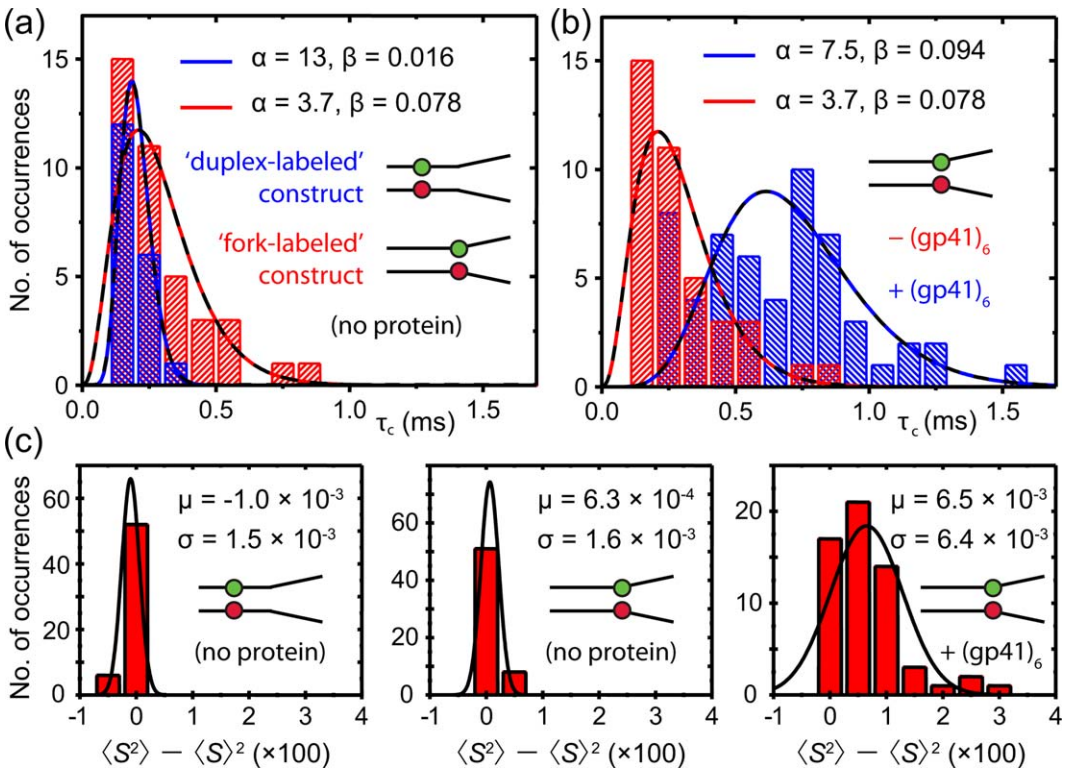


FIGURE 24 Dynamics of breathing fluctuations at the replication fork in the absence and presence of helicase. (a) Histograms of relaxation times (τ_c) obtained from the analysis of smFRET/smFLD trajectories for DNA fork constructs, which were labeled with iCy3/iCy5 placed deep in a duplex region (blue) or at the replication fork junction (red), in the absence of DNA unwinding proteins. (b) A comparison is shown for a fork labeled construct in the absence (red) and the presence (blue) of 300 n*M* gp41 and 6 μM GTPγS. (c) A comparison is shown of histograms of the relative magnitudes $<S^2> - <S^2>$ of the fluctuating smFLD$_r$ signal for duplex labeled DNA (left panel), fork labeled DNA (middle), and fork labeled DNA + (gp41 · GTPγS)$_6$ (right). Histograms of the relaxation times were characterized using the Gamma distribution function, with skewedness and width parameters α (dimensionless) and β (in units of ms), respectively. Adapted, with permission, from Ref. 64.

that of a similarly formed unstable conformation near the ss-dsDNA fork junction. Nevertheless, both the time scale and the amplitude of breathing fluctuations at sites near the fork can be dramatically lengthened by the intermittent presence of the gp41 helicase hexamer bound weakly to the replication fork junction.

Our observation that the presence of the helicase greatly alters both the distributions of the lifetime and the magnitude of disordered conformations at the fork junction implies that the presence of the helicase at the fork perturbs the breathing fluctuations at that position and suggests that these thermally populated minority states of the DNA substrate may well play a role in the helicase binding and activity mechanisms. The significance of this early result for our present purposes lies in the context of understanding the energetic factors that contribute to the functional assembly of biological macromolecular machines. An open question is whether the initial steps of helicase-primase (primosome) assembly occur at the replication fork junction through a cooperative mechanism in which the binding of the gp41 hexameric helicase fundamentally alters the nature of DNA fluctuations at the fork junction and thereby facilitates subsequent binding of the gp61 primase. The alternative interpretation, of course, is that DNA breathing and helicase binding do not influence one another, in which case the formation of the primosome-DNA complex must be considered to occur by a random coincidence of events.

