The Triplex-Hairpin Transition in Cytosine-Rich DNA

Anton S. Petrov, Gene Lamm, and George R. Pack
Department of Chemistry, University of Louisville, Louisville, Kentucky

ABSTRACT We present a theoretical study of the self-complementary single-stranded 30-mer ***d***(TC*TTC*C*TTTTCCTTCTC*-CCGAGAAGGTTTT) (PDB ID: 1b4y) that was designed to form an intramolecular triplex by folding back twice on itself. At neutral pH the molecule exists in a duplex hairpin conformation, whereas at acidic pH the cytosines labeled by an asterisk (*) are protonated, forming Hoogsteen hydrogen bonds with guanine of a GC Watson-Crick basepair to generate a triplex. As a first step in an investigation of the energetics of the triplex-hairpin transition, we applied the Bashford-Karplus multiple site model of protonation to calculate the titration curves for the two conformations. Based on these data, a two-state model is used to study the equilibrium properties of transition. Although this model properly describes the thermodynamics of the protonation-deprotonation steps that drive the folding-unfolding of the oligomer, it cannot provide insight into the time-dependent mechanism of the process. A series of molecular dynamics simulations using the ff94 force field of the AMBER 6.0 package was therefore run to explore the dynamics of the folding/unfolding pathway. The molecular dynamics method was combined with Poisson-Boltzmann calculations to determine when a change in protonation state was warranted during a trajectory. This revealed a sequence of elementary protonation steps during the folding/unfolding transition and suggests a strong coupling between ionization and folding in cytosine-rich triple-helical triplexes.

INTRODUCTION

Since the first triple-helical nucleic acid complex containing two strands of poly(uridylic acid) and one strand of poly(adenylic acid) was discovered in 1957 (Felsenfeld et al., 1957) numerous other examples have been found and investigated (Frank-Kamenetskii and Mirkin, 1995). Among these structures much attention has been given to the H-DNA molecule, which contains both triple- and single-stranded regions (Mirkin et al., 1987; Lyamichev et al., 1985). H-DNA has been found in the genomes of vertebrates and invertebrates (Htun et al., 1984; Glikin et al., 1983) and is thought to play a regulatory role in replication (Rao et al., 1988; Brinton et al., 1991) and transcription (Shimizu et al., 1989; Kato and Shimizu, 1992). Thermodynamic and structural studies of triplex formation use intramolecular sequences of 20–30 nucleotides that fold back twice on themselves, forming a paperclip-like structure called a hairpin (Sklenar and Feigon, 1990; Chen, 1991). There are several reasons for this: first, the strands of these triplexes are always present in stoichiometric amounts; second, because the folded tertiary structure is defined by the primary sequence (Volker and Klump, 1994) there is only a small chance of the formation of competing structures; and third, these triplexes are more stable compared to intermolecular ones and may be formed in the absence of stabilizing agents over a wider range of temperature (Soyfer and Potaman, 1996).

Generally, a nucleic acid duplex composed of one homopyrimidine and one homopurine strand can form inter- as well as intramolecular triplexes by binding a homopyrimidine strand parallel to the homopurine strand (Riley et al., 1966). Antiparallel binding of a purine strand to the pyrimidine strand of such a duplex is also possible (Arnott and Selsing, 1974). The first mechanism involves the formation of the TA·T and CG·C^+ triads stabilized by conventional Watson-Crick pairing in the original duplex and Hoogsteen pairing to the third strand. The formation of a GC^+ Hoogsteen pair requires cytosine to be protonated at the N3 position, thus making the stability of these triads pH-dependent. It has been demonstrated that CG·C triads are more stable than TA·T triads at low pH (Roberts and Crothers, 1996) due to the formation of stronger hydrogen bonds in the Hoogsteen pair, electrostatic stabilization of the positively charged triad by the negative backbone (Sun et al., 1991), and favorable secondary interactions (Jorgensen and Pranata, 1990). An increase in pH induces cytosine deprotonation and destabilizes the CG·C triads, leading to dissociation of the third strand. Under this condition, intramolecular ''paperclip'' triplexes will unfold to hairpin duplexes. One of the factors determining the stability of triplexes is the pK_a of cytosine N3 atoms of the third strand. These pK_a values can be estimated by various methods including ultraviolet absorption spectroscopy, circular dichroism (CD) spectroscopy, fluorescence (Soyfer and Potaman, 1996), and nuclear magnetic resonance (NMR) (Leitner et al., 1998;Weisz et al., 2000). All of these methods indicate that the pK_a depends upon the position of the particular cytosine within the molecule as well as the global molecular conformation.

Protonation constants of the titratable groups of a macromolecule in an electrolyte solution can be estimated computationally by solving the Poisson-Boltzmann (PB) equation (Pack et al., 1993; Nicholls and Honig, 1991; Madura et al.,

Submitted March 29, 2004, and accepted for publication September 3, 2004.

Address reprint requests to George R. Pack, E-mail: george.pack@louisville.edu.

© 2004 by the Biophysical Society

0006-3495/04/12/3954/20 $2.00

doi: 10.1529/biophysj.104.043752

1995). Bashford and Karplus (1990) have presented the standard approach for calculation of pK_a values of titratable sites of amino acids in proteins based on the Debye-Hückel (DH) approximation to the PB equation. The method may be applied with reasonable accuracy to systems containing polyelectrolyte molecules that are not highly charged and, although neglecting polarization of the protonation sites (Sham et al., 1997), provides satisfactory results for many systems (Antosiewicz et al., 1996, 1994). A few methods that account for the local conformational flexibility of proteins have been proposed (You and Bashford, 1995; Ripoll et al., 1996; Zhou and Vijayakumar, 1997; Georgescu et al., 2002; Lim et al., 1991; Li et al., 2002). However, triple-helical DNA systems pose two difficulties when using the Bashford-Karplus approach to calculate pK_a values. First, nucleic acids are highly charged macromolecules and the DH approximation severely overestimates the electrostatic potential at the surface (Pack et al., 1998; Lamm, 2003). Second, the Bashford-Karplus method, as typically applied to proteins, assumes that protonation and folding processes are not strongly coupled—an approximation which, although valid in those cases, fails for hairpin-triplex transitions. Details of the PB calculations are presented in Methods.

Because protonation is very sensitive to local and global conformation, structural changes are an important consideration in understanding triple-helix formation. In computer studies of folding, conformational changes can, in principle, be taken into account by molecular dynamics (MD) methods. In practice, major structural reorganization occurs on a timescale of micro- to milliseconds that is much longer than the typical nanosecond range of simulations. Early simulations were also confronted with problems in the parameter sets: the standard AMBER (Cornell et al., 1995) and CHARM (MacKerell et al., 1995) parameter sets were unable to describe the formation of a GC^+ Hoogsteen pair due to absence of parameters for a protonated cytosine. MD simulations on triple helices were first performed on structures that contained only TA·T triads (Kiran and Bansal, 1995; Shields et al., 1997) or on triplexes stabilized by GC·G and CG·C triads containing unprotonated cytosines (Ojha and Tiwari, 2002). The development of additional parameters (Sponer et al., 1997) for the iminohydrogen of cytosine (H3) presented an opportunity to simulate nucleic acid molecules containing three (Spackova et al., 1998; Basye et al., 2001) or four (Csaszar et al., 2001) strands stabilized by GC^+ Hoogsteen basepairs. These studies confirmed the stability of $CG{\cdot}C^+$ triads. In these calculations the cytosines were protonated a priori, regardless of their pK_a values, and the number of protons was fixed.

Hairpin tetraloops are important structural fragments of RNA and DNA molecules and are often involved in the formation of tertiary structures from secondary units (Moore, 1999; Brion and Westhof, 1997). It has been proposed that tetraloops regulate the activity of macromolecules by shifting the equilibrium between alternate structures (Gluck et al., 1994; Wool et al., 1992) and that they may serve as recognition elements for proteins and nucleic acids (Michel and Westhof, 1990; Zwieb, 1992; Murphy and Cech, 1994). According to a classification proposed by Hilbers et al. (1991) tetraloops can be divided into three groups depending on the position of the second and third bases of the loop. Particularly, loops belonging to group II are of major interest because they can be found in both DNA and RNA (van Dongen et al., 1999; Jucker and Pardi, 1995). The conformation of loops in this group is characterized by the following feature: a base of the second residue at the 5′ end of a molecule is turned into or toward the minor groove, whereas a base of the third residue stacks over a (non-canonical) basepair formed by the two remaining residues of the tetraloop. It has been shown that the formation of group II loops is determined predominantly by the nature of the second residue in a loop rather than by the conformation of underlying helical stem. This folding pattern will be realized if a base of this residue is a pyrimidine (van Dongen et al., 1996).

Elucidation of the mechanism of formation of C-containing triplexes requires an understanding of how pH change alters the population of protonation substates and how this in turn affects the global conformation of the molecule. This article presents calculations on the self-complementary single-stranded DNA 30-mer ***d***(TC*TTC*C*TTTTCCTTCTC*CCGAGAAGGTTTT), where C* indicates a protonation site. This sequence folds back twice on itself forming GCCC and TTTT tetraloops in an intramolecular triple helix (PDB ID: 1b4y). NMR studies have shown the triple-helical structure to be stable at pH 4.5 and that an increase to pH 7.0 induces unfolding into a double-helical hairpin structure (van Dongen et al., 1999). The first part of this article discusses an equilibrium two-state model that illuminates the conformational transition by taking into account the pK_a values of protonation sites for the hairpin and triplex endstates. In the second part of the article we describe MD simulations which show the effect of protonation of titratable sites on the local and global conformation of DNA. Particularly, we tested the influence of the protonated incoming third strand on the conformation of the CCCG tetraloop. The calculations indicated that it is necessary to treat the conformational and electrostatic factors as highly coupled to obtain a quantitative description of the folding mechanism.

METHODS

Charge derivation procedure

When atom N3 on cytosine (C) is protonated, the added charge is distributed over the entire ring so a recalculation of charges from the standard AMBER data set is necessary. A cytosine base was extracted from a canonical B-DNA structure and terminated by hydrogenating N1. A similar procedure was performed to obtain the geometry of a deoxycytosine nucleotide (*dC*) in which the O5′ atom bound to phosphate was methylated and another O5′

atom bound to carbon C5′ hydrogenated to terminate the nucleotide (Fig. 1). ESP charges were derived according to the Merz-Kollman scheme (Singh and Kollman, 1984) at the RHF/6-31G(d) level for the unprotonated and protonated forms of dC and C. These charges were used as input to calculate RESP charges using the RESP program implemented in the AMBER package (Pearlman et al., 1995). RESP charges for C and C^+ were calculated without any constraints. To ensure an integral amount of charge on each nucleotide, a constraint was applied to the nucleotides in their unprotonated and protonated forms (i.e., $Q_{\mathrm{I}} - Q_{\mathrm{II}} = 0$ in Fig. 1). The calculation of pK_a values according to the Bashford-Karplus scheme (discussed below) requires the same set of charges for the model compound (in our case cytosine) and the protonation site dC in DNA. Two extra constraints were therefore applied for dC and dC^+ ($Q_{\mathrm{III}} = 0$ and $Q_{\mathrm{IV}} = 1$ in Fig. 1) and the charges of these fragments were replaced by those obtained for C and C^+, respectively. The derived sets of RESP charges for nucleotides dC and dC^+ with constrained fragments are presented in Fig. 1. The AMBER ff94 parameter set was used for all other nucleotide charges (Cornell et al., 1995).

Electrostatic calculations

We used the finite difference method as implemented in the University of Houston Brownian Dynamics program to calculate the electrostatic potential at the titratable sites (Davis et al., 1991). Determination of the potential ϕ due to a highly-charged molecule requires solving the full nonlinear Poisson-Boltzmann equation

$$\nabla\varepsilon(r)\nabla\phi(r) - \varepsilon(r)\kappa^2 \sinh(\phi(r)) = -\frac{4\pi L_{\mathrm{B}}}{e_0}\rho^{\mathrm{f}}, \qquad (1)$$

where $\varepsilon(r)$ is the local dielectric coefficient, $\kappa^2 = 8\pi L_{\mathrm{B}} C$ is the square of the Debye constant, C is the concentration of added salt (the experimental value 0.1 M was used; van Dongen et al., 1999), $L_{\mathrm{B}} = e_0^2/k_{\mathrm{B}}T$ is the Bjerrum length, e_0 is the proton charge, $k_{\mathrm{B}}T$ is the temperature, and ρ^{f} is the distribution of fixed charges. For comparison we also tested the (linear) Debye-Hückel (DH) version:

$$\nabla\varepsilon(r)\nabla\phi(r) - \varepsilon(r)\kappa^2\phi(r) = -\frac{4\pi L_{\mathrm{B}}}{e_0}\rho^{\mathrm{f}}. \qquad (2)$$

The potential at the boundary of a coarse grid ($65 \times 65 \times 65$) with grid spacing 3.0 Å was set to zero. A focusing procedure was performed at each titratable site using a finer grid (0.4 Å) with the same number of grid points. Calculations were performed using dielectric coefficients of 20 and 78.5 for the DNA and solvent, respectively. The set of atomic radii of the AMBER 94 force field was used in all electrostatic calculations.

pK_a calculations

In the present work a titratable site consists of all cytosine base atoms including either C1′ (for the polynucleotide) or H1 (for the model compound); this assignment conforms to the reduced site approximation (Bashford and Karplus, 1991) in which protonation of phosphate ions are neglected (their pK_{intr} is lower by 2–3 units than that of cytosine and outside the range of the simulations), but allows the charge in protonating cytosine atom N3 to be distributed throughout the base.

