

Comparative analysis of inosine-substituted duplex DNA by circular dichroism and X-ray crystallography

Justin P. Peters, Ewa A. Kowal, Pradeep S. Pallan, Martin Egli & L. James Maher III

To cite this article: Justin P. Peters, Ewa A. Kowal, Pradeep S. Pallan, Martin Egli & L. James Maher III (2017): Comparative analysis of inosine-substituted duplex DNA by circular dichroism and X-ray crystallography, Journal of Biomolecular Structure and Dynamics, DOI: [10.1080/07391102.2017.1369164](https://doi.org/10.1080/07391102.2017.1369164)

To link to this article: <https://doi.org/10.1080/07391102.2017.1369164>

 View supplementary material [↗](#)

 Accepted author version posted online: 18 Aug 2017.
Published online: 04 Sep 2017.

 Submit your article to this journal [↗](#)

 Article views: 27

 View related articles [↗](#)

 View Crossmark data [↗](#)



Comparative analysis of inosine-substituted duplex DNA by circular dichroism and X-ray crystallography

Justin P. Peters^a, Ewa A. Kowal^b, Pradeep S. Pallan^b, Martin Egli^b and L. James Maher III^{a*} 

^aDepartment of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, 200 First St. SW, Rochester, MN 55905, USA; ^bDepartment of Biochemistry, Vanderbilt University School of Medicine, 607 Light Hall, Nashville, TN 37232, USA

Communicated by Ramaswamy H. Sarma

(Received 30 June 2017; accepted 4 August 2017)

Leveraging structural biology tools, we report the results of experiments seeking to determine if the different mechanical properties of DNA polymers with base analog substitutions can be attributed, at least in part, to induced changes from classical B-form DNA. The underlying hypothesis is that different inherent bending and twisting flexibilities may characterize non-canonical B-DNA, so that it is inappropriate to interpret mechanical changes caused by base analog substitution as resulting simply from 'electrostatic' or 'base stacking' influences without considering the larger context of altered helical geometry. Circular dichroism spectra of inosine-substituted oligonucleotides and longer base-substituted DNAs in solution indicated non-canonical helical conformations, with the degree of deviation from a standard B-form geometry depending on the number of I-C pairs. X-ray diffraction of a highly inosine-substituted DNA decamer crystal (eight I-C and two A-T pairs) revealed an A-tract-like conformation with a uniformly narrow minor groove, reduced helical rise, and the majority of sugars adopting a C1'-*exo* (southeastern) conformation. This contrasts with the standard B-DNA geometry with C2'-*endo* sugar puckers (south conformation). In contrast, the crystal structure of a decamer with only four I-C pairs has a geometry similar to that of the reference duplex with eight G-C and two A-T pairs. The unique crystal geometry of the inosine-rich duplex is noteworthy given its unusual CD signature in solution and the altered mechanical properties of some inosine-containing DNAs.

Keywords: A-tracts; B-DNA; circular dichroism; DNA topology; inosine; metal ion coordination; sequence-dependent helix geometry; X-ray crystallography

1. Introduction

The earliest X-ray diffraction studies of DNA fibers revealed two different flavors of DNA, an A form found at low water activity (~75% relative humidity) and a B form found at high water activity (Franklin & Gosling, 1953; Watson & Crick, 1953). Subsequent efforts have sought to understand the biological significance of duplex DNA polymorphism. On one level, DNA structural polymorphism manifests because of the ability of identical DNA sequences to favor different forms in different environments (solvents, ions, and ligands), with B-DNA the favored conformation under physiological conditions.

On a second level, different DNA sequences may adopt different preferred conformations under set conditions. The classic example is the left-handed Z-DNA conformation observed for some alternating purine-pyrimidine sequences (Wang et al., 1979). However, the interplay of base sequence, composition, structure, backbone flexibility, counter-ion condensation, and solvation in determining these conformations has been difficult to resolve. In one prominent model, DNA helical structure

is determined by a 'tug of war' between the two grooves for cation localization (Hud & Plavec, 2003; Hud & Polak, 2001). Sequences with one groove that dominates the other for preferential cation localization tend to assume a helical structure other than the canonical B-form (e.g. B'-form or A-form), and sequence-directed axial curvature is connected to sequence-specific cation localization by a 'cation-dependent junction model' (Hud & Plavec, 2003; Hud & Polak, 2001). The interchange of sequence and shape is also key for biological function, e.g. the specific interaction of a protein with DNA requires recognition of the chemical signature of DNA sequence as well as intrinsic DNA structure (El Hassan & Calladine, 1996; Harteis & Schneider, 2014; Okonogi, Alley, Reese, Hopkins, & Robinson, 2000).