While the latter scenario clearly can describe the chemical kinetics of small molecular systems, which are generally characterized by "stationary" free energy landscapes, our current results appear to support the more complex possibility in which a highly mutable time-evolving energy landscape can respond to instantaneously changing conditions. The studies

in this area on which we are currently embarking using the methods outlined in this section of this overview article should begin to provide specific answers to such questions for defined macromolecular complexes engaged in biologically-significant manipulations of the DNA genome.

CONCLUDING THOUGHTS ABOUT DNA BREATHING FLUCTUATIONS AND THEIR RELEVANCE TO DNA-PROTEIN INTERACTION MECHANISMS

In this article we started by reviewing old experiments that we and others had used to try to understand DNA breathing fluctuations by examining the rates with which various simple chemical probes interacted with functional groups buried inside the ground-state structure of duplex DNA. We demonstrated that these probes, for most of the conditions examined, reacted with their target sites in a Kk_3–dependent reaction (chemical probes section), in which the rate constant for opening the cooperative breathing unit exceeded the rate constant of the subsequent chemical reaction with the open state. As a consequence the observed reaction rates could be simply modeled as product of the intrinsic reaction rate constant (k_3) and the equilibrium fraction of the time that the cooperative breathing unit was open (K). By varying the solution conditions (probe concentrations, temperature, salt concentration, pH) we were able to separate solution changes that altered the intrinsic reaction/exchange rate from those that perturbed the fraction open, and as a consequence were able to define simple elementary models (Figure 5) that were consistent with our knowledge of the structure of dsDNA and the competing thermodynamic interactions that control its stability.

The dominant breathing states at low temperature (modeled in Figure 5c) seemed to be accessible only to hydrogen exchange probes and were characterized by low, essentially temperature independent, transition state barriers. It seemed (and seems) unlikely that these fluctuations could provide a pathway for proteins to access the DNA interior. On the other hand we should consider that these ubiquitous "underlying" breathing modes might well "nucleate" others at higher temperatures that could provide access to proteins, such as the local (unstacked, unhydrogen-bonded) fluctuations schematized in Figure 5d. Other possible fluctuations that might provide openings for intercalating groups were schematized in Figure 5b.

It is important to remember that most proteins that eventually bind to functional groups on the dsDNA interior (for example RNA polymerases or DNA helicases) first recognize their ultimate interior target loci by specific binding interactions with the duplex DNA exterior, and only then transition into the duplex interior via one or more subsequent steps. Since such specific exterior binding might be expected to *stabilize* the local sequence against opening reactions, we must further speculate that in these cases the presence of the bound proteins themselves might help to "shape" or nucleate otherwise less likely fluctuations that then help to provide appropriate access to the duplex interior. Furthermore the transition of the protein or protein complex from an exterior to the interior bound state must involve assembly processes or conformational rearrangements whose rates are limited by diffusion. Thus it seems likely that breathing fluctuations of the sort that might afford entry to the duplex interior must be characterized by relaxation times in the microsecond time range. This gives us some confidence that the fluctuations we are likely to be able to explore with the dynamic spectroscopy methods described in the previous section could be biologically relevant.

In closing it seems clear that, despite the 50 year history of DNA breathing fluctuation studies, we are still just "scratching the surface" in beginning to understand the fluctuational processes that must also be coded into the base pair sequences of the DNA genome and guide the interactions of the macromolecular machines of gene expression with their DNA targets. Clearly we, and our scientific descendants, should still find much of interest to pursue in this area in the next 50 years.

The authors acknowledge the detailed research studies carried out by the many skillful and creative graduate students and postdoctoral fellows whose work is summarized in this "Reflections" article, as well as the 50 years of discussions with laboratory and faculty colleagues whose collective insights and thinking were instrumental in shaping the ideas summarized here. P.H.v.H. also acknowledges specifically the influence of countless discussions with John Schellman over the years that contributed so much to his ongoing scientific education.