According to the Tanford-Kirkwood approach (Tanford and Kirkwood, 1957), the intrinsic pK_a of a titratable site μ ($pK_{\mathrm{intr},\,\mu}$) of a macromolecule can be calculated from the pK_a of the site in the absence of the molecule (for cytosine, $pK_{\mathrm{model},\,\mu} = 4.3$) by taking into account the effect of the dielectric property of the macromolecule of the electrostatic free energy of the complex. That is,

$$pK_{\mathrm{intr},\,\mu} = pK_{\mathrm{model},\,\mu} - \beta\Delta\Delta G(\mu)/2.3, \qquad (3)$$

where $\Delta\Delta G(\mu)$ is the difference in the free energy changes between the process of dissolving cytosine μ in DNA in its protonated (H) and unprotonated (0) forms. In the (linear) Bashford-Karplus approach, the free energy difference can be expressed as the sum of two terms, $\Delta\Delta G_{\mathrm{Born}}(\mu)$ and $\Delta\Delta G_{\mathrm{back}}(\mu)$, where each free energy term can be calculated provided that both the site μ of DNA and the model compound have the same internal geometry, the same set of charges, and the same arrangement of grid points (to eliminate the self energy term). Thus, following Bashford and Karplus, but with the understanding that protonation of a single site μ may alter the charges of multiple atoms i within that site (to give the charge set Q_μ^{H}), we have (Lamm, 2003)

$$\beta\Delta\Delta G_{\mathrm{Born}}(\mu) = \frac{1}{2}\sum_{\mathrm{i}\in\mu}^{\text{titratable atoms}} Q_{\mathrm{i}}^{\mathrm{H}}\left[\phi_{\mathrm{DNA}}^{\mathrm{H}}(Q_\mu^{\mathrm{H}}|r_{\mathrm{i}}) - \phi_{\mathrm{model}}^{\mathrm{H}}(Q_\mu^{\mathrm{H}}|r_{\mathrm{i}})\right] - \frac{1}{2}\sum_{\mathrm{i}\in\mu}^{\text{titratable atoms}} Q_{\mathrm{i}}^{0}\left[\phi_{\mathrm{DNA}}^{0}(Q_\mu^{0}|r_{\mathrm{i}}) - \phi_{\mathrm{model}}^{0}(Q_\mu^{0}|r_{\mathrm{i}})\right]. \qquad (4)$$

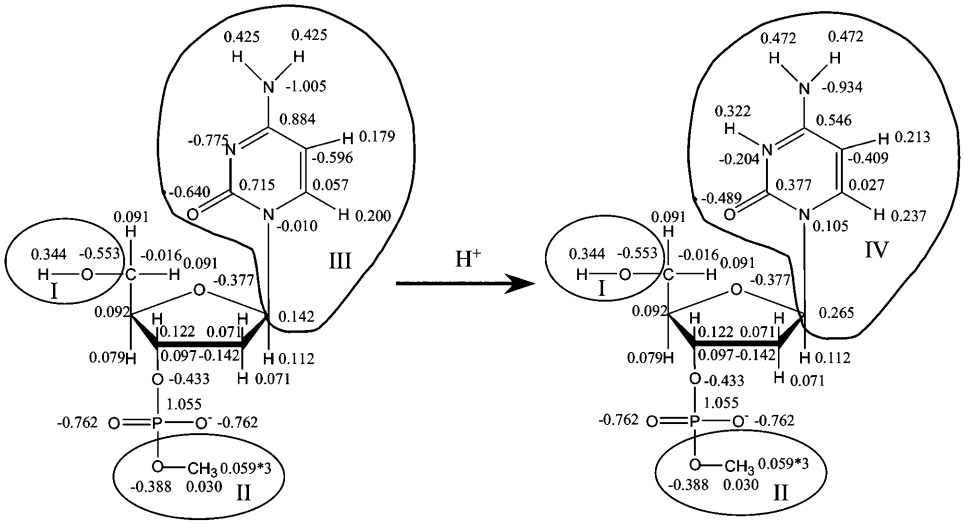

FIGURE 1 Scheme used to derive the RESP charges for dC and dC^+; charges derived at HF/6-31G(***d***) level and used in the MD simulations are shown (see Methods for details).

The notation $\phi^{H}_{DNA}(Q^{H}_{\mu}|r_i)$ denotes the value of the electrostatic potential in the DNA (as opposed to that in the model compound) at position r_i of atom i belonging to site μ due to the protonated charge set Q^{H}_{μ}. The Born term accounts for the difference in the energies of transferring the protonated and unprotonated forms of the cytosine site into a low dielectric DNA cavity. (Note that all atoms of a site are titratable in this site-distributed charge approach.)

Because the DNA oligomer is highly charged, to calculate the influence of the background charges upon the pK_a, we used an approach that differs slightly from one proposed by Bashford and Karplus (1990). In their approach the background free energy contribution is calculated as the product of the background charges and the potential due to a model compound embedded into the dielectric cavity of the macromolecule. This assumes that the energy is linear in the background charges, as obtained, for example, in a Guntelberg-type charging process in which the linear Poisson-Boltzmann equation is used. However, because DNA is highly charged, we choose instead to charge the much smaller protonatable sites in the presence of the nonlinear PB-determined background potential. We thus write the background free energy contribution as

$$\beta\Delta\Delta G_{back}(\mu) = \sum_{i\in\mu}^{\text{titratable atoms}} (Q^{H}_{i} - Q^{0}_{i})\phi^{back}(q^{back}|r_i), \qquad (5)$$

where $\phi^{back}(q^{back}|r_i)$ is the potential due to background charges q^{back} of the DNA molecule calculated at the positions r_i of the charges Q_i of the model compound.

We must also take into account the presence of other mutually titratable sites in the molecule. The protonation of site μ in the presence of a proton on site ν costs an energy $W_{\mu\nu}$ due to the electrostatic interaction between these sites:

$$\beta W_{\mu\nu} = \sum_{i\in\mu}^{\text{titratable atoms}} (Q^{H}_{i} - Q^{0}_{i}) \times \left[\phi_{DNA}(Q^{H}_{\nu}|r_i) - \phi_{DNA}(Q^{0}_{\nu}|r_i)\right]. \qquad (6)$$

Site-site interaction energies have been obtained explicitly for each of the 16 possible protonation states. In the spirit of Tanford and Roxby (1972) and Bashford and Karplus (1990), an ''approximate'' or ''apparent'' pK_a (or pK_{app}) can be defined by

$$pK_{app,\mu} = pK_{intr,\mu} - \frac{\beta}{2.3}\sum_{\nu\neq\mu} x_\mu x_\nu W_{\mu\nu}, \qquad (7)$$

where x_μ is either 1 or 0 depending on whether site μ is protonated or not. These pH-independent pK_{app} values are useful as a more accurate indicator of site protonation (at a given pH) than the intrinsic values. By knowing the $pK_{intr,\mu}$ for each site μ and the interaction energy $W_{\mu\nu}$ between sites, we can calculate the fraction of molecules having site σ protonated as a function of pH. The titration curve for site σ is given by

$$\theta_\sigma = \frac{\sum_{i,j} x_{ij}\exp\left[\sum_\mu x_\mu 2.3(pK_{intr,\mu} - pH) - \frac{\beta}{2}\sum_{\mu,\nu} x_\mu x_\nu W_{\mu\nu}\right]}{\sum_{i,j}\exp\left[\sum_\mu x_\mu 2.3(pK_{intr,\mu} - pH) - \frac{\beta}{2}\sum_{\mu,\nu} x_\mu x_\nu W_{\mu\nu}\right]}, \qquad (8)$$

where the sum over i and j accounts for all possible protonation states (with elements x_{ij}, see Appendix for details). The total titration curve is then the sum of the titration curves for all sites.

MD simulations

Molecular dynamics (MD) simulations were carried out using the AMBER 6.0 suite of programs (Pearlman et al., 1995) and the force field developed by Cornell et al. (1995) Additional force field parameters for protonated cytosine were taken from Sponer et al. (1997) The set of charges previously derived for the PB calculations was used for protonated cytosines.

Each system was prepared using the XLeaP module by surrounding the molecule in a water box that typically contained ~6500 water molecules and 25–29 Na^+ ions, depending on the number of protonated cytosines. MD simulations were performed in the NTP ensemble with a time step of 2 fs using the SHAKE algorithm (Ryckaert et al., 1977). Coupling to an external heat bath and volume scaling were applied every 0.2 ps to keep the system at 300 K and 1 atm; a 9.0 Å cutoff was applied for van der Waals interactions.

Water molecules and counterions were equilibrated by minimizing the system (1000 steps) under positional restraints on the DNA with a force constant of 500 kcal mol^{-1} $Å^{-2}$. This was followed by a restrained MD run of 25 ps with heating from 100 K to 300 K during the first ps. The restrained MD run was continued for 25 ps more using the particle-mesh Ewald method (Darden et al., 1993). After this the restraints on the DNA molecule were reduced by the following scheme: 1000 minimization steps with 25 kcal mol^{-1} $Å^{-2}$ position restraints; then 3 ps dynamics at 300 K with 25 kcal mol^{-1} $Å^{-2}$ position restraints with five consecutive rounds of 600 steps minimization, removing positional restraints by 5 kcal mol^{-1} $Å^{-2}$ for each run. The system with no positional restraints was heated from 100 K to 300 K during a 20-ps MD run followed by the 1-ns production run; snapshots of the trajectory were saved every 1 ps.

Trajectories were analyzed using the Carnal module of the AMBER suite and CURVES 5.3 (Lavery and Sklenar, 1988, 1989) was used to obtain global and local helix parameters, sugar pucker conformations, and backbone dihedral angles.

RESULTS

The two-state equilibrium model

Initial structures

The DNA oligomer schematically shown in Fig. 2 *a* exists in a triple-helical conformation (the τ-state, Fig. 2 *b*) at low pH due to protonation of the N3 atoms of C_{18}, C_{21}, C_{22}, and C_7, which form Hoogsteen basepairs with guanines (PDB ID:1b4y, *Structure 10*). The first three cytosines are located in the triple-helical stem and form $CG{\cdot}C^+$ triads; C_7 is located in a CCCG tetraloop and forms a Hoogsteen basepair with G_{10}. At neutral pH all cytosines are unprotonated and the DNA molecule exists as a hairpin duplex (the δ-state; for example, see Fig. 2 *c*). Overall, the equilibrium can be expressed as

$$\delta + 4H^+ \rightleftharpoons \tau. \qquad (9)$$

Within the framework of a two-state model, we assume that the molecule adopts either the τ- or δ-conformation and that each of these states has four protonation sites with $2^4 = 16$ protonation substates. Only the four aforementioned cytosine bases are considered protonatable in this reduced-site approximation (see Methods, above; see also Bashford and Karplus, 1991). The triplex structure predominates at low pH, where all four cytosines favor protonation, whereas the hairpin form predominates at neutral (and high) pH where a shift toward unprotonated bases occurs. The population of

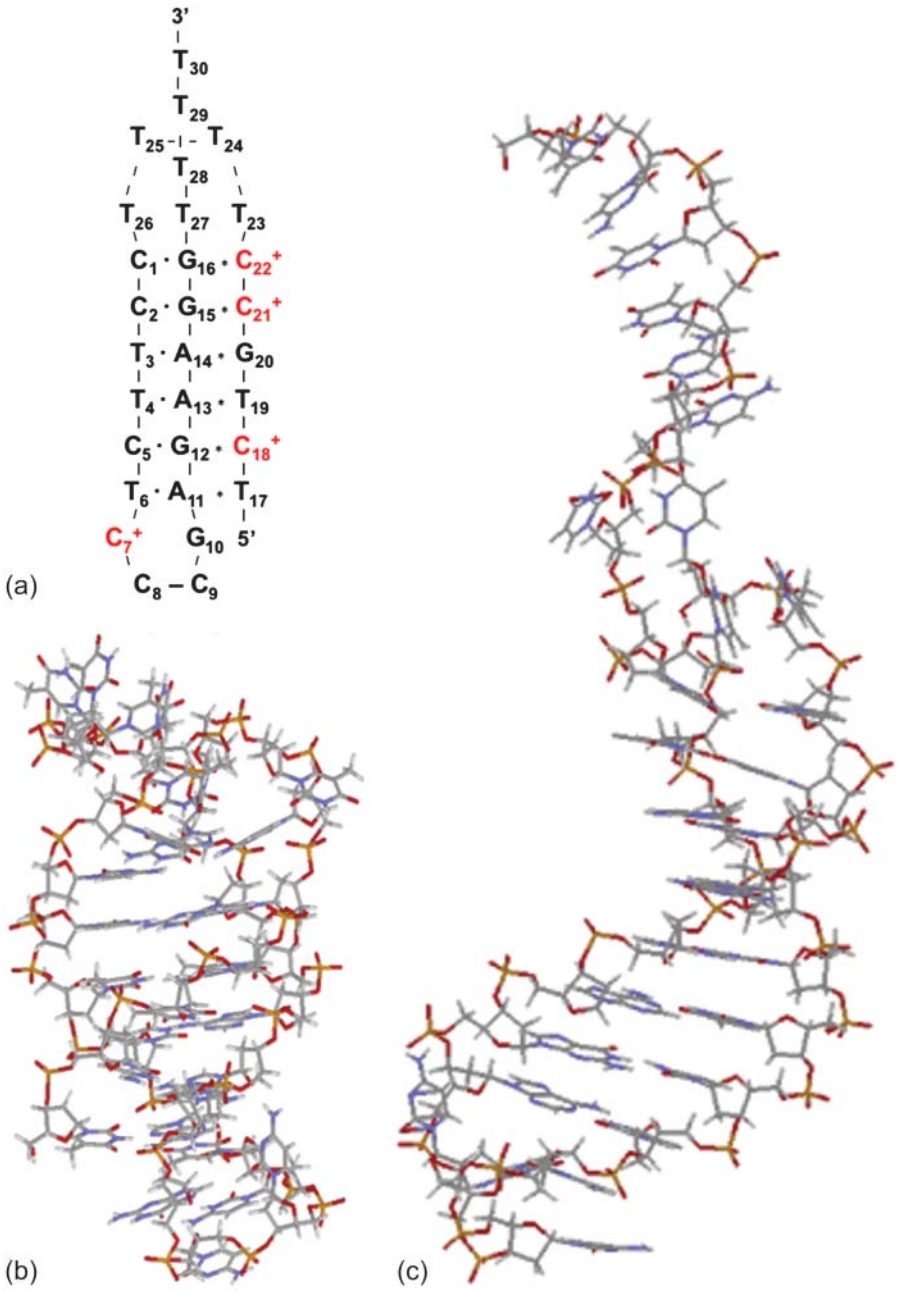

FIGURE 2 (*a*) Schematic representation of triple-helical DNA (PDB ID: 1b4y). Protonated cytosines are highlighted in red. The original (NMR) triplex (*b*) and relaxed duplex (*c*) conformations used in the two-state model.

protonation substates within each conformation, in addition to the τ–δ structural free energy difference, determines the relative concentration of conformations in the equilibrium mixture. Most observables, such as ultraviolet absorbance, depend only on the ratio of the concentrations of the two conformations, not on the population of protonation substates. The calculation of pK_a values and titration and absorbance curves are discussed in the following sections.

The structure of the triplex conformation used for the equilibrium calculations was that of the PDB file. All atomic charges were obtained from the AMBER ff94 parameter set except those for cytosine for which an RESP recalculation was performed (see Methods, above). The hairpin duplex structure was generated from the triplex by changing the dihedral angles of T_{23}, T_{24}, and T_{25} originally located in the TTTT loop, and rearranging the Hoogsteen G_{10}–C_7^+ pair into a Watson-Crick G_{10}–C_7 pair by flipping the guanine base. This significant structural difference in the CCCG tetraloop was noted previously (van Dongen et al., 1996). A 1-ns MD trajectory was generated to obtain the equilibrated hairpin conformation. Conformations of the NMR triplex (denoted τ_4^{NMR} in Eq. 14 below) and hairpin duplex after equilibration (denoted δ_0^0 in Eq. 15 below) are shown in Fig. 2, *b* and *c*, respectively.