On yet a third level, recent work has revealed that B-DNA consists of two distinct nucleotide conformations, B-I and B-II (Imeddourene et al., 2016). The B-II sub-conformation comprises only about 10% of the B-DNA nucleotides in the Protein Data Bank (PDB) (Balaceanu et al., 2017). Yet, these structures differ from canonical B-I structures with respect to inclination and destacking

*Corresponding author. Email: maher@mayo.edu

of the bases, presence of a broad and very shallow major groove due to displacement of the bases towards the major groove, and presence of a deep and narrow minor groove (Hartmann, Piazzola, & Lavery, 1993; van Dam & Levitt, 2000).

Biophysical and structural investigations of chemically modified DNA can provide insight into the underlying forces driving the properties of native DNA. In particular, nucleic acid base analogs have been utilized to investigate DNA mechanics and structure (Berger, Tereshko, Ikeda, Marquez, & Egli, 1998; Ding, Gryaznov, & Wilson, 1998; Heath, Clendenning, Fujimoto, & Schurr, 1996; Peters, Mogil, McCauley, Williams, & Maher, 2014; Peters, Yelgaonkar, Srivatsan, Tor, & Maher, 2013; Tereshko et al., 1998). Dozens of high-resolution structures of chemically modified DNAs have been determined. One well-characterized nucleotide substitution is inosine for guanosine (i.e. the base hypoxanthine replacing guanine, removing the guanine exocyclic amino group). The resulting I·C base pairs have reduced stacking energy and form two hydrogen-bond interactions, rather than three in G·C base pairs. Hybridization studies suggest that the stabilities of inosine base pairs are $G·C > I·C > I·A > I·G \approx I·T$ (Case-Green & Southern, 1994; Janke, Riechert-Krause, & Weisz, 2011). Previous fiber diffraction studies have shown that poly[d(I)]·poly[d(C)] forms B-DNA, similar to the right-handed B-DNA form of poly[d(A)]·poly[d(T)] or poly[d(G)]·poly[d(C)] (Leslie, Arnott, Chandrasekaran, & Ratliff, 1980). These studies also examined poly[d(IC)], poly[d(AC)]·poly[d(IT)], poly[d(AI)]·poly[d(CT)], poly[d(AIT)]·poly[d(ACT)], poly[d(AIC)]·poly[d(ICT)], and poly[d(IIT)]·poly[d(ACC)], revealing that, in many regards, I·C base pairs behave as A·T base pairs (Leslie et al., 1980). Under the appropriate conditions, these fibers can exist in several conformational families, i.e. A-, B-, Z-DNA. This polymorphism has also been observed in solution using circular dichroism (CD) spectroscopy (Kypur, Kejnovska, Renciuik, & Vorlickova, 2009). Poly[d(IC)] for instance, in addition to having A-DNA and B-DNA CD signatures under appropriate conditions (Vorlickova & Sagi, 1991), has also been shown to display an unusual negative band at high wavelength indicating a left-handed helical structure (Mitsui et al., 1970) reminiscent of the CD spectrum of the left-handed Z-DNA of poly[d(G[5BrC])] (Fishel, Anziano, & Rich, 1990).

Our interest in the physical properties of I-substituted duplex DNA arose from our previous experiments to explore how replacement of thymine by charged and uncharged base analogs impacts the physical properties (melting temperature, helical geometry, bend and twist stiffnesses, etc.) of duplex DNA (Peters et al., 2013, 2014). This prior work was inspired by the proposal that DNA bend stiffness is dominated by DNA charge (Manning, 2006), suggesting that charge-modified polymers

should have distinct bending persistence lengths. In contrast, we were surprised to find that charged and uncharged base analogs had only small effects on bending persistence length, but larger effects on DNA twist flexibility measured in ligase-catalyzed cyclization experiments. Moreover, for an experimental ~400-bp sequence, substitution of some neutral base analogs (such as I for G) altered DNA CD spectra and increased double-helical DNA twist flexibility without large effects on global thermal stability or bend persistence length (Peters et al., 2013, 2014). These observations create a puzzle: what conformational differences might explain altered CD spectra and twist flexibility without large changes in thermal stability or bend persistence length for long I-substituted duplexes (Figure 1)?