REFERENCES

1. Watson, J. D.; Crick, F. H. Nature 1953, 171, 737-738.
2. Lehman, I. R.; Bessman, M. J.; Simms, E. S.; Kornberg, A. J Biol Chem 1958, 233, 163-170.
3. Weiss, S., Gladstone, L. J. Am. Chem. Soc. 1959, 81, 4118–4119.
4. Benson, E. E.; Linderstrom-Lang, K. Biochim Biophys Acta 1959, 32, 579-581.
5. Linderstrom-Lang, K. Biochim Biophysi Acta 1955, 18, 308.
6. Hvidt, A.; Linderstrom-Lang, K. Biochim Biophys Acta 1954, 14, 574-575.
7. Kalman, S. M.; Linderstrom-Lang, K.; Ottesen, M.; Richards, F. M. Biochim Biophys Acta 1955, 16, 297-299.
8. Englander, S. W. Biochemistry 1963, 2, 798-807.
9. Printz, M. P.; von Hippel, P. H. Biochemistry 1968, 7, 3194-3206.
10. Printz, M. P.; von Hippel, P. H. Proc Natl Acad Sci USA 1965, 53, 363-370.
11. McConnell, B.; von Hippel, P. H. J Mol Biol 1970, 50, 317-332.

12. McConnell, B.; von Hippel, P. H. J Mol Biol 1970, 50, 297-316.
13. Teitelbaum, H.; Englander, S. W. J Mol Biol 1975, 92, 79-92.
14. Teitelbaum, H.; Englander, S. W. J Mol Biol 1975, 92, 55-78.
15. Gueron, M.; Kochoyan, M.; Leroy, J. L. Nature 1987, 328, 89-92.
16. Gueron, M.; Leroy, J. L. Methods Enzymol 1995, 261, 383-413.
17. Hu, W.; Walters, B. T.; Kan, Z. Y.; Mayne, L.; Rosen, L. E.; Marqusee, S.; Englander, S. W. Proc Natl Acad Sci USA 2013, 110, 7684-7689.
18. Bothe, J. R.; Nikolova, E. N.; Eichhorn, C. D.; Chugh, J.; Hansen, A. L.; Al-Hashimi, H. M. Nat Methods 2011, 8, 919-931.
19. Delagoutte, E.; von Hippel, P. H. Quat Rev Biophys 2002, 35, 431-478.
20. Inman, R. B. J Mol Biol 1966, 18, 464-476.
21. Morris, C. F.; Sinha, N. K.; Alberts, B. M. Proc Natl Acad Sci USA 1975, 72, 4800-4804.
22. Nossal, N. G. J Biol Chem 1979, 254, 6026-6031.
23. Nossal, N. G.; Peterlin, B. M. J Biol Chem 1979, 254, 6032-6037.
24. McGhee, J. D.; von Hippel, P. H. Biochemistry 1975, 14, 1297-1303.
25. McGhee, J. D.; von Hippel, P. H. Biochemistry 1975, 14, 1281-1296.
26. McGhee, J. D.; von Hippel, P. H. Biochemistry 1977, 16, 3267-3276.
27. McGhee, J. D.; von Hippel, P. H. Biochemistry 1977, 16, 3276-3293.
28. Bohnuud, T.; Beglov, D.; Ngan, C. H.; Zerbe, B.; Hall, D. R.; Brenke, R.; Vajda, S.; Frank-Kamenetskii, M. D.; Kozakov, D. Nucleic Acids Res 2012, 40, 7644-7652.
29. Kitayner, M.; Rozenberg, H.; Rohs, R.; Suad, O.; Rabinovich, D.; Honig, B.; Shakked, Z. Nat Struct Mol Biol 2010, 17, 423-429.
30. Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Nature 2011, 470, 498-502.
31. Lepecq, J. B.; Yot, P.; Paoletti, C. C R Hebd Seances Acad Sci 1964, 259, 1786-1789.
32. Lerman, L. S. J Mol Biol 1961, 3, 18-30.
33. Waring, M. J. Annu Rev Biochem 1981, 50, 159-192.
34. Wilhelm, M.; Mukherjee, A.; Bouvier, B.; Zakrzewska, K.; Hynes, J. T.; Lavery, R. J Am Chem Soc 2012, 134, 8588-8596.
35. Sung, R. J.; Zhang, M.; Qi, Y.; Verdine, G. L. J Biol Chem 2013, 288, 10012-10023.
36. von Hippel, P. H.; Delagoutte, E. Cell 2001, 104, 177-190.