Calculation of intrinsic pK$_a$ values

Tanford and Kirkwood introduced the concept of an intrinsic pK_a (or pK_{intr}), defined as the pK_a of a site on a molecule when all other sites are unprotonated (Tanford and Kirkwood, 1957). The pK_{intr} values of the titratable sites for the triplex and hairpin conformations (Table 1) were calculated as described in Methods, above, according to the method of Bashford and Karplus (1991) by solving the PB equation using the University of Houston Brownian Dynamics program (Davis et al., 1991). The pK_{intr} values of the third-strand cytosines of the triplex conformation lie between 8.4 and 9.5 and increase as the position of the titratable site moves from one end of the molecule toward the middle. This variation of pK_a with position is readily understood by noting that the electrostatic potential is more negative in the center of the triplex than at an end. Asensio et al. (1998) have performed NMR measurements of the pK_a values of triplexes containing a single CG·C$^+$ triad with this triad positioned at different locations along the strand and have shown that the pK_a of cytosine N3 of the triad ''must be at least 9.5 for internal positions.'' The pK_a values in these experiments are identical to pK_{intr} values because the molecules have only a single protonation site. These experimental data are in a good agreement with theoretical calculations of pK_{intr} values performed by Pack et al. (1998).We mention that experimental studies of similar systems using indirect methods, in which the measured pK_a values are in the 6–7 range (Volker and Klump, 1994; Asensio et al., 1998; Plum and Breslauer, 1995; Sugimoto et al., 2001), actually record the average pK_a values of the triplex and hairpin conformations.

Our calculations show that the pK_{intr} values calculated for cytosines located in the single-strand region of the hairpin conformation lie between 5.2 and 5.5 and reflect the same position dependence as in the triplex. Even in the absence of complementary guanines, the acidity of these cytosines is still less than that of free cytosine ($pK_a = 4.3$) because of the negative electrostatic field of the molecule.

The calculation of the pK_a using Eqs. 6 and 7 assumes that the potential due to a set of charges equals the sum of the potentials of the individual charges, a property inherent in the linear Debye-Hückel equation but not necessarily applicable to the full Poisson-Boltzmann equation (Eq. 1). To test the accuracy of Eqs. 6 and 7, we calculated the pK_{intr} of C_{18} in the hairpin conformation using a more accurate two-step approach. First, the Born and background terms were calculated in the absence of electrolyte by solving the (linear) Poisson equation. After that, corrections due to added salt were obtained separately by solving the nonlinear Poisson-Boltzmann equation. (A detailed description of this method will appear in a forthcoming article, Petrov et al., 2005.) The pK_{intr} of C_{18} calculated by this method is 5.0, which is only 0.2 units less than that calculated obtained by using Eqs. 6 and 7, suggesting that the interactions are still within (or near) the linear regime, and that the approximation is tolerable.

TABLE 1 Intrinsic and apparent pK$_a$ values calculated for site N3 of the protonatable cytosine bases for the triplex (τ) and hairpin (δ) conformations used in the two-state model

Site	$pK_{intr}(\tau)$	$pK_{intr}(\delta)$	$pK_{app}(\tau)$	$pK_{app}(\delta)$
C_{18}	8.4	5.2	8.1	5.1
C_{21}	9.5	5.4	8.2	4.7
C_{22}	9.0	5.5	7.9	4.9
C_7	7.7	4.2	7.5	4.2

Apparent pK_a values were calculated from the intrinsic pK_a values according to Eq. 7 with site-site interaction energies given in Table 2.

Influence of neighboring cytosines

The definition of an intrinsic pK_a assumes that two titratable sites in a molecule do not interact, and thus pK_{intr} may be considered as an approximation to the apparent pK_a if the sites are shielded from one another by distance, solvent, and/or ions. In the present oligomer, C_{21} and C_{22} are neighbors, whereas C_{18} has two nucleotides separating it from C_{21}; in the triplex form C_7 is within one base triad of C_{18}. NMR measurements have shown that the pK_a of cytosine in a CG·C$^+$ triad in a triplex drops from 9.5 to 8.0 when an adjacent CG·C triad becomes protonated (Asensio et al., 1998). Tanford and Roxby modified the original Tanford-Kirkwood method for calculating pK_a values by taking into account site-site interactions (Tanford and Roxby, 1972). Accordingly, protonation of a particular site μ decreases the pK_a of a nearby site ν by $\beta W_{\mu\nu}/2.3$ units from its intrinsic value, with $W_{\mu\nu}$ being the site-site interaction energy. These site-site interaction energies can be estimated from the solution of the PB equation using Eq. 6; values for these energies are presented in Table 2.

The strongest interaction is found between adjacent cytosines C_{21} and C_{22} ($W_{C21,C22}$ are 1.10 and 0.57 for the triplex and hairpin states, respectively) and this interaction decays rapidly as the distance between two titratable sites increases. It is seen that all site-site interactions energies are higher in the triplex than in the hairpin, although the distances between the titratable sites (except those involving C_7) are almost identical. The titratable sites in the triplex are contained a low dielectric cavity, whereas in the hairpin they are shielded by the solvent. The site-site interaction energy is almost symmetric with respect to site indexes because the main factor in determining its magnitude is the distance between the sites; minor corrections due to nonlinearity of the free energy on the potential, as noted above, account for the slight asymmetry.

The site-site interactions of the fully protonated molecules were calculated by two different methods. According to the first method these energy terms are assumed to be additive, therefore, the total interaction $\sum_{\nu\neq\mu}\beta W_{\mu,\nu}/2.3$ was calculated as the sum of the pairwise terms $\beta W_{\mu\nu}/2.3$. However, these terms are strictly additive only within the linear DH

TABLE 2 Pairwise site-site interactions $W_{\mu\nu}$, the total influence of the neighboring sites calculated as a sum of all pairwise site-site interactions and the resulting apparent pK_a values calculated according to Eq. 7 used in the two-state model

Site $\nu \rightarrow$	*C*18	*C*21	*C*22	*C*7	$(\sum_{\nu\neq\mu}\beta W_{\mu,\nu})/(2.3)$	$\beta W_{\mu,total}/2.3$	pK_{app}
Site $\mu\downarrow$	Triplex						
C_{18}		0.10	0.03	0.18	0.33	0.32	8.05
C_{21}	0.10		1.10	0.02	1.22	1.23	8.23
C_{22}	0.03	1.08		0.01	1.13	1.13	7.88
C_7	0.18	0.02	0.01		0.21	0.23	7.49
	Hairpin						
C_{18}		0.07	0.03	3.87×10^{-5}	0.10	0.10	5.13
C_{21}	0.07		0.57	1.00×10^{-4}	0.64	0.63	4.72
C_{22}	0.03	0.58		1.00×10^{-4}	0.61	0.61	4.87
C_7	3.60×10^{-5}	1.01×10^{-4}	1.40×10^{-4}		2.7×10^{-4}	3×10^{-3}	4.15

All values are given in pH units.

theory. To test this approximation, we also calculated the corrections to the pK_a of each site $\beta W_{\mu,\text{total}}/2.3$ explicitly, assuming that the other three sites were simultaneously protonated. These values are also presented in Table 2, columns 6 and 7. The site-site interactions calculated by these two methods are in extremely good agreement, indicating that the linear DH approximation is valid for these interactions.

Titration and absorbance curves

The titration curves were calculated according to Eq. 8 using the intrinsic pK_a values of Table 1 and the pairwise site-site interaction energies of Table 2 by averaging the Boltzmann-weighted site occupancies over all possible protonation states. This use of fixed conformations assumes that the free energy change upon protonation is much larger than that due to any subsequent structural changes within a particular (triplex or hairpin) conformation and that structural free energy changes may be described solely by the free energy difference between the unprotonated triplex and hairpin conformations.

The shift between the hairpin and triplex titration curves (Fig. 3 *a*) reflects the differences in the pK_{intr} values of their titratable sites (Table 1). The fact that the two curves are not exactly parallel results from different shifts between pK_a values for the titratable sites in the two conformations and in the conformational dependence of $W_{\mu\nu}$. The midpoints of the titration curves are 5.7 and 8.0 for the hairpin and triplex conformations, respectively.

The equilibrium constant $K_{\tau-\delta}$ for the triplex-hairpin duplex (τ–δ) conformational transition in our pH-dependent two-state model is given by Eqs. A5–A7 in the Appendix. This leads to an expression for the absorbance A as a function of DNA concentration (C_{DNA}), triplex (ε_τ), and duplex (ε_δ) extinction coefficients, and pH (Eq. A13) as

$$A(\text{pH}) = C_{\text{DNA}}\frac{\varepsilon_\tau + \varepsilon_\delta K_0 K(\text{pH})}{1 + K_0 K(\text{pH})}, \tag{10}$$

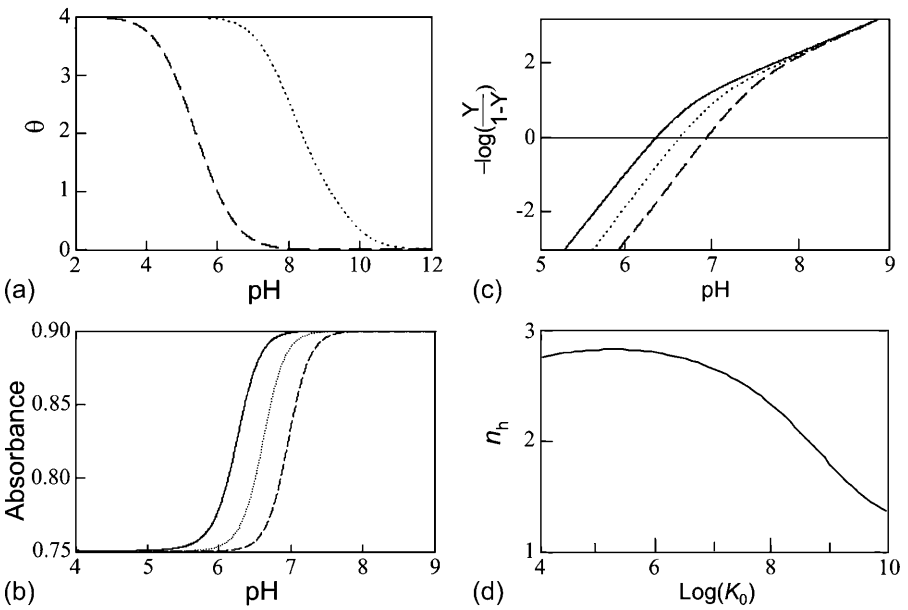

FIGURE 3 (*a*) The theoretical titration curves (fraction of protonated cytosines, θ versus pH) calculated for the triplex (*dotted*) and duplex (*dashed*) conformations, respectively. (*b*) The titration curves (absorbance versus pH) for which the observable accounts for the triplex-duplex transition calculated using Eq. A10 with different K_0: 10^8 (*solid*), 10^7 (*dotted*), and 10^6 (*dashed*). (*c*) Hill plots ($-\log Y/(1-Y)$) calculated using Eq. A14 with different K_0: 10^8 (*solid*), 10^7 (*dotted*), and 10^6 (*dashed*). (*d*) Dependence of the Hill coefficient, n_h, on folding constant K_0.

where

$$K_{\delta-\tau} = \frac{\Delta}{\mathrm{T}} = K_0 K(\mathrm{pH}) \tag{11}$$

is the relative amount of DNA in the duplex (δ) and triplex (τ) conformations in the two-state model. In Eq. 10 K_0 is the equilibrium constant for the unprotonated triplex-hairpin conformational transition and, although difficult to estimate, can be obtained by fitting an experimental absorption curve. In Fig. 3 *b* we show the absorbance as a function of pH for three values of K_0 (10^6, 10^7, and 10^8) using the calculated $\mathrm{pK_a}$ values and the extinction coefficients $\varepsilon_\delta = 0.76$ and $\varepsilon_\tau = 0.90$ (Asensio et al., 1998). The predicted absorbance is seen to depend strongly on K_0. Given that the duplex-triplex transition occurs between 4.5 and 7.0 pH (van Dongen et al., 1999), the absorbance curves suggest that K_0 lies in the range 10^6–10^8. It then follows that the fully unprotonated oligomer is more stable in the hairpin than in the triplex conformation by $\ln(K_0)/\beta \sim 10$ kcal/mol. By contrast, the difference in free energy change gained by protonating all cytosines in the triplex versus those in the hairpin is $\beta\Delta\Delta G = 2.3\sum_{\mu}^{\mathrm{N}}(\mathrm{p}K_{\tau,\mu} - \mathrm{p}K_{\delta,\mu}) = \sim 20\,\mathrm{kcal/mol}$.

To investigate the cooperativity of the triplex-hairpin transition, we performed a Hill analysis (see Eqs. A14 and A15 in the Appendix). Fig. 3 *c* shows a plot of $-\log(Y/(1-Y))$ versus pH, where Y is the fraction of molecules with *any* cytosine protonated (Voet and Voet, 1995), for several values of K_0. The slope of this curve at $Y = 0.5$ (zero on the ordinate) gives the Hill coefficient (n_h), which is shown in Fig. 3 *d* as a function of $\log(K_0)$. Values for n_h are seen to fall in the range 2.3–2.8 for K_0 between 10^6 and 10^8, indicating that protonation of the molecule is highly cooperative (and similar to that in hemoglobin-oxygen binding) and that protonation and conformation processes are strongly coupled. Fig. 3 *d* also shows that cooperativity vanishes at extreme values of K_0 due to the uncoupling of the protonation and folding processes.

Simulations of unfolding

Overview

The simple two-state model presented above is useful in verifying experimental conclusions about the influence of protonation site position as well as site-site interactions on $\mathrm{pK_a}$ values. The calculations also demonstrate a strong dependence of the $\mathrm{pK_a}$ values on macromolecular conformation. Furthermore, this approach explains the general behavior of experimental titration curves obtained for triple-helical DNA. The two-state model, however, neglects any conformational difference between different protonation substates and so is unable to shed light on the folding or unfolding pathway, a simple example of which is the opening of a $\mathrm{GC^+}$ Hoogsteen pair preceding, and allowing, cytosine deprotonation (Gueron et al., 1987). Therefore, we have used molecular dynamics simulations to investigate the relationship between protonation and subsequent conformational changes.