To further pursue this question, we here consider X-ray crystallography. The well-studied Dickerson–Drew dodecamer (DDD) is the classic example of B-DNA conformation (Dans et al., 2016; Drew et al., 1981; Tereshko, Minasov, & Egli, 1999a). Based on the sequence of the self-complementary DDD, d(CGCGAATTCGCG), several I-containing dodecamers have previously been studied to determine the conformation of I·C and I·A base pairs in d(CGCIAATTCGCG) (Xuan & Weber, 1992), d(CG CIAATTAGCG) (Corfield, Hunter, Brown, Robinson, & Kennard, 1987), and d(CGCAAATTIGCG) (Leonard, Booth, Hunter, & Brown, 1992), relative to the unmodified d(CGCGAATTAGCG) (Brown, Hunter, Kneale, & Kennard, 1986), d(CGCAAATTGGCG) (Brown, Leonard, Booth, & Chambers, 1989), and d(CGCAAATTTGGCG) (Woods et al., 2004), respectively. The structures of the decamers d(CCAACGTTGG) and d(CCAAGATTGG) (Privé, Yanagi, & Dickerson, 1991) were also studied with I·C base pairs, i.e. d(CCAACITTGG) and d(CCAAIATTGG), respectively (Lipanov, Kopka, Kaczor-Grzeskowiak, Quintana, & Dickerson, 1993). Other published pairs of crystal structures (with and without I substitution) include the Holliday junction-forming d(CCIGTAC[5meC]GG) (Hays, Jones, & Ho, 2004) and the parent d(CCGGTAC[5meC]GG) (Vargason & Ho, 2002); the A-DNA forming d(CCGGGCC[5meC]GG) (Tippin & Sundaralingam, 1997) and its I variant d(CCIGGCC[5meC]GG) (Ramakrishnan & Sundaralingam, 1995); the I·T wobble base pairs in the A-DNA octamer d(GGGGCTCC) (Hunter, Brown, & Kennard, 1986) and its I variant d(GGIGCTCC) (Cruse et al., 1989); and d(CGCICICG) (Kumar, Harrison, Andrews, & Weber, 1992), which resembled the Z-DNA crystal structure of d(CGCGCG) (Gessner, Frederick, Quigley, Rich, & Wang, 1989; Malinina et al., 1998). Structures of I-containing duplexes that include other binding partners, e.g. proteins or the minor groove binder netropsin (Chen, Mitra, Rao, Sekar, & Sundaralingam, 1998; Shi, Mitra, & Sundaralingam, 2002), or without a reference structure were not considered here.

These cases create a database of examples showcasing the effects of I substitution on DNA crystal structures. However, the availability of high-resolution three-dimensional structures with corresponding solution CD spectra is limited. Expanding this collection is one objective of this work. The other motivation is to understand our previous observation that various base analog substitutions, including I, alter DNA conformation and physical properties as shown by CD spectroscopy and cyclization kinetics (Peters et al., 2013, 2014). We hypothesized that I-induced changes in DNA conformation would be detectable in X-ray crystallography experiments (Figure 1).

Here we show that a high-resolution X-ray structure of a decamer DNA duplex with four I substitutions (four I·C pairs in place of G·C pairs) exhibits relatively close resem-

blance to the reference duplex with G·C and A·T pairs only, despite moderate differences between the solution CD spectra of the two duplexes. In contrast, the duplex with all eight G·C pairs replaced by I·C pairs adopts an A-trait-like geometry in the crystal, with a minor groove that remains very narrow along the entire length of the decamer. Additional characteristics of this I-rich duplex include a 5% reduction in helical rise and southeastern C1'-*exo* puckering by more than half the sugars. These features distinguish it from the reference duplex where we observe canonical B-form geometry, i.e. a 3.4 Å rise and mostly south C2'-*endo* puckered sugars. The unusual conformation of the I-rich duplex in the crystal may at least partially account for its non-B form character observed by CD spectroscopy in solution, even under crystallization conditions. We discuss the similarities between the I-rich duplex