37. Alberts, B. M.; Frey, L. Nature 1970, 227, 1313-1318.
38. Jose, D.; Weitzel, S. E.; von Hippel, P. H. Proc Natl Acad Sci USA 2012, 109, 14428-14433.
39. Johnson, N. P.; Baase, W. A.; von Hippel, P. H. Proc Natl Acad Sci USA 2004, 101, 3426-3431.
40. Johnson, N. P.; Baase, W. A.; von Hippel, P. H. J Biol Chem 2005, 280, 32177-32183.
41. Johnson, N. P.; Baase, W. A.; von Hippel, P. H. Proc Natl Acad Sci USA 2005, 102, 7169-7173.
42. Datta, K.; Johnson, N. P.; von Hippel, P. H. J Mol Biol 2006, 360, 800-813.
43. Datta, K.; von Hippel, P. H. J Biol Chem 2008, 283, 3537-3549.
44. Baase, W. A.; Jose, D.; Ponedel, B. C.; von Hippel, P. H.; Johnson, N. P. Nucleic Acids Res 2009, 37, 1682-1689.
45. Jose, D.; Datta, K.; Johnson, N. P.; von Hippel, P. H. Proc Natl Acad Sci USA 2009, 106, 4231-4236.
46. Datta, K.; Johnson, N. P.; von Hippel, P. H. Proc Natl Acad Sci USA 2010, 107, 17980-17985.
47. Datta, K., Johnson, N. P., Villani, G., Marcus, A. H., von Hippel, P. H. Nucleic Acids Res 2012, 40, 1191-1202.
48. Jose, D.; Weitzel, S. E.; Jing, D.; von Hippel, P. H. Proc Natl Acad Sci USA 2012, 109, 13596-13601.
49. Lee, W.; Jose, D.; Phelps, C.; Marcus, A. H.; von Hippel, P. H. Biochemistry 2013, 52, 3157-3170.
50. Callen, H. B. Thermodynamics and an Introduction to Thermostatistics; John Wiley & Sons: New York, 1985.
51. Hamdan, S. M.; van Oijen, A. M. J Biol Chem 2010, 285, 18979-18983.
52. Neuman, K. C. J Biol Chem 2010, 285, 18967-18971.
53. McKinney, S. A.; Freeman, A. D. J.; Lilley, D. M. J.; Ha, T. Proc Natl Acad Sci USA 2005, 102, 5715-5720.
54. Ha, T.; Rasnik, I.; Cheng, W.; Babcock, H. P.; Gauss, G. H.; Lohman, T. M.; Chu, S. Nature 2002, 419, 638-641.
55. Altan-Bonnet, G.; Libchaber, A.; Krichevsky, O. Phys Rev Lett 2003, 90, 138101-138104.
56. Peyrard, M.; Cuesta-Lopez, S.; James, G. J Biol Phys 2009, 35, 73-89.
57. Cheatham, T. E., III; Kollman, P. A. Annu Rev Phys Chem 2000, 51, 435-471.
58. Jonas, D. M. Science 2003, 300, 1515-1517.
59. Widom, J. R., Johnson N. J., von Hippel P. H., Marcus A. H. New J Phys 2013, 15, 025029-025016.
60. Lott, G. A.; Perdomo-Ortiz, A.; Utterbach, J. K.; Widom, J. R.; Aspuru-Guzik, A.; Marcus, A. H. Proc Natl Acad Sci USA 2011, 108, 16521-16526.
61. Perdomo-Ortiz, A.; Widom, J. R.; Lott, G. A.; Aspuru-Guzik, A.; Marcus, A. H. J Phys Chem B 2012, 116, 10757-10770.
62. Widom, J. R.; Lee, W.; Perdomo-Ortiz, A.; Rappoport, D.; Molinski, T. F.; Aspuru-Guzik, A.; Marcus, A. H. J Phys Chem A, 2013, 117, 6171–6184.
63. Tekavec, P. F.; Lott, G. A.; Marcus, A. H. J Chem Phys 2007, 127, 214307-214321.
64. Phelps, C.; Lee, W.; Jose, D.; von Hippel, P. H.; Marcus, A. H. Proc Natl Acad Sci USA, (2013) in press.
65. Richardson, R. W.; Nossal, N. J Biol Chem 1989, 264, 4725-4731.
66. Dong, F.; Gogol, E. P.; von Hippel, P. H. J Biol Chem 1995, 270, 7462.
67. Dong, F.; von Hippel, P. H. J Biol Chem 1996, 271, 19625-19631.

Reviewing Editor: Kenneth J. Breslauer