Basepair opening

The importance of going beyond a two-state model and considering the effect of protonation or deprotonation on an initial conformation is illustrated by the following example. It has been suggested that deprotonation of cytosine at N3, which forms a Hoogsteen bond with guanine N7 in the triplex, is caused by, rather than causes, disruption of the bond (Leitner et al., 2000). This process can be described by the scheme

$$[\mathrm{CG\cdot C^+}]_{\mathrm{closed}} \overset{K_{\mathrm{open}}}{\rightleftharpoons} [\mathrm{CG\cdots C^+}]_{\mathrm{open}} \overset{K_\mathrm{a}}{\rightleftharpoons} \mathrm{CG\cdots C + H^+} \tag{12}$$

and

$$K_{\mathrm{app}} = K_{\mathrm{open}} K_\mathrm{a}. \tag{13}$$

If basepair opening is a prerequisite for the triplex-duplex transition, it should be taken into account in a description of triplex unfolding. Basepair opening has been studied experimentally by measuring the exchange rate between iminoprotons using NMR (Gueron et al., 1987; Leroy et al., 1988) and has been applied to a 31-mer paperclip DNA molecule (Macaya et al., 1992a), revealing the following features (Powell et al., 2001). First, the opening constants K_{open} obtained for AT Hoogsteen pairs are higher than that for $\mathrm{GC^+}$ pairs, indicating the greater stability of the latter due to background (phosphate) charges stabilizing the positively charged basepair. Second, K_{open} for $\mathrm{GC^+}$ basepairs located in the center of the molecule are lower than those for $\mathrm{GC^+}$ at the ends, showing increased stability at the center of the triplex. This observation is consistent with our calculation of triplex $\mathrm{pK_{intr}}$ values discussed above. The increased stability of the central $\mathrm{GC^+}$ basepairs also results from positional constraints imposed by stacking interactions involving the two neighboring basepairs. This basepair opening has also been observed in MD simulations (Cieplak et al., 1997; van Aalten et al., 1999).

Coupling $\mathrm{pK_a}$ calculations to dynamic structural changes

The two-state model of the triplex-hairpin transition discussed in a previous section cannot provide insight into the dynamic mechanism of the process. To couple protonation with time-dependent structural changes, such as basepair opening, we combined electrostatic $\mathrm{pK_a}$ calculations, as described above, with MD simulations of DNA in an electrolyte plus water bath. To investigate the triplex-hairpin unfolding process we started with the molecule in the fully protonated triplex NMR conformation (denoted τ_4^{NMR}

in Eq. 14 below) and, after adding and relaxing solvent (water) and counterions, performed a 20-ps equilibration run to obtain the starting geometry for the system (τ_4^0). A complete 1-ns trajectory showed the NMR triplex conformation to be stable with structural changes within ~2.8 Å of the initial NMR structure. However, some local conformational changes, e.g., reversible basepair opening, especially for the C_7^+–G_{10} Hoogsteen pair located in the CCCG tetraloop, was observed. Apparent pK_a calculations for the complex (Eq. 7 and Table 2) indicate that at pH 4.5 all Hoogsteen cytosines should be protonated, consistent with the observed NMR structure.

To find a specific conformation for which one might reasonably deprotonate C_7 (or another cytosine), we ran a 1-ns simulation starting from τ_4^0 and analyzed the resulting trajectory by plotting the N3(C)–N7(G) hydrogen bond distances as function of time. A large fluctuation in this distance would indicate when a basepair opened and one of the bases becomes more exposed to solvent. This would suggest when the simulation could be interrupted and the pK_{app} values recalculated to determine if a change in the protonation state of a base was warranted. A proton was therefore removed (concomitant with charge rearrangement; see Methods, above) and a new trajectory generated.

The entire process can be summarized by the scheme

$$\tau_4^{\mathrm{NMR}} \xrightarrow[20\,\mathrm{ps\ equil.}]{\mathrm{MD}} \tau_4^0 \xrightarrow[600\,\mathrm{ps}]{\mathrm{MD}} \tau_4^1 \xrightarrow[-\mathrm{H}^+]{} \tau_3^1 \xrightarrow[100\,\mathrm{ps}]{\mathrm{MD}} \tau_3^2 \xrightarrow[-\mathrm{H}^+]{} \tau_2^2 \times \xrightarrow[300\,\mathrm{ps}]{\mathrm{MD}} \tau_2^3 \xrightarrow[-\mathrm{H}^+]{} \tau_1^3 \xrightarrow[1980\,\mathrm{ps}]{\mathrm{MD}} \tau_1^4 \xrightarrow[-\mathrm{H}^+]{} \tau_0^4 \times \xrightarrow[2000\,\mathrm{ps}]{\mathrm{MD}} \tau_0^5 \longrightarrow \ldots \longrightarrow \delta_0^0, \quad (14)$$

where superscripts denote a particular conformation, subscripts indicate the number of protonated cytosine bases, and τ and δ refer to the triplex and hairpin structures. This alternating MD/PB(pK_a) procedure was continued until the molecule was fully deprotonated. This general approach, for which specifics are related below, was also used to investigate the initial steps in the hairpin-to-triplex folding pathway. Our approach explores the behavior of single MD trajectory and is thus different from the MM/PBSA (Massova and Kollman, 2000) and MCCE (Georgescu et al., 2002) approaches that perform electrostatic calculations on macromolecules in various conformations. A detailed account of the triplex-hairpin unfolding investigation now follows.

Instability and deprotonation of the C^+CCG tetraloop

Analysis of the triplex structure τ_4^0 indicated that all four basepairs containing protonated cytosines, including those in the C^+CCG loop (Fig. 4 *a*), were in the closed state (van Dongen et al., 1999). As mentioned above, analysis of a 1-ns trajectory displayed major structural changes in the C^+CCG tetraloop (Fig. 4 *b*) with the N3(C_7^+)–N7(G_{10}) and N3(C_7^+)–O6(G_{10}) distances (Fig. 4 *c*) undergoing a significant increase of ~7 Å first at 400 ps and again at 600 ps. A comparison of snapshots of the initial conformation (τ_4^0, almost identical to τ_4^{NMR} displayed in Fig. 4 *a*) and that after 600 ps (τ_4^1) showed a distinct opening of the C_7^+–G_{10} Hoogsteen basepair (Fig. 4 *b*). Analysis of the N3–N7 distances in the CG·C^+ triads in the triple-helical region also illustrated the instability of this region. Poisson-Boltzmann calculations performed on the τ_4^1 structure gave a pK_{app} of 6.7 for site N3(C_7^+), the lowest value among all protonated

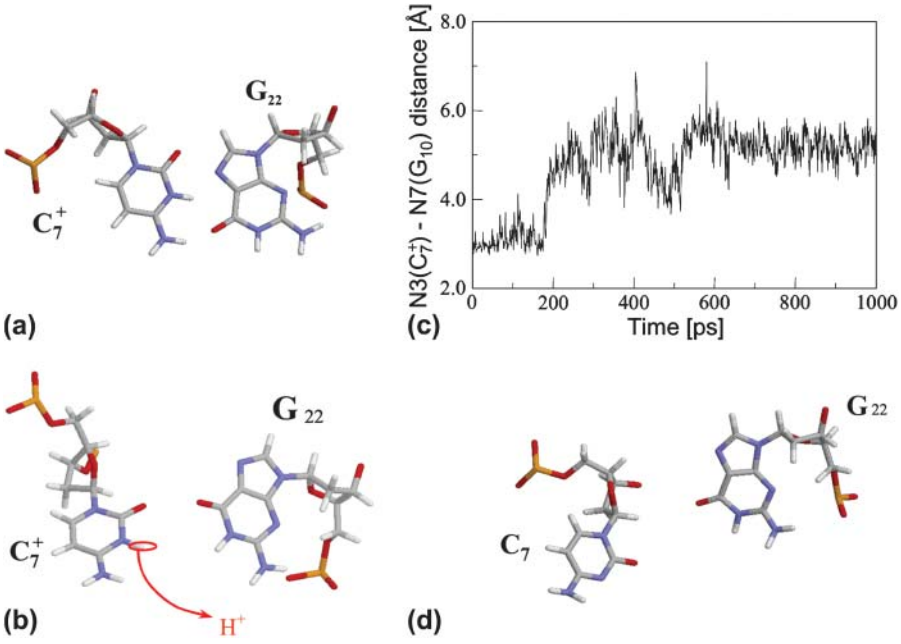

FIGURE 4 (*a*) The NMR conformation of C_7^+–G_{10} in the C^+CCG tetraloop (τ_4^{NMR}) of the triplex. (*b*) Snapshot of the protonated C^+CCG tetraloop at 600 ps (τ_4^1). (*c*) The N3(C_7^+)–N7(G_{10}) distance during a 1-ns MD simulation after a 20-ps equilibrium run on τ_4^{NMR} (that is, starting from conformation τ_4^0). (*d*) Snapshot of the deprotonated CCCG tetraloop after an additional 1-ns simulation (that is, τ_3^1+1 ns) to illustrate the separation between bases. Conformational notation follows that of Eq. 14 for this figure and for Figs. 5–8.

cytosines (Table 3). The C_7^+–G_{10} basepair opening indicated by the large fluctuation in the N3–N7 distance would expose the cytosine N3 atom to solvent, deprotonating it based on its calculated pK_{app} value.

Deprotonation of N3(C_7) (to yield conformation τ_3^1), along with the addition of an equilibrated sodium counterion to maintain system electroneutrality, followed by a 1-ns simulation, revealed that the CCCG tetraloop expanded during the first hundred picoseconds due to electrostatic repulsion between N3(C_7) and the N7 and O6 atoms of G_{10}. This repulsion led to tilting of G_{10} as well as severe changes in the backbone dihedral angles during the first 100 ps of the trajectory; a snapshot of basepair C_7–G_{10} of the CCCG loop conformation obtained after 1 ns is shown in Fig. 4 *d*. An extremely interesting feature of C_7–G_{10} is the switch between Hoogsteen pairing in the fully protonated triplex and Watson-Crick pairing in the fully deprotonated hairpin duplex (van Dongen et al., 1999). Although we were unable to observe the flipping of the G_{10} base necessary for this switch to occur, either during this or later trajectories, undoubtedly due to the limited nanosecond time regime investigated by the dynamics, we feel that the noted base tilting and backbone angle changes strongly hint at the instability of the loop.

Deprotonation of C_{22}^+

The conformational changes induced by deprotonation of C_7^+ are not confined to the CCCG loop but are transmitted to the rest of the molecule. The result of these structural changes, as well as the influence of solvent interactions, is to affect the N3–N7 distances of the remaining protonated G–C^+ Hoogsteen pairs. Fig. 5 displays these distances for the G_{16}–C_{22}^+ (Fig. 5 *a*) and G_{12}–C_{18}^+ (Fig. 5 *b*) pairs. After 100 ps the N3–N7 distance in the G_{16}–C_{22}^+ pair, adjacent to the TTTT tetraloop at the top of the molecule (Fig. 2 *a*), increased significantly from 2.9 Å to ~5.0 Å (Fig. 5 *a*). Snapshots of the G_{16}–C_{22}^+ pair at the beginning of the trajectory (τ_3^1) and after 100 ps (τ_3^2) are shown in Fig. 5, *c* and *d*, respectively, and show that base C_{22}^+ has opened out into the solvent. In contrast, the corresponding distance in the G_{12}–C_{18}^+ basepair temporarily increased much less, from 2.8 Å to 3.2 Å, returning to its initial value at 250 ps (Fig. 5 *b*). The same parameter for the G_{15}–C_{21}^+ basepair remained unchanged, fluctuating around its initial value of 2.9 Å. As a control, we monitored these distances in the initial system with C_7 protonated and found that they fluctuated around their initial values with no significant changes. These observations strongly suggest that deprotonation of C_7 not only induces changes in the tetraloop conformation but is also responsible for structural changes throughout the entire molecule and is a key event in initiating the subsequent transformation toward full deprotonation and the unfolded hairpin duplex state (van Dongen et al., 1999). In a study of the reactivity of intermolecular triplexes, Shimizu et al. (1994) showed that a triad (either TA·T or CG·C^+) adjacent to a loop displayed increased reactivity to OsO_4 and so concluded that this is an unstable region. This would explain an enhanced sensitivity of G_{16}–C_{22}^+ to local structural perturbations.

TABLE 3 Apparent pK_a values for site N3 of the protonatable cytosine bases for the triplex conformation during unfolding (at pH 7.0)

Site	τ_4^{NMR}	τ_4^0	τ_4^1	τ_3^1	τ_3^2	τ_2^2	τ_2^3	τ_1^3	τ_1^4	τ_0^4
C_{18}	8.1	8.0	8.0	8.1	7.7	7.7	7.6	**7.6**	**6.5**	**6.6**
C_{21}	8.2	7.9	8.2	8.2	7.7	7.8	8.8	8.9	7.4	**7.4**
C_{22}	7.9	8.2	8.3	8.3	7.2	**7.2**	**6.8**	**6.8**	**6.1**	**7.2**
C_7	7.5	7.4	6.7	**6.7**	**5.5**	**5.5**	**5.4**	**5.6**	**5.9**	**5.9**

The pK_{app} values for the unprotonated sites are shown in bold (the notation corresponds to that in Eq. 14).

The energy of deprotonation of C_7^+ induces conformational changes in the tetraloop. Relaxation of the loop in turn leads to energy transfer either along the dihedral angles of the sugar-phosphate backbones or through the base triads by altering the helical parameters, or both. Full rotation about the most flexible dihedral angles α, β, and ξ is capable of transferring ~2.5–2.8 kcal/mol, as found from the AMBER ff94 force field (Cornell et al., 1995). These values correspond to the maximal energy transfer through the sugar-phosphate backbone. The force constants for slide and shift, the two principal helical degrees of freedom for stacked Watson-Crick bases (Gardiner et al., 2003), are in the range 1–2 kcal/(mol per $Å^2$) (Olson et al., 1998). Analysis of the same trajectory showed the largest change occurred for the slide parameter, which is a translation of one basepair with respect to its neighbor toward the backbone (shift is a translation perpendicular to slide and moves a basepair into the major or minor groove in B-DNA). Fig. 6 displays the slide parameter for the Watson-Crick and Hoogsteen basepairs as a function of time for trajectory τ_3^1 (0–100 ps) before deprotonation of C_{22}^+ and for the first 100 ps after deprotonation (trajectory τ_3^2). (As helical slide has not yet been parameterized for Hoogsteen stacking, Watson-Crick values—see Olson et al., 1998—were used.) During the trajectory the slide changed by 1.5–2.0 Å, corresponding to a basepair-basepair interaction energy of 6–8 kcal/mol. There is a slight indication (Fig. 6) that this structural perturbation propagates primarily along the duplex stem, reaching the TTTT loop at ~70 ps and just before the observed base flipping of C_{22}^+. As the amount of slide energy is much larger than that which could be transmitted along the sugar-phosphate backbone, energy transfer through basepair stacking appears to be responsible for disruption of the weak G_{16}–C_{22}^+ basepair, and leads to conditions for the deprotonation of C_{22}^+.