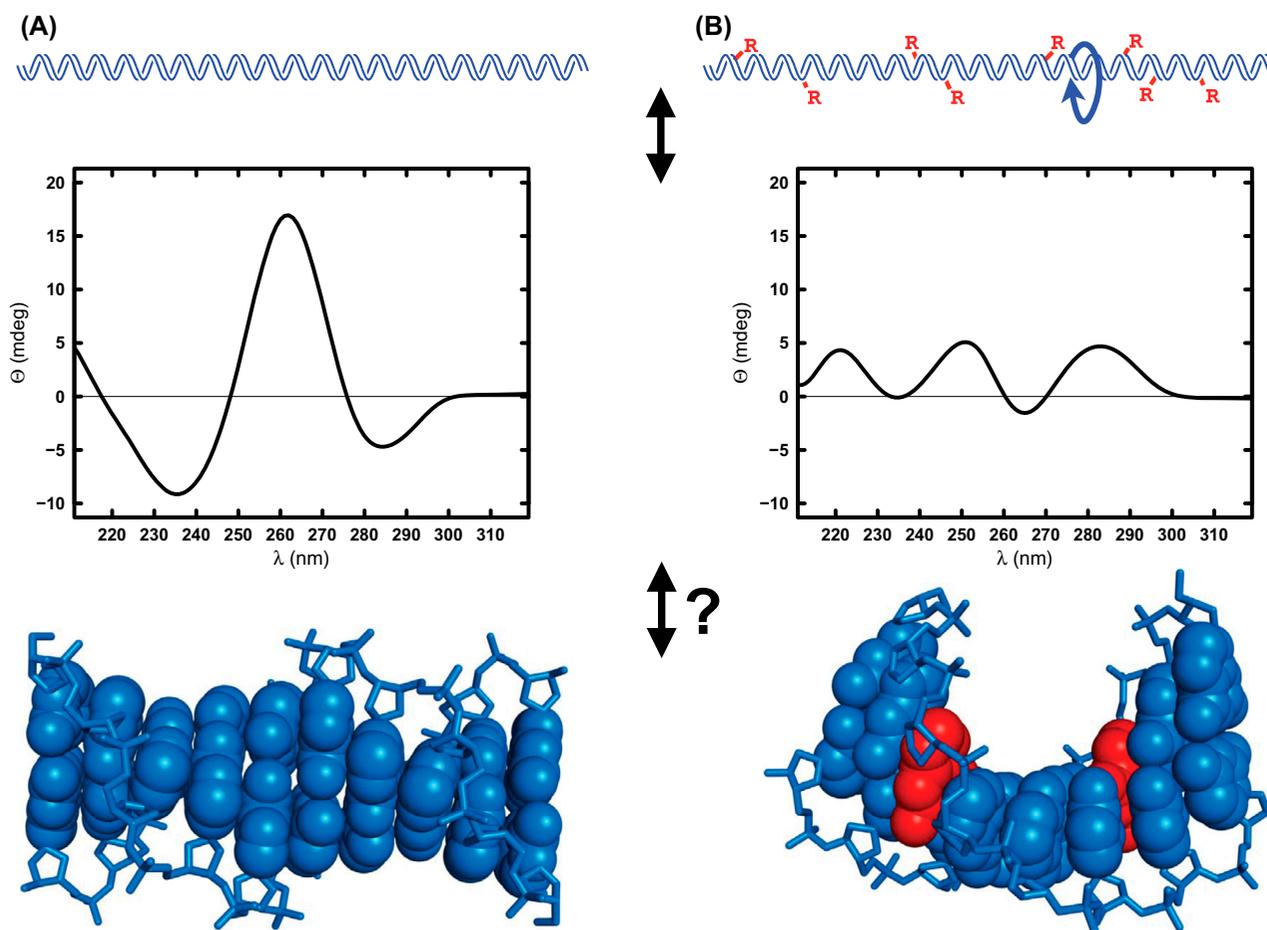


Figure 1. Overall strategy. Previous work characterized the mechanical properties of DNA molecules greater than 400 bp in length and incorporating DNA base analogs, such as inosine (Peters et al., 2013, 2014). These DNA molecules exhibited altered DNA twist flexibilities that correlated with altered circular dichroism (CD) spectra (Peters et al., 2013). We hypothesized that altered CD spectra reflected altered helical conformation and sought to test if altered helical conformations were detected by X-ray crystallography. (A) DNA containing only the four common natural bases (top) displays a conventional B-DNA helical signature in CD experiments (middle). Structural studies of short DNA duplexes (10–12 bp) allow direct visualization of this canonical B-form DNA conformation (bottom). (B) DNA molecules incorporating modified bases (red R) display altered twist flexibility (top), correlating with altered CD spectra (middle). This work explores whether the implied altered helical conformations are detectable in the crystal structures of inosine-substituted DNA (bottom).

and A-tract DNA and the challenges of understanding properties of duplexes when comparing X-ray crystallography and CD spectroscopy in solution.

2. Materials and methods

2.1. Oligonucleotides

All DNA oligonucleotides were purchased with standard desalting from either Integrated DNA Technologies (Coralville, IA) or Trilink Biotechnologies (San Diego, CA).

2.2. CD spectroscopy

CD spectroscopy was performed using a Jasco J-810 spectropolarimeter (Jasco, Easton, MD). Briefly, ultraviolet-CD spectra were acquired from 320 to 210 nm, taking measurements every 0.5 nm with a scanning speed of 10 nm/min. Sample temperature was maintained at 4°C throughout, unless otherwise indicated. Samples were monitored three times with the average of the three scans reported. Samples were prepared by diluting ~7.5 nmol (~50 µg) of DNA into 300 µL of buffer and analyzed in a 0.1 cm cuvette. Buffer samples were scanned in a similar manner and these spectra subtracted from sample spectra.