Returning to the observation concerning the increased G_{16}–C_{22}^+ distance after 100 ps of trajectory τ_3^2, we then extracted the DNA coordinates and recalculated the pK_{app} values (Table 3). The pK_{app} of C_7 in the CCCG tetraloop was

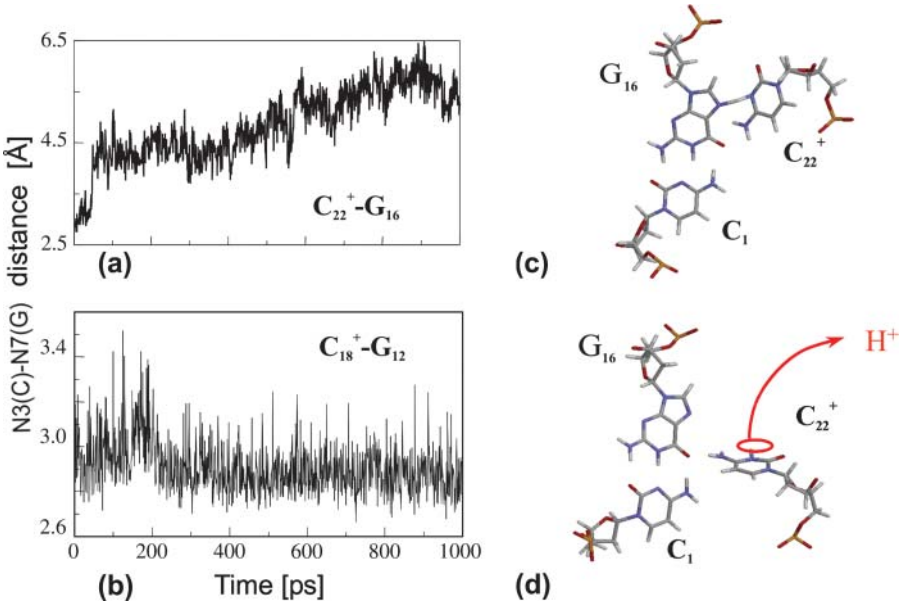

FIGURE 5 The N3–N7 distance for the (*a*) G_{16}–C_{22}^{+} and (*b*) G_{12}–C_{18}^{+} basepairs of the triplex during a 1-ns MD simulation after deprotonation of C_7^{+} (starting from conformation τ_3^1). Snapshots of the C_1–G_{16}–C_{22}^{+} triad at (*c*) 0 ps (τ_3^1) and (*d*) 100 ps (τ_3^2).

found to have been further lowered from 6.7 to 5.5, indicating that reprotonation of C_7 and reformation of the Hoogsteen-paired loop structure is unlikely at this point. The pK_{app} values of the third-strand cytosines also decreased with the largest decrease observed for C_{22}^{+}, from 8.3 to 7.2 and in agreement with the N3–N7 distance analysis discussed earlier. The preference among the three third-strand cytosines for C_{22}^{+} to deprotonate first is not, as might first be thought, due to its position near one end of the molecule where the negative electrostatic potential is weaker. If this were so then the pK_{app} for this base would be the lowest of the three for structure τ_3^1, when in fact it is the highest (or at least equal, within error). Rather this preference is due to basepairing instability near the TTTT loop, as mentioned above. Also, structural changes resulting from deprotonation of the C^{+}CCG loop were transmitted to the other end of the molecule, further destabilizing the G_{16}–C_{22}^{+} basepair. The result after 100 ps of dynamics is a rolling and tilting of C_{22}^{+} from its original position, exposing the protonation site to the solvent (Fig. 5, *c* and *d*). The instability of CG·C^{+} triads adjacent to the TTTT loop has also been inferred from measurements of cytosine pK_{app} values (Leitner et al., 2000). These values were found to be lower by a few units than those measured for cytosines near the interior of DNA structures, suggesting that the H3 atom of a cytosine adjacent to a loop may no longer be involved in Hoogsteen pairing but exposed to solvent.

Deprotonation of C_{18}^{+}

On the basis of the sudden and stable increase in the $N3(C_{22}^{+})$–$N7(G_{16})$ distance after 100 ps of the $\tau_3^1 \rightarrow \tau_3^2$ trajectory, we interrupted the simulation at this point and deprotonated C_{22}^{+} (to give τ_2^2). Recalculated pK_{app} values are given in Table 3. After updating the charges (including an added counterion), the simulation was restarted from these coordinates and a 1-ns trajectory generated. The N3–N7 distances for the two remaining Hoogsteen pairs (G_{12}–C_{18}^{+} and G_{12}–C_{22}^{+}) were then analyzed as before; Fig. 7 *a* displays these data for the G_{12}–C_{18}^{+} pair. The transient instability of G_{12}–C_{18}^{+} observed during the first 200 ps of the previous trajectory (Fig. 5 *b*) is perhaps evident in the more severe fluctuations seen in Fig. 7 *a*. However, an obvious point at which to interrupt the simulation was much more difficult to discern in this case; we chose, somewhat arbitrarily, 300 ps (τ_2^3). Recalculation of the pK_{app} values showed that for this conformation C_{18}^{+} was indeed more likely to be deprotonated than C_{21}^{+}, although most (80%) would remain protonated. The slight conformational changes in the triad coordinates may be gleaned by comparing Fig. 7, *b* (τ_2^2) and *c* (τ_2^3). To model those molecules for which C_{18}^{+} was deprotonated at this point, we updated the charges based on the neutral base (τ_1^3), added a counterion, and generated a 2-ns trajectory, starting at the interrupted coordinates.

Deprotonation of C_{21}^{+}

The remaining protonated Hoogsteen pair is G_{15}–C_{21}^{+} and the N3–N7 distances during the penultimate simulation are shown in Fig. 8 *a*, where a significant increase can be seen at 1980 ps. This conformational change was reversible as long as the protonation state of C_{21}^{+} remained unchanged. Snapshots of the C_2–G_{15}–C_{21}^{+} triad geometry at the start (τ_1^3) and end (τ_1^4) of the run are shown in Fig., 8 *b* and *c*, respectively. To check if the implied base opening was fortuitous, we generated a second trajectory starting from conformation τ_3^2 but with bases C_{18}^{+} and C_{22}^{+} simultaneously deprotonated. A similar increase in the $N3(C_{21}^{+})$–$N7(G_{15})$ displacement was observed at 1450 ps (data not shown). Apparent pK_a values for $N3(C_{21}^{+})$ calculated at the start (τ_1^3)

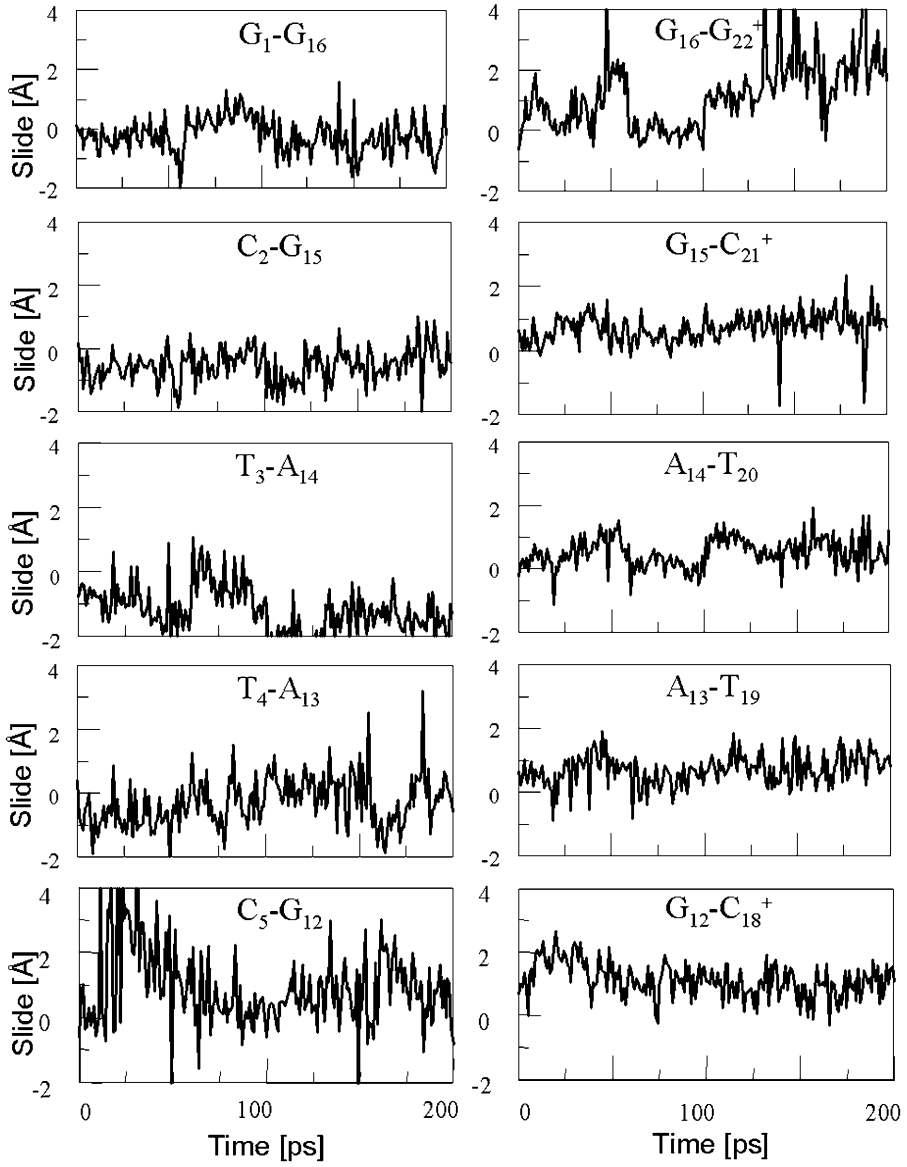

FIGURE 6 The helical parameter slide (in Å) for the indicated basepairs during trajectory τ_3^1 before deprotonation of C_{22}^+ (0–100 ps) and the first 100 ps of trajectory τ_2^2 after deprotonation (100–200 ps).

and end (τ_1^4) of the trajectory are listed in Table 3, with the substantial decrease from 8.9 to 7.4 evident. Also listed are the pK_{app} values after deprotonation of C_{21}^+ (τ_0^4), which shows a unit increase in the pK_{app} for N3(C_{22}) resulting from a strong site-site interaction with neighboring C_{21}. A final 2-ns trajectory was generated after the last deprotonation to verify stability of the fully deprotonated structure. Fig. 9 compares the NMR structure (τ_4^{NMR}, Fig. 9 *a*) with that obtained at the end of the final simulation (τ_0^5, Fig. 9 *b*), illustrating the separation of the third strand as well as expansion of the CCCG tetraloop.

Summary of unfolding

A picture of the initial part of the unfolding dynamics may be obtained by splicing together the separate $\tau_4^0 \rightarrow \tau_4^1/\tau_3^1 \rightarrow \tau_3^2/\tau_2^2 \rightarrow \tau_2^3/\tau_1^3 \rightarrow \tau_1^4/\tau_0^4 \rightarrow \tau_0^5$ segments. The presence of the third strand influences the double-helical region by pushing the Watson-Crick basepairs toward the minor groove. This distortion is characterized by an *x* displacement of -2.5 ± 0.9 Å (versus the averages -5.4 Å and -0.7 Å for the canonical *A* and *B* forms, respectively; Cieplak et al., 1997). Also, the analysis of this super-trajectory indicated that the sugar puckers remain in the range of $126 \pm 23°$ (versus $\sim 190°$ and $\sim 12°$), which corresponds to C1′-exo and C2′-endo conformations, and the values of the dihedral angle δ are found to be in a range of $117 \pm 25°$ (versus $\sim 80°$ and $\sim 130°$). These observations suggest that triple-helical DNA assumes a conformation intermediate between the canonical *A* and *B* forms. Our data are in a good agreement with experimental results (Radhakrishnan and Patel, 1994; Macaya et al., 1992b).

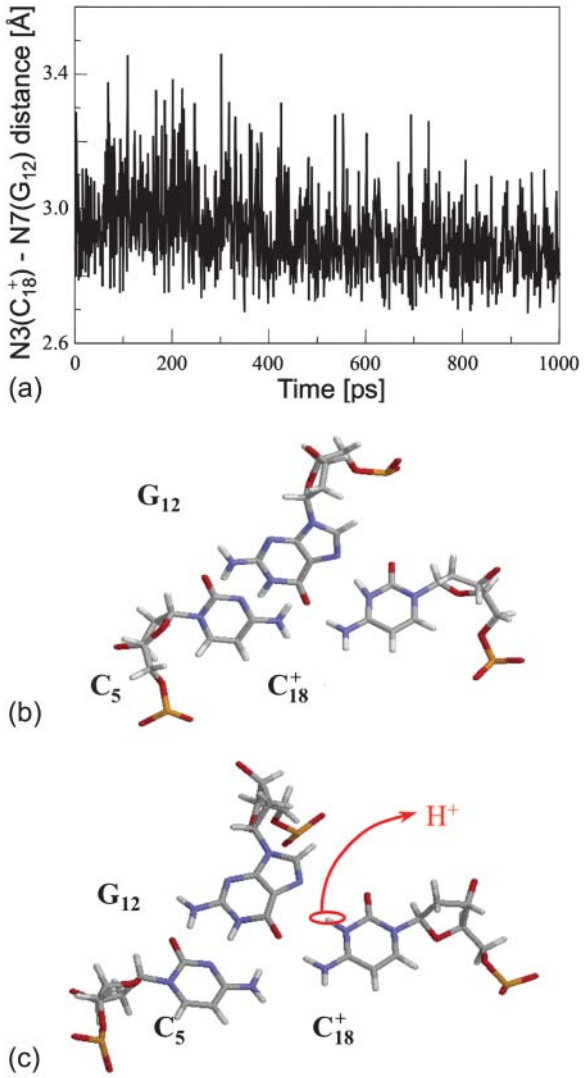

FIGURE 7 (*a*) The N3(C_{18}^+)–N7(G_{12}) distance during a 1-ns MD simulation after deprotonation of C_{22}^+ (starting from conformation τ_2^2). Snapshots of the C_5–G_{12}–C_{18}^+ triad at (*b*) 0 ps (τ_2^2) and (*c*) 300 ps (τ_2^3).

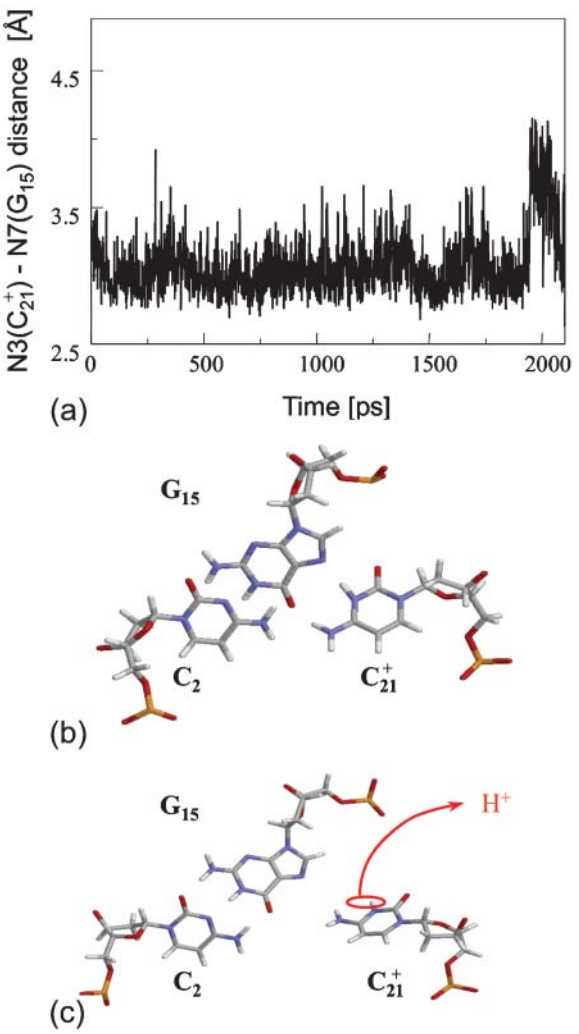

FIGURE 8 (*a*) The N3(C_{21}^+)–N7(G_{15}) distance during a 1-ns MD simulation after deprotonation of C_{18}^+ (starting from conformation τ_1^3). Snapshots of the C_2–G_{15}–C_{21}^+ triad at (*b*) 0 ps (τ_1^3) and (*c*) 1980 ps (τ_1^4).

This combined MD/PB approach suggests that the C^+CCG tetraloop in the particular triplex plays the role of a conformational switch, supporting the contention of van Dongen et al. (1996) Deprotonation of C_7 not only induces local changes in the loop structure but, we believe, also leads to conformational changes in the helical stem. The data show that the G_{16}–C_{22}^+ Hoogsteen pair is much less stable than either the G_{15}–C_{21}^+ and G_{12}–C_{18}^+ pairs and that prior deprotonation of C_7^+ could also help influence G_{16}–C_{22}^+ to open first. Subsequent deprotonation of C_{22}^+ then aids the opening first of G_{15}–C_{18}^+ and then of G_{15}–C_{21}^+, with their loss of protons. Deprotonation of these four cytosines is seen as cooperative in the sense that structural changes induced by Hoogsteen pair repulsion after loss of the binding proton as well as site-site interactions play lead roles. In fact, the presence of the latter accounts for the nonzipper order of basepair opening.

The simulation of unfolding was not continued after the fourth and final cytosine deprotonation as it appeared, as indicated by a few sample trajectories that full or even partial extension of the third strand was now governed by electrostatic and solvent-mediated effects in the diffusive time regime and, as such, were too long for the nanosecond dynamics investigated here. To obtain a more nearly complete view of the entire pathway, we therefore repeated the above MD/PB procedure but from the point of view of folding by beginning with the unprotonated hairpin duplex structure of Fig. 2 *c*.

Dynamics simulations of folding

Summary of folding

The main question we wanted to address in simulating the folding pathway is whether the sequence of cytosine pro-

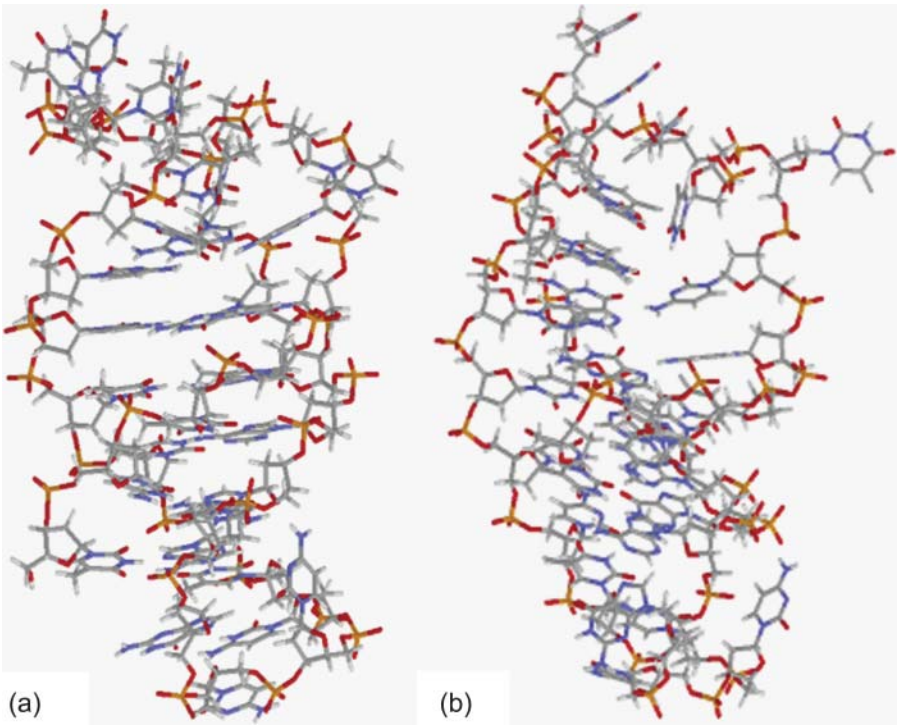

FIGURE 9 (*a*) The NMR triplex conformation τ_4^{NMR} compared with (*b*) the final conformation τ_0^5 obtained after a 2-ns simulation on the fully deprotonated structure τ_0^4.

tonation would be the reverse of that found for deprotonation during unfolding. We were also interested in the stability of the CCCG tetraloop. The folding simulation began with the hairpin duplex coordinates used for the equilibrium studies (labeled δ_0^0 in Eq. 15 below and shown as a cartoon in Fig. 10 *a* and in detail in Fig. 2 *c*). In analogy with the unfolding scheme shown in Eq. 14, we give that for folding

$$\begin{aligned}\delta_0^0(10a) &\xrightarrow[+\,3H^+]{} \delta_3^0(10a)\\ &\times \xrightarrow[3\,ns]{MD} \delta_3^1(10b)\longrightarrow\delta_3^2(10c)\longrightarrow\delta_3^3(10d)\\ &\times \xrightarrow[165\,ps]{MD} \delta_3^4 \xrightarrow[+\,H^+]{} \delta_4^4\longrightarrow\ldots\longrightarrow\tau_4^0,\end{aligned} \tag{15}$$

where the notation in parentheses refers to the cartoons of Fig. 10. Calculated pK_{app} values for particular conformations of the hairpin are listed in Table 4. These show that, for the extended hairpin structure (δ_0^0), all cytosines remain unprotonated at neutral pH, hence implying the stability of the hairpin duplex under this condition. Lowering the pH to 4.5 to initiate folding, results in the protonation of all cytosines in the extended third strand. However, C_7 in the CCCG tetraloop would remain unprotonated (67%) and retain its Watson-Crick pairing with G_{10}. Because protonation of most of the titratable sites occurs immediately upon lowering of the pH (the three sites are already exposed to solvent), we may already conclude that the sequence of events involved in folding do not mirror those of unfolding.

To investigate the effect of third-strand protonation on the folding dynamics, we ran a 3-ns MD simulation starting from conformation δ_3^0, at the end of which the extended third strand of the molecule was bent by ~90° with respect to the double-helical region (δ_3^1 and shown in Fig. 10 *b*). To verify that this bend was due to reduced electrostatic repulsion between the extended strand and the duplex, we also generated a 2-ns trajectory based on the fully unprotonated structure δ_0^0 and observed no significant global conformational change during the run. Analysis of the helix parameters for the δ_0^0 duplex region showed the *x* displacement to lie in the range -1.6 ± 0.8 Å. This value is more positive than that calculated for the folded triple-helical structure and indicates that the hairpin conformation at neutral pH is closer to the canonical B-form of DNA than the triplex structure. However, this value, along with data of the time-averaged sugar puckers (118 ± 34°) and dihedral angles (123 ± 21°), still suggests that the hairpin structure lies between the *A* and *B* forms.

As complete protonation of the extended strand followed by three nanoseconds of simulation had moved the pathway into the diffusive regime, we suspended the MD/PB procedure at this point. However, the results of the method have allowed us to elicit the essential features of the folding mechanism. To summarize, the unprotonated hairpin is stable at neutral pH. Lowering of the pH to 4.5 immediately leads to protonation of all extended-strand cytosines and a reduction in the electrostatic repulsion between this strand and the duplex. This decrease appears to be enough for the extended arm to bend approximately at right angle to the

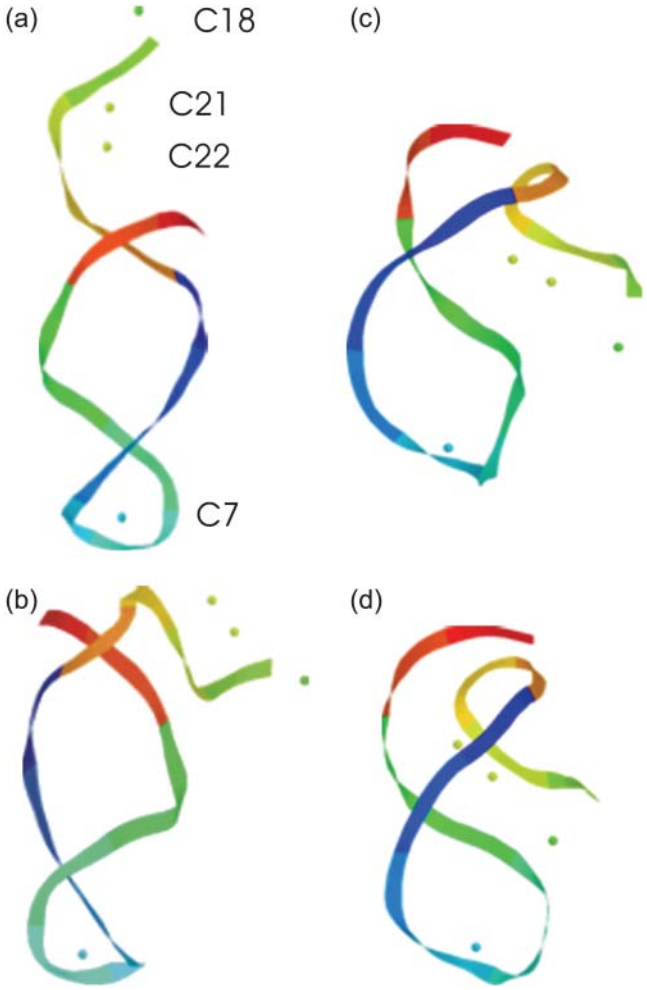

FIGURE 10 The artificially generated (*a*) hairpin and (*b–d*) partially folded hairpin conformations used to investigate the folding pathway. The hairpin conformation (*a*, δ_0^0) was constructed as described in the text in connection with Fig. 2 *c*; conformation (*b*, δ_3^1) was obtained by protonating the three cytosines on the extended arm of the hairpin and running a 3-ns MD simulation; conformations (*c*, δ_3^2) and (*d*, δ_3^3) were artificially constructed from the hairpin conformation by changing dihedral angles in the TTTT tetraloop and equilibrating the structures (see the text for additional details).

duplex, at which point diffusive fluctuations cause it to approach closer to the duplex, ultimately providing a conformation suitable for Hoogsteen pairing between bases of the third strand and the duplex to occur. Before or after this pairing, G_{10} in the CCCG loop flips over and Hoogsteen-pairs with a protonated C_7^+. To gain insight into the sequence of events involved in the reorganization of the CCCG loop, we performed several additional MD simulations which we now describe.

Conformational changes in the CCCG tetraloop

To investigate the stability of the unprotonated CCCG loop in the Watson-Crick conformation, we artificially created

TABLE 4 Apparent pK_a values for site N3 of the protonatable cytosine bases for the hairpin during folding (at pH 4.5)

Site	δ_0^0 (10*a*)	δ_3^0 (10*a*)	δ_3^1 (10*b*)	δ_3^2 (10*c*)	δ_3^3 (10*d*)	δ_3^4	δ_4^4
C_{18}	5.2	**5.1**	**5.2**	**5.3**	**6.3**	**6.5**	**6.4**
C_{21}	5.4	**4.7**	**5.1**	**5.4**	**5.8**	**5.9**	**5.9**
C_{22}	5.5	**4.9**	**5.6**	**6.0**	**6.3**	**6.4**	**6.4**
C_7	4.2	4.2	4.2	4.3	4.3	4.4	4.4

The pK_{app} values for the protonated sites are shown in bold (the notation corresponds to that in Eq. 15).

two hairpin structures intermediate between the triply-protonated δ_3^1 conformation and the NMR triplex (τ_4^{NMR}) in the same manner as the extended hairpin δ_0^0 was made (including 20 ps of MD equilibration). These conformations, δ_3^2 and δ_3^3, in which the angles between the third strand and the duplex are ~20° and 5°, are shown in Fig. 10, *c* and *d*, respectively, and were intended to model the later stages of folding. In the sequence $\{\delta_3^0, \delta_3^1, \delta_3^2, \delta_3^3\}$ of comparative structures, C_7–G_{10} forms a Watson-Crick pair and the third-strand bases C_{18}^+, C_{21}^+, and C_{22}^+ are protonated. Calculated pK_{app} values for these conformations are listed in Table 4 and show that bending the protonated third strand with respect to the duplex from 180° (δ_3^0) to 90° (δ_3^1) to 20° (δ_3^2) has little effect on the proclivity for N3(C_7) to protonate. A 1-ns MD simulation on conformation δ_3^2 also showed the C_7–G_{10} pair to remain intact in the Watson-Crick bonding pattern with a final, essentially unchanged pK_{app} for N3(C_7) of 4.23. From these data we infer that it is only when the extended strand approaches to <5° of the duplex that N3(C_7) is protonated at pH values near 4.5.