Thermal denaturation experiments were conducted by sample heating from 5°C to 85°C with a temperature ramp rate of 0.5°C/min. Wavelength scans between 320 and 210 nm were taken every 10°C at a data pitch of 1 nm, a scan rate of 100 nm/min, an accumulation number of 3, a bandwidth of 1 nm, and a response rate of 2 s. Single-wavelength data were collected every 0.5°C at the wavelength of a spectral maximum at 4°C.

All variable wavelength curves were smoothed using Friedman's super smoother (Friedman & Tibshirani, 1984).

2.3. Thermal denaturation

Thermal denaturation data were processed following a two-state transition model (Mergny & Lacroix, 2003). Briefly, the two states are referred to as 'folded' (fully associated) and 'unfolded' (fully dissociated) for simplicity. The data were converted from ellipticity (Θ) as a function of temperature (T) to fraction folded (θ) as a function of temperature

$$\theta(T) = \frac{\Theta(T) - u(T)}{f(T) - u(T)} \quad (1)$$

where the unfolded (u) and folded (f) baselines were determined from the data by linear regression.

First, a sigmoid model was fit to the fraction folded Θ data

$$m(T) = \frac{1}{1 + e^{a(T-b)}} \quad (2)$$

In this model, $T_m = b$ since this is the value where $m(T) = 0.5$.

For the bimolecular reaction of self-complementary oligonucleotides ($2S \rightleftharpoons D$), the total strand concentration (C_t) is constant (and equal to the starting concentration of the DNA)

$$C_t = [S] + 2[D] \quad (3)$$

The fraction annealed or 'folded' (i.e. duplex strand proportion of total strands) is:

$$\theta = \frac{2[D]}{C_t} \quad (4)$$

so that the equilibrium constant is:

$$K_a = \frac{[D]}{[S]^2} = \frac{\theta}{2C_t(1-\theta)^2} \quad (5)$$

and at the T_m it follows that $[S] = C_t/2$ and $[D] = C_t/4$, so that $K_a = 1/C_t$.

From the definition of Gibbs free energy

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (6)$$

and the well known relationship between Gibbs free energy and the equilibrium constant K_a

$$\Delta G^\circ = -RT \ln(K_a) \quad (7)$$

follows the classic form of the van't Hoff equation

$$\ln(K_a) = -\frac{\Delta H^\circ}{R} \left(\frac{1}{T} \right) + \frac{\Delta S^\circ}{R} \quad (8)$$

The change in enthalpy (ΔH°) and entropy (ΔS°) were determined from linear regression using this van't Hoff equation within the analysis interval ($0.2 < \theta < 0.8$) where the plots are reasonably linear. The T_m is calculated by rearranging this last equation when $K_a = 1/C_t$

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln C_t} \quad (9)$$

2.4. Analysis of PDB files

Unless otherwise noted, conformational parameters of the DNA structures were determined from the PDB files using Web 3DNA (Lu & Olson, 2003; Zheng, Lu, & Olson, 2009). Molecular graphics images were rendered using open source PyMOL (Schrodinger, 2015).

2.5. Crystallization, diffraction data collection and data processing

Crystals of d(CCAIICCTGG) (I2G2) were grown by the hanging-drop vapor diffusion technique from droplets (2 µL) containing oligonucleotide (0.23 mM), sodium cacodylate (25 mM, pH 6.5), magnesium acetate (12.5 mM), and 2-methyl-2,4-pentanediol (MPD, 20% v/

v) that were equilibrated against a 1 mL reservoir of sodium cacodylate (50 mM, pH 6.5), magnesium acetate (25 mM), and MPD (40% v/v). Crystals of d(CCAIICC [5BrU]II) (I4) were obtained from droplets (2 μ L) containing oligonucleotide (0.6 mM), sodium cacodylate (20 mM, pH 6.0), magnesium acetate (12.5 mM), and MPD (20% v/v) that were equilibrated against a 1 mL reservoir of sodium cacodylate (50 mM, pH 6.5), magnesium acetate (25 mM), and MPD (40% v/v). All crystals were mounted without further cryo-protection and flash-cooled in liquid nitrogen. I2G2 crystals were grown at 20°C and I4 crystals were grown at 4°C and mounted in the cold room as well.

Diffraction data were collected on the 21-ID-F or 21-ID-D beam lines of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS), located at Argonne National Laboratory (Argonne, IL). For I2G2 crystals, native data-sets were collected on ID-F at 0.97625 Å. For I4 crystals, anomalous data-sets were collected on 21-ID-D by tuning the wavelength to 0.9184 Å (bromine K edge). Crystals were kept at 100 K during data collection using MARCCD 225 or MARCCD 300 detectors. Diffraction data were integrated, scaled and merged with HKL2000 (Otwinowski & Minor, 1997). A summary of selected crystal data and data collection parameters is provided in Table 3.

2.6. Phasing and refinement

The I2G2 structure was solved by molecular replacement using the program MOLREP (Collaborative Computational Project; Vagin & Teplyakov, 1997) and the native decamer d(CCAGGCCTGG) (G4, PDB ID 3U89) (Mae-higashi et al., 2012) as the starting model. The initial refinement was carried out using REFMAC5 (Collaborative Computational Project, 1994; Murshudov et al., 2011) and water molecules were added after manually examining the electron density maps. An additional seven cycles of refinements were carried out in SHELX (Schneider & Sheldrick, 2002). Ions were added after visualizing the electron density and their coordination geometries were examined carefully after each refinement cycle. The final three refinement cycles were carried out using REFMAC5 (Collaborative Computational Project, 1994; Murshudov et al., 2011).

The I4 structure was phased by the single-wavelength anomalous dispersion (SAD) technique using the program PHENIX (Adams et al., 2010). The resulting experimental electron density maps allowed visualization of 32 nucleotides of the 40 nucleotides per unit cell and AutoBuild (Terwilliger et al., 2008) was used to build two partial duplexes. The initial orientations of the duplexes were optimized by rigid body refinement. Repeated rounds of difference Fourier synthesis in combination with manual building using the program COOT

(Emsley & Cowtan, 2004) were performed to complete the duplex models. All refinements were carried out with the program REFMAC5 (Collaborative Computational Project, 1994; Murshudov et al., 2011) and water molecules were added after inspecting sum and difference Fourier electron density maps. Water molecules and ions were gradually added as described for the I2G2 structure. Final refinement parameters are listed in Table 3.

2.7. Data deposition

Final coordinate and structure factor files for the I2G2 and I4 crystal structures have been deposited in the Protein Data Bank (www.rcsb.org): PDB ID codes 5W20 and 5W1Z, respectively.

3. Results and discussion

3.1. CD signatures of long vs. short base-substituted duplexes: uracil and diaminopurine examples

Our study of inosine substitution was initially motivated by results of previous studies characterizing 417-base pair (bp) DNA duplexes with 54.2% GC content where one of the common natural bases was completely substituted by a DNA base analog (Peters et al., 2013, 2014). Modifications included seven charged and uncharged thymine analogs, or 2,6-diaminopurine (D) substituted for adenine, or hypoxanthine (the base of inosine, I) substituted for guanine (Peters et al., 2013, 2014). We did not observe meaningful differences in the absorbance spectra of these 417-bp substituted DNA molecules (Figure S1). Figure 2(A) shows a comparison of CD spectra previously reported (Peters et al., 2013) for 417-bp natural (■), diaminopurine-substituted (Δ), and uracil-substituted (\times) DNA. While the CD signature of uracil-substituted DNA is quite similar to that of natural DNA, the CD signature of diaminopurine-substituted DNA is distinct from both. The observed CD spectra could not be attributed to the mere presence of a modified residue dangling from the terminus of a model hexamer duplex (Figure S2). We therefore interpreted these CD spectra as evidence of diaminopurine-dependent DNA polymorphism, indicating transitions to helical conformations different from canonical B-form DNA.

These results led to the hypothesis that even short (10–12 bp) DNA duplexes with base-substitutions might model the alternate DNA conformations implicated in the studies of 417-bp duplexes. To test this, we first examined the Dickerson–Drew dodecamer (DDD) composed only of the common natural bases. CD spectra for this parent molecule d(CGTGAATTCACG) (■) along with d(CGTGDDTTCD CG) (Δ) and d(CGUGAAUUCACG) (\times) are shown in Figure 2(B). The distinct CD signatures of these dodecamers are remarkably reminiscent of the longer substituted DNA molecules, indicating

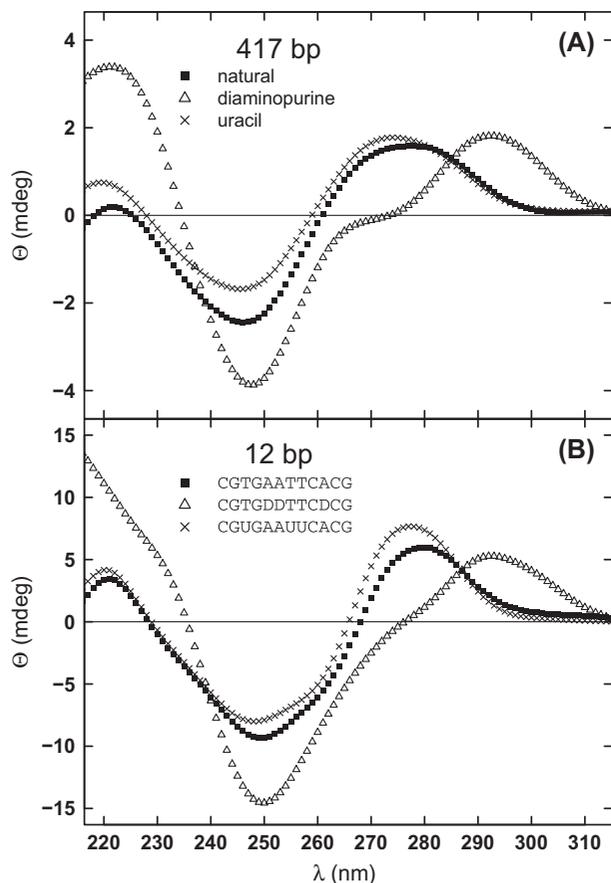


Figure 2. CD spectroscopy showing ellipticity (Θ) as a function of wavelength (λ). (A). CD spectra previously reported (Peters et al., 2013) for 417-bp DNA molecules collected in 10 mM phosphate buffer, pH 7.0, containing 1 M NaCl and comparing natural (■), diaminopurine-substituted (Δ), and uracil-substituted (\times) DNA. (B). CD spectra of dodecamers collected in 10 mM sodium cacodylate, pH 6.6, and comparing natural (■), diaminopurine-substituted (Δ), and uracil-substituted (\times) DNA.

that the preferred helical conformations of each are conserved even in dodecamer duplexes.

3.2. CD signatures of 10-bp base-substituted duplexes: inosine examples

These previous studies focused on DNA molecules where the fraction of substituted bases ranged from 21 to 29%. We hypothesized that increasing the density of base modifications would more dramatically perturb the helical conformation of the resulting DNA. To address this question, we examined CD spectra of duplexes bearing an increasing number of G to I substitutions. The results are shown in Figure 3. When compared to the unsubstituted parent decamer (bold, solid line), a single I substitution per strand (dotted line) does not substantially alter the observed CD spectrum. The spectral effect of

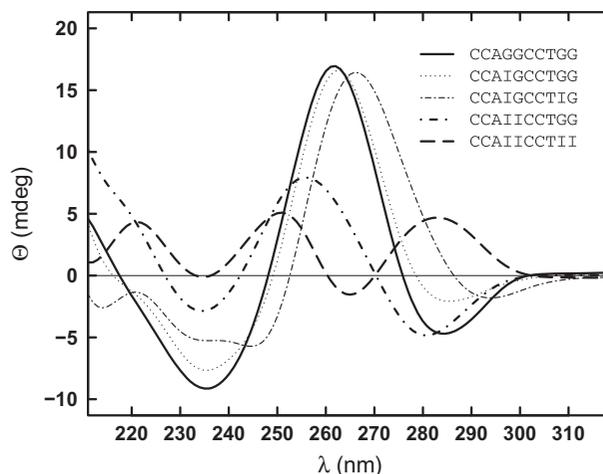


Figure 3. CD spectroscopy of inosine-substituted DNA showing ellipticity (Θ) as a function of wavelength (λ). CD spectra of a series of decamers with increasing number of I substitutions collected in 10 mM sodium cacodylate, pH 6.6. When compared to the unsubstituted parent decamer (bold, solid line), a single I substitution (dotted line) does not substantially alter the observed CD spectrum. The effect of two I substitutions on the observed CD spectra is dependent on whether the modifications are spaced (two dashed line) or adjacent (bold, dot dashed line). Four I substitutions of the parent decamer (bold, long dashed line) result in the most altered CD spectrum.

two I substitutions per strand is dependent on whether the modifications are spaced (two dashed line) or adjacent (bold, dot dashed line). Finally, four I substitutions per strand (bold, long dashed line) resulted in the most altered CD spectrum. We ruled out differences in absorbance spectra (Figure S3) or the contribution of CD spectra from unpolymerized nucleotides (Figure S4) as possible explanations for these distinct CD spectra. This result is consistent with the hypothesis that increasing base substitution drives modified DNA molecules from the classical B-DNA conformation.

3.3. Inosine substitution effects on DNA stability

We sought to rule out that altered CD spectra might reflect duplex denaturation due to the destabilizing effects of I substitution. We therefore performed thermal denaturation experiments for I-substituted DNA molecules in 10 mM sodium cacodylate buffer (pH 6.6). Along with the parent, two decamer duplexes displaying the most unusual CD signatures at 4°C were chosen for further study: d(CCAGGCCTGG), d(CCAIICCTGG), and d(CCAIICCTII). CD spectra were collected at 10°C intervals from 5°C to 85°C (Figure 4). Each of the decamers produced similar CD spectra in the denatured state at 85°C (Figure 4(D)), further indicating that the duplex conformations are responsible for the observed CD spectra in the native state.

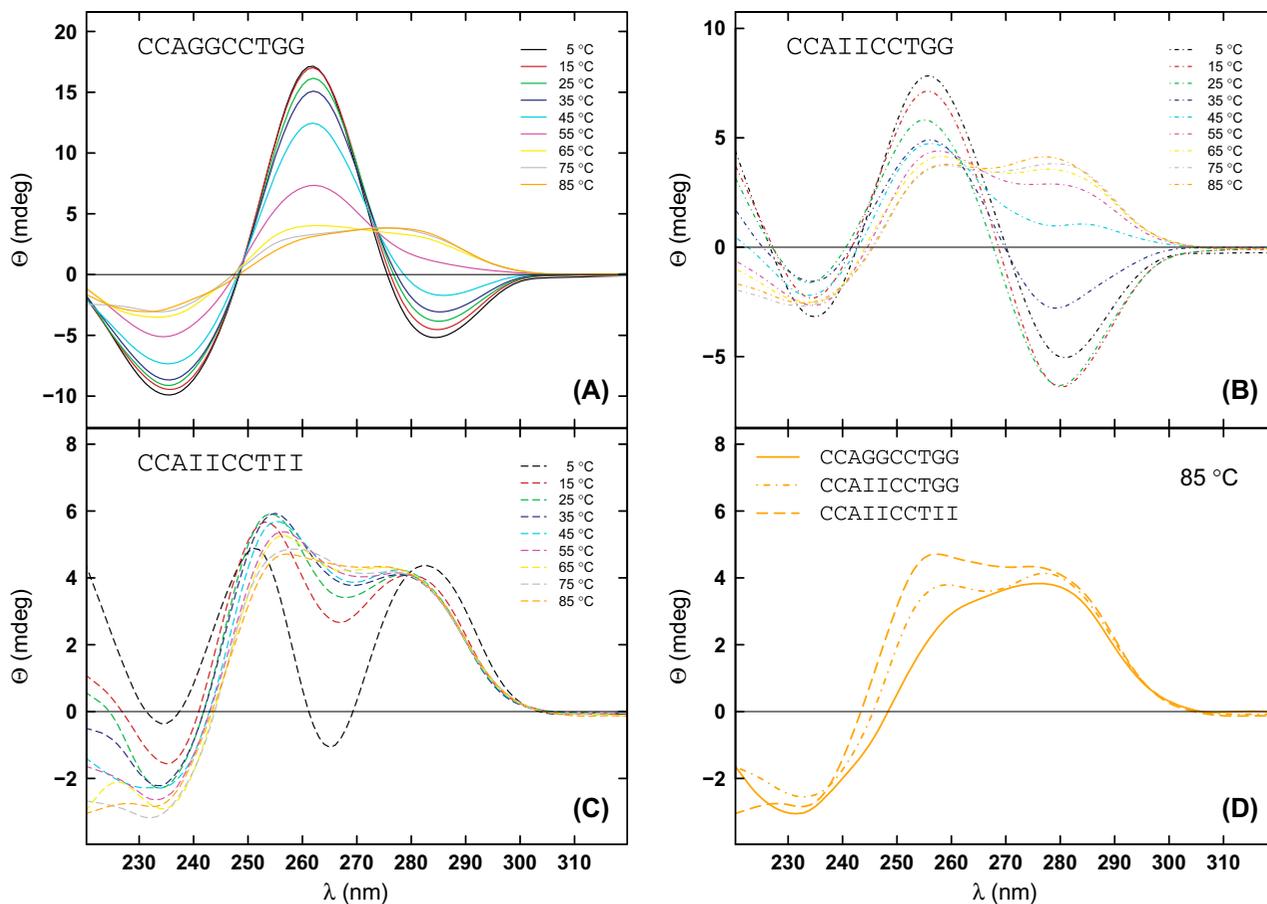


Figure 4. CD thermal denaturation analysis of inosine