We therefore focused more closely on a comparison between the CCCG loop structure in the unprotonated hairpin conformation (δ_0^0) and that with the 5° third-strand-duplex bend (δ_3^3). Two 1-ns MD trajectories were generated, one starting from the unprotonated extended conformation (δ_0^0) and a second starting from the triply-protonated, partially-bent conformation (δ_3^3) and the N3(C_7)–N1(G_{10}) distances compared (Fig. 11, *a* and *b*). For the latter trajectory the analysis of the data disclosed large fluctuations, indicating that H-bonds in the C_7–G_{10} pair had become weaker. This results from the approach of T_{17} to basepair T_6–A_{11} adjacent to pair C_7–G_{10}, inducing the former to open by a few degrees (Fig. 11 *c*). This then leads to an unwinding of the double-helical fragment as shown by a change in the twist measured for C_7–G_{10} relative to T_6–A_{11} (Fig. 11 *d*), consequently weakening the H-bonds in the CCCG loop. In the first significant fluctuation during the δ_3^3 trajectory, this distance has increased from 2.95 Å to 3.17 Å by 165 ps (Fig. 11 *b*), compared to the average value of 2.96 Å for the δ_0^0 (Fig. 10 *a*). At this point the distance decreases back down, only to repeat this cycle ~160 ps later. As seen in Fig. 10 *d*, there is some correlation with the change in twist angle. Extracting the coordinates of the conformation at 165 ps

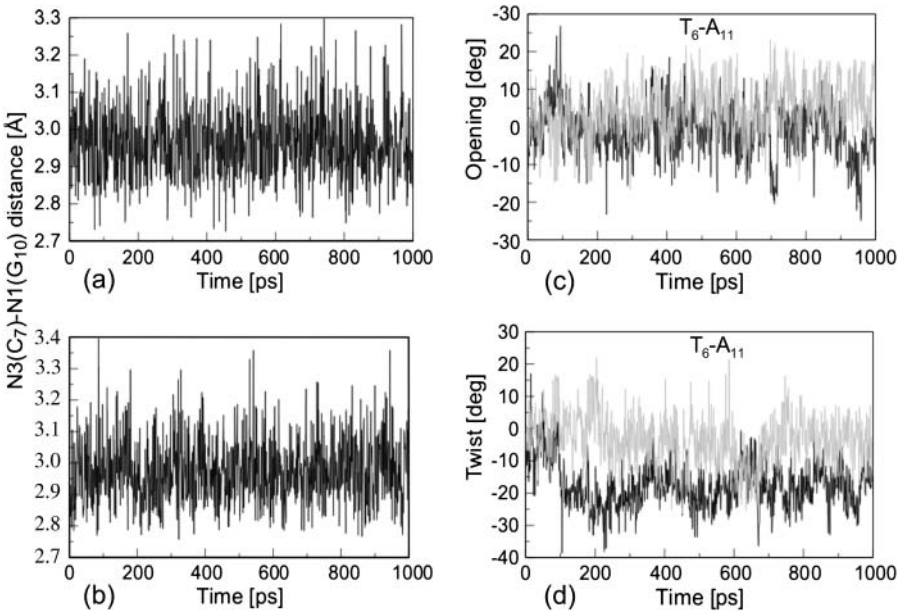

FIGURE 11 The N3(C_7)–N7(G_{10}) distance during a 1-ns MD simulation of (*a*) the unprotonated extended hairpin (starting from conformation δ_0^0) and (*b*) the triply-protonated, partially bent hairpin (starting from conformation δ_3^3). Also shown are comparisons of the two helix parameters (*c*) opening (symbol σ in Lavery and Sklenar, 1989) and (*d*) twist (symbol Ω in Lavery and Sklenar, 1989) of the T_6–A_{11} pair during these two simulations (δ_0^0 in *black* and δ_3^3 in *shading*).

(δ_3^4) provided a slight increase in the pK_{app} for N3(C_7) to 4.4. Protonating this structure (δ_4^4) and running a 2-ns MD simulation revealed a partial rearrangement of the C^+CCG loop (G_{10} flips out of the stack, but does not turn over); however, the complete Watson-Crick-to-Hoogsteen transition for C_7^+–G_{10} was not observed.

From these observations, and in agreement with the sequence of events observed in the folding study discussed earlier, the data suggest that transformation from an unprotonated Watson-Crick C_7–G_{10} pair to a protonated Hoogsteen C_7^+–G_{10} pair is probably coupled to structural changes induced by binding of the third strand to the duplex. This would suggest that although deprotonation of the C^+CCG tetraloop might act as a conformational switch during unfolding of the triplex by inducing the third strand to dissociate from the duplex, binding of the third strand to the duplex during folding is the *switch* that induces opening of C_7–G_{10}, leading to protonation of C_7 and rebinding with G_{10} as a Hoogsteen pair. This is, of course, speculative but consistent with the data.

Van Dongen et al. (1996) have hypothesized two possible mechanisms for the formation of the C_7^+–G_{10} basepair. According to their first mechanism the Watson-Crick-to-Hoogsteen transition results directly from the lowering of the pH and is subsequently followed by protonation of C_7. In this case the conformational changes in the CCCG loop would be independent of third-strand folding. Our dynamics data tend to discount this. Alternatively, they suggested that the transition might be mediated by conformational changes in the duplex that are induced by folding of the third strand, in agreement with the view posited above. However, longer timescale studies appear necessary to clarify this picture.

DISCUSSION

The hairpin/triplex transition of DNA is a complex process that is driven by the exchange between the electrostatic energy of protonation/deprotonation and the internal (bond) energy of the macromolecule, aided by interactions with the solvent. A simple two-state equilibrium model is useful in explaining the general behavior of the experimental titration curves and can in fact be used to provide an estimate of the free energy difference between the unprotonated (or protonated) endstates. This model also shows that protonation and folding of this H-DNA oligomer are coupled and highly cooperative. The observation that evolutionary processes commonly take advantage of cooperativity lends support for a biological role for H-DNA.

A more detailed approach is required to investigate intermediate states along the folding pathway and the combined MD/PB method used here seems noteworthy. Of particular interest is the asymmetry in the protonation/deprotonation sequence for the pH-induced folding/unfolding of this molecule. Also, it appears that the C^+CCG tetraloop plays a central role in the unfolding transition. Containing the first cytosine to deprotonate, the loop may act as a conformational switch, initiating structural changes throughout the molecule, thus inducing deprotonation and dissociation of the third strand in preparation for unfolding. Perhaps a better description of the CCCG loop would be as a *conformational lock*. Only when the third strand is protonated and in close proximity to the duplex, and perhaps Hoogsteen-paired with it, does it appear that the CCCG loop protonates, with the guanine base flipping from its Watson-Crick position to form a Hoogsteen pair, essentially locking

the entire structure in place. Unfolding the molecule then requires that the structure first be unlocked through deprotonation of the C^+CCG loop.

Finally, we make one comment regarding the lack of a C^+CCG tetraloop in H-DNA for which the system studied here was used as a model (van Dongen et al., 1996). The effect of a conformational change in the tetraloop leading to deprotonation of the first cytosine in the third strand (C_{22}^+) may be viewed as a structural perturbation due to the binding by, or at least influence of, another molecule (protein, drug, or cation) six-basepairs distant, and may thus not be entirely irrelevant to the regulatory role of H-DNA in transcription and replication.

In summary, the complementary use of a two-state equilibrium model and a combined MD/PB dynamic approach has provided a fairly detailed picture of this simple yet rich system. Extension of these techniques to similar problems such as the magnesium-induced folding of RNA should be equally rewarding.

APPENDIX

Consider a molecule (DNA) that can assume two conformations, which we designate as τ (triplex) and δ (duplex). Consider also that this molecule and hence each conformation possesses N (necessarily nonidentical) protonatable sites. We identify the *protonation state* of a conformation by the total number of sites which are protonated, thus giving $N + 1$ possible states for each conformation and ranging from fully unprotonated to fully protonated. Within each of these protonation states are $N_i \equiv (N!)/((N-i)!i!)$ *protonation substates* determined by the specific sites that are protonated. The τ–δ conformational transition can be described by the two-state equilibrium constant

$$K_{\tau-\delta} = \frac{[\Delta]}{[T]} = \frac{\sum_{i=0}^{N}\sum_{j=1}^{N_i}[\delta_{ij}]}{\sum_{i=0}^{N}\sum_{j=1}^{N_i}[\tau_{ij}]}, \tag{A1}$$

where Δ and T are total concentrations of duplex and triplex, and i and j indicate the protonation state and substate of the particular conformation, respectively. To express $K_{\tau-\delta}$ in terms of pH, we require the free energy change as a function of pH between the fully unprotonated state ($\delta_0 \equiv \delta_{01}$; no second subscript is necessary since this state is unique) and a particular protonation substate. (We could also start from the fully protonated state δ_4.) That is, we want the equilibrium constant K_{ij}^{δ} for the protonation scheme

$$\delta_0 + i\mathrm{H}^+ \overset{K_{ij}^{\delta}}{\rightleftharpoons} \delta_{ij} \tag{A2}$$

or

$$K_{ij}^{\delta} = \frac{[\delta_{ij}]}{[\mathrm{H}^+]^i[\delta_0]}, \tag{A3}$$

as well as the corresponding expression for the τ-conformation. Assuming that no structural changes contribute to the free energy upon protonation, the Bashford-Karplus method can be used to obtain the equilibrium constant for reaction A2 (Bashford and Karplus, 1990):

$$K_{ij}^{\delta} = \frac{1}{[\mathrm{H}^+]^i}\exp\left[2.3\sum_{\mu} x_{ij}^{\mu}(\mathrm{p}K_{\mathrm{intr},\mu}^{\delta} - \mathrm{pH}) - \frac{\beta}{2}\sum_{\mu\nu} x_{ij}^{\mu}x_{ij}^{\nu}W_{\mu\nu}^{\delta}\right], \tag{A4}$$

where x^{μ} denotes a protonation state ''vector'' for a site μ which takes the value 1 or 0 depending on whether substate ij of site μ is protonated or not. Equation A4 also contains a correction term $W_{\mu\nu}$ ($W_{\mu\mu} \equiv 0$) that accounts for the interaction between mutually protonated sites in a substate. Combining Eqs. A1, A3, and A4, along with the corresponding expression for the τ-conformation, gives the desired result for the τ–δ equilibrium constant of

$$K_{\tau-\delta} = K_0 K(\mathrm{pH}), \tag{A5}$$

where

$$K_0 = \frac{[\delta_0]}{[\tau_0]} \tag{A6}$$

is the high-pH limit of $K_{\tau-\delta}$ corresponding to the free energy change ($\beta\Delta G_0 = -2.3\ \mathrm{pK_0}$) for the transition between unprotonated conformations, and the pH-dependent part of the equilibrium constant is contained in the quantity

$$K(\mathrm{pH}) = \frac{\sum_{i,j}\exp\left[2.3\sum_{\mu} x_{ij}^{\mu}(\mathrm{p}K_{\mathrm{intr},\mu}^{\delta} - \mathrm{pH}) - \frac{\beta}{2}\sum_{\mu\nu} x_{ij}^{\mu}x_{ij}^{\nu}W_{\mu\nu}^{\delta}\right]}{\sum_{i,j}\exp\left[2.3\sum_{\mu} x_{ij}^{\mu}(\mathrm{p}K_{\mathrm{intr},\mu}^{\tau} - \mathrm{pH}) - \frac{\beta}{2}\sum_{\mu\nu} x_{ij}^{\mu}x_{ij}^{\nu}W_{\mu\nu}^{\tau}\right]}. \tag{A7}$$

The equilibrium constant has the obvious limits corresponding to the fully protonated and unprotonated states of

$$\lim_{\text{low pH}} K_{\tau-\delta} = \frac{[\delta_4]}{[\tau_4]} < \lim_{\text{high pH}} K_{\tau-\delta} = \frac{[\delta_0]}{[\tau_0]}. \tag{A8}$$

Proton exchange NMR measurements can be used to obtain $K_{\tau-\delta}$ at different pH. The apparent equivalence point is the pH at which the triplex and duplex concentrations are equal, and corresponds to the condition $K_{\tau-\delta} = 1$. By using Eqs. A1 and A5, and along with the conservation requirement

$$\Delta + T = C_{\mathrm{DNA}}, \tag{A9}$$

where C_{DNA} is the total concentration of DNA, the total concentration of each conformation is given by

$$\Delta(\mathrm{pH}) = \frac{C_{\mathrm{DNA}}K_0K(\mathrm{pH})}{1 + K_0K(\mathrm{pH})} \tag{A10}$$

and

$$T(\mathrm{pH}) = \frac{C_{\mathrm{DNA}}}{1 + K_0K(\mathrm{pH})}. \tag{A11}$$

For a measured property of the system that depends linearly on the concentrations of the conformations *and is independent of protonation substates*, such as the absorbance A, we can write

$$A = \varepsilon_{\delta}\Delta + \varepsilon_{\tau}T, \tag{A12}$$

where ε_{δ} and ε_{τ} are the extinction coefficients of the duplex and triplex conformations, respectively, and are assumed to be independent of pH. Equations A10–A12 allow us to express the absorbance in terms of pH:

$$A(\mathrm{pH}) = C_{\mathrm{DNA}}\frac{\varepsilon_\tau + \varepsilon_\delta K_0 K(\mathrm{pH})}{1 + K_0 K(\mathrm{pH})}. \tag{A13}$$

The low- and high-pH limits of Eq. A8 show that the absorbance curve displays the expected sigmoidal shape. Using Eq. A13 to fit an absorbance curve allows one to determine the sole fitting parameter K_0 and hence the free energy for the unprotonated conformational transition.

A similar analysis can be used to calculate the Hill coefficient (n_h) describing the cooperativity of the transition. The Hill coefficient is given by the slope of $-\log(Y/(1-Y))$ versus pH at $Y = 0.5$, where Y is the protonated fraction of DNA (Voet and Voet, 1995). Following the notation above, the protonated fraction is given by

$$\begin{aligned} Y &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N_i}[\delta_{ij}] + \sum_{i=1}^{N}\sum_{j=1}^{N_i}[\tau_{ij}]}{\delta_0 + \sum_{i=1}^{N}\sum_{j=1}^{N_i}[\delta_{ij}] + \tau_0 + \sum_{i=1}^{N}\sum_{j=1}^{N_i}[\tau_{ij}]} \\ &\approx \frac{\sum_{i=1}^{N}\sum_{j=1}^{N_i}[\delta_{ij}] + \sum_{i=1}^{N}\sum_{j=1}^{N_i}[\tau_{ij}]}{\delta_0 + \sum_{i=1}^{N}\sum_{j=1}^{N_i}[\delta_{ij}] + \sum_{i=1}^{N}\sum_{j=1}^{N_i}[\tau_{ij}]}, \end{aligned} \tag{A14}$$

with the second line of Eq. A14 following from $K_0 = \delta_0/\tau_0 \gg 1$. Use of Eqs. A3, A4, and A6 then give

$$\begin{aligned} \frac{Y}{1-Y} = &\sum_{i,j}\exp\left[2.3\sum_{\mu} x_{ij}^{\mu}(\mathrm{p}K^{\delta}_{\mathrm{intr},\mu} - \mathrm{pH}) - \frac{\beta}{2}\sum_{\mu\nu} x_{ij}^{\mu}x_{ij}^{\nu}W^{\delta}_{\mu\nu}\right] \\ &+ K_0^{-1}\sum_{i,j}\exp\left[2.3\sum_{\mu} x_{ij}^{\mu}(\mathrm{p}K^{\tau}_{\mathrm{intr},\mu} - \mathrm{pH}) - \frac{\beta}{2}\sum_{\mu\nu} x_{ij}^{\mu}x_{ij}^{\nu}W^{\tau}_{\mu\nu}\right], \end{aligned} \tag{A15}$$

which, when plotted as $-\log(Y/(1-Y))$ versus pH, yields the Hill coefficient.

We thank Professors S.S. Wijmenga, C.W. Hilbers, and A.N. Lane for helpful insights and discussions.

This work was supported in part by National Institutes of Health grant GM29079.

REFERENCES

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* 238:415–436.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1996. The determinants of pK_as in proteins. *Biochemistry.* 35:7819–7833.

Arnott, S., and E. Selsing. 1974. Structures for polynucleotide complexes poly(dA)·poly(dT) and poly(dT)·poly(dA)·poly(dT). *J. Mol. Biol.* 88:509–521.

Asensio, J. L., A. N. Lane, J. Dhesi, S. Bergqvist, and T. Brown. 1998. The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J. Mol. Biol.* 275:811–822.

Bashford, D., and M. Karplus. 1990. pK_as of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry.* 29: 10219–10225.

Bashford, D., and M. Karplus. 1991. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* 95:9556–9561.

Basye, J., J. O. Trent, D. Q. Gao, and S. W. Ebbinghaus. 2001. Triplex formation by morpholino oligodeoxyribonucleotides in the HER-2/neu promoter requires the pyrimidine motif. *Nucleic Acids Res.* 29:4873–4880.

Brinton, B. T., M. S. Caddle, and N. H. Heintz. 1991. Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA-replication in COS-7 cells. *J. Biol. Chem.* 266:5153–5161.

Brion, P., and E. Westhof. 1997. Hierarchy and dynamics of RNA folding. *Annu. Rev. Bioph. Biom.* 26:113–137.

Chen, F. M. 1991. Intramolecular triplex formation of the purine·purine·pyrimidine type. *Biochemistry.* 30:4472–4479.

Cieplak, P., T. E. Cheatham, and P. A. Kollman. 1997. Molecular dynamics simulations find that 3′ phosphoramidate-modified DNA duplexes undergo a B to A transition and normal DNA duplexes an A to B transition. *J. Am. Chem. Soc.* 119:6722–6730.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

Csaszar, K., N. Spackova, R. Stefl, J. Sponer, and N. B. Leontis. 2001. Molecular dynamics of the frame-shifting pseudoknot from Beet Western Yellow virus: the role of non-Watson-Crick basepairing, ordered hydration, cation binding and base mutations on stability and unfolding. *J. Mol. Biol.* 313:1073–1091.

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N·Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

Davis, M. E., J. D. Madura, B. A. Luty, and J. A. McCammon. 1991. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 62:187–197.

Felsenfeld, G., D. R. Davis, and A. Rich. 1957. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* 79:2023–2024.

Frank-Kamenetskii, M. D., and S. M. Mirkin. 1995. Triplex DNA structures. *Annu. Rev. Biochem.* 64:65–95.

Gardiner, E. J., C. A. Hunter, M. J. Packer, D. S. Palmer, and P. Willett. 2003. Sequence-dependent DNA structure: a database of octamer structural parameters. *J. Mol. Biol.* 332:1025–1035.

Georgescu, R. E., E. G. Alexov, and M. R. Gunner. 2002. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys. J.* 83:1731–1748.

Glikin, G. C., G. Gargiulo, L. Renadescalzi, and A. Worcel. 1983. *Escherichia coli* single-strand binding: protein stabilizes specific denatured sites in superhelical DNA. *Nature.* 303:770–774.

Gluck, A., Y. Endo, and I. G. Wool. 1994. The ribosomal-RNA identity elements for ricin and for α-sarcin: mutations in the putative CG pair that closes a GAGA tetraloop. *Nucleic Acids Res.* 22:321–324.

Gueron, M., M. Kochoyan, and J. L. Leroy. 1987. A single-mode of DNA basepair opening drives imino proton-exchange. *Nature.* 328:89–92.

Hilbers, C. W., M. J. J. Blommers, F. J. M. Vandeven, J. H. van Boom, and G. A. van der Marel. 1991. High-resolution NMR: studies of DNA hairpins with four nucleotides in the loop region. *Nucleos. Nucleot.* 10:61–80.

Htun, H., E. Lund, and J. E. Dahlberg. 1984. Human U1 RNA genes contain an unusually sensitive nuclease S1 cleavage site within the conserved 3′ flanking region. *Proc. Natl. Acad. Sci. USA.* 81:7288–7292.

Jorgensen, W. L., and J. Pranata. 1990. Importance of secondary interactions in triply hydrogen-bonded complexes: guanine-cytosine vs. uracil-2,6-diaminopyridine. *J. Am. Chem. Soc.* 112:2008–2010.

Jucker, F. M., and A. Pardi. 1995. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry.* 34:14416–14427.

Kato, M., and N. Shimizu. 1992. Effect of the potential triplex DNA region on the in vivo expression of bacterial β-lactamase gene in superhelical recombinant plasmids. *J. Biochem.* 112:492–494.

Kiran, M. R., and M. Bansal. 1995. Structural polymorphism in ***d***(T)(12)·D(A)(12)*d*(T)(12) triple helices. *J. Biomol. Struct. Dyn.* 13:493–505.

Lamm, G. 2003. The Poisson-Boltzmann equation. *In* Reviews in Computational Chemistry. K.B. Lipkowitz, R. Larter, and T. R. Cundari, editors. Wiley & Sons, Hoboken, NJ. 147–365. We point out two typographic errors in Lamm, 2003, in Eq. 407; the superscript of the free energy in the first line should read *B* and that of the second line should read *C*. Also, the sign in front of the summation over background charges in Eq. 412 should be positive. Additional errata will appear on the *Reviews in Computational Chemistry* web site.

Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* 6:655–667.

Leitner, D., W. Schroder, and K. Weisz. 1998. Direct monitoring of cytosine protonation in an intramolecular DNA triple helix. *J. Am. Chem. Soc.* 120:7123–7124.

Leitner, D., W. Schroder, and K. Weisz. 2000. Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. *Biochemistry.* 39:5886–5892.

Leroy, J. L., M. Kochoyan, T. Huynhdinh, and M. Gueron. 1988. Characterization of basepair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* 200:223–238.

Li, H., A. W. Hains, J. E. Everts, A. D. Robertson, and J. H. Jensen. 2002. The prediction of protein pK_as using QM/MM: the pK_a of lysine 55 in Turkey ovomucoid third domain. *J. Phys. Chem. B.* 106:3486–3494.

Lim, C., D. Bashford, and M. Karplus. 1991. Absolute pK_a calculations with continuum dielectric methods. *J. Phys. Chem.* 95:5610–5620.

Lyamichev, V. I., S. M. Mirkin, and M. D. Frank Kamenetskii.1985. A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J. Biomol. Struct. Dyn.* 3:327–338.

Macaya, R., E. Wang, P. Schultze, V. Sklenar, and J. Feigon. 1992a. Proton nuclear-magnetic resonance assignments and structural characterization of an intramolecular DNA triplex. *J. Mol. Biol.* 225:755–773.

Macaya, R. F., P. Schultze, and J. Feigon. 1992b. Sugar conformations in intramolecular DNA triplexes determined by coupling constants obtained by automated simulation of P·COSY cross peaks. *J. Am. Chem. Soc.* 114:781–783.

MacKerell, A. D., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975.

Madura, J. D., J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 91:57–95.

Massova, I., and P. A. Kollman. 2000. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov.* 18:113–135.

Michel, F., and E. Westhof. 1990. Modeling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216:585–610.

Mirkin, S. M., V. I. Lyamichev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, and M. D. Frank Kamenetskii. 1987. DNA H-form requires a homopurine homopyrimidine mirror repeat. *Nature.* 330:495–497.

Moore, P. B. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* 68:287–300.

Murphy, F. L., and T. R. Cech. 1994. GAAA tetraloop and conserved bulge stabilize tertiary structure of a group-I intron domain. *J. Mol. Biol.* 236:49–63.

Nicholls, A., and B. Honig. 1991. A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* 12:435–445.

Ojha, R. P., and R. K. Tiwari. 2002. Molecular dynamics simulation study of DNA triplex formed by mixed sequences in solution. *J. Biomol. Struct. Dyn.* 20:107–126.

Olson, W. K., A. A. Gorin, X. J. Lu, L. M. Hock, and V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA.* 95:11163–11168.

Pack, G. R., G. A. Garrett, L. Wong, and G. Lamm. 1993. The effect of a variable dielectric coefficient and finite ion size on Poisson-Boltzmann calculations of DNA-electrolyte systems. *Biophys. J.* 65:1363–1370.

Pack, G. R., L. Wong, and G. Lamm. 1998. pK_a of cytosine on the third strand of triplex DNA: preliminary Poisson-Boltzmann calculations. *Inter. J. Quantum. Chem.* 70:1177–1184.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91:1–41.

Petrov, A. S., G. Lamm, and G. R. Pack. 2005. Calculation of binding free energy for magnesium-RNA interactions. *Biopolymers.* In press.

Plum, G. E., and K. J. Breslauer. 1995. Thermodynamics of an intramolecular DNA triple helix: a calorimetric and spectroscopic study of the pH and salt dependence of thermally induced structural transitions. *J. Mol. Biol.* 248:679–695.

Powell, S. W., L. H. Jiang, and I. M. Russu. 2001. Proton exchange and basepair opening in a DNA triple helix. *Biochemistry.* 40:11065–11072.

Radhakrishnan, I., and D. J. Patel. 1994. DNA Triplexes: solution structures, hydration sites, energetics, interactions, and function. *Biochemistry.* 33:11405–11416.

Rao, B. S., H. Manor, and R. G. Martin. 1988. Pausing in Simian virus-40 DNA: replication by a sequence containing (dG-dA)27·(dT-dC)27. *Nucleic Acids Res.* 16:8077–8094.

Riley, M., B. Maling, and M. J. Chamberlin. 1966. Physical and chemical characterization of two- and three-stranded adenine·thymine and adenine·uracil homopolymer complexes. *J. Mol. Biol.* 20:359–389.

Ripoll, D. R., Y. N. Vorobjev, A. Liwo, J. A. Vila, and H. A. Scheraga. 1996. Coupling between folding and ionization equilibria: effects of pH on the conformational preferences of polypeptides. *J. Mol. Biol.* 264: 770–783.

Roberts, R. W., and D. M. Crothers. 1996. Prediction of the stability of DNA triplexes. *Proc. Natl. Acad. Sci. USA.* 93:4320–4325.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comp. Phys.* 23:327–341.

Sham, Y. Y., Z. T. Chu, and A. Warshel. 1997. Consistent calculations of pK_as of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem. B.* 101:4458–4472.

Shields, G. C., C. A. Laughton, and M. Orozco. 1997. Molecular dynamics simulations of the d(T·A·T) triple helix. *J. Am. Chem. Soc.* 119:7463–7469.

Shimizu, M., J. C. Hanvey, and R. D. Wells. 1989. Intramolecular DNA triplexes in supercoiled plasmids. I. Effect of loop size on formation and stability. *J. Biol. Chem.* 264:5944–5949.

Shimizu, M., K. Kubo, U. Matsumoto, and H. Shindo. 1994. The loop sequence plays crucial roles for isomerization of intramolecular DNA triplexes in supercoiled plasmids. *J. Mol. Biol.* 235:185–197.

Singh, U. C., and P. A. Kollman. 1984. An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* 5:129–145.

Sklenar, V., and J. Feigon. 1990. Formation of a stable triplex from a single DNA strand. *Nature.* 345:836–838.

Soyfer, V., and V. N. Potaman. 1996. Triple-Helical Nucleic Acids. Springer, New York.

Spackova, N., I. Berger, M. Egli, and J. Sponer. 1998. Molecular dynamics of hemiprotonated intercalated four-stranded I-DNA: stable trajectories on a nanosecond scale. *J. Am. Chem. Soc.* 120:6147–6151.

Sponer, J., J. V. Burda, P. Mejzlik, J. Leszczynski, and P. Hobza. 1997. Hydrogen-bonded trimers of DNA bases and their interaction with metal cations: ab initio quantum-chemical and empirical potential study. *J. Biomol. Struct. Dyn.* 14:613–628.

Sugimoto, N., P. Wu, H. Hara, and Y. Kawamoto. 2001. pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. *Biochemistry.* 40:9396–9405.

Sun, J. S., J. L. Mergny, R. Lavery, T. Montenaygarestier, and C. Helene. 1991. Triple helix structures: sequence dependence, flexibility and mismatch effects. *J. Biomol. Struct. Dyn.* 9:411–424.

Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. I. General equations for impenetratable spheres. *J. Am. Chem. Soc.* 79: 5333–5339.

Tanford, C., and R. Roxby. 1972. Interpretation of protein titration curves: application to lysozyme. *Biochemistry.* 11:2192–2198.

van Aalten, D. M. F., D. A. Erlanson, G. L. Verdine, and L. Joshua-Tor. 1999. A structural snapshot of basepair opening in DNA. *Proc. Natl. Acad. Sci. USA.* 96:11809–11814.

van Dongen, M. J. P., J. F. Doreleijers, G. A. van der Marel, J. H. van Boom, C. W. Hilbers, and S. S. Wijmenga. 1999. Structure and mechanism of formation of the H-y5 isomer of an intramolecular DNA triple helix. *Nat. Struct. Biol.* 6:854–859.

van Dongen, M. J. P., S. S. Wijmenga, G. A. van der Marel, J. H. van Boom, and C. W. Hilbers. 1996. The transition from a neutral-pH double helix to a low-pH triple helix induces a conformational switch in the CCCG tetraloop closing a Watson-Crick stem. *J. Mol. Biol.* 263:715–729.

Voet, D., and J. G. Voet. 1995. *Biochemistry*. J. Wiley & Sons, New York.

Volker, J., and H. H. Klump. 1994. Electrostatic effects in DNA triple helices. *Biochemistry.* 33:13502–13508.

Weisz, K., D. Leitner, C. Krafft, and T. Welfle. 2000. Structural heterogeneity in intramolecular DNA triple helices. *Biol. Chem.* 381: 275–283.

Wool, I. G., A. Gluck, and Y. Endo. 1992. Ribotoxin recognition of ribosomal RNA and a proposal for the mechanism of translocation. *Trends Biochem. Sci.* 17:266–269.

You, T. J., and D. Bashford. 1995. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys. J.* 69:1721–1733.

Zhou, H. X., and M. Vijayakumar. 1997. Modeling of protein conformational fluctuations in pK_a predictions. *J. Mol. Biol.* 267:1002–1011.

Zwieb, C. 1992. Recognition of a tetranucleotide loop of signal recognition particle RNA by protein-SRP 19. *J. Biol. Chem.* 267:15650–15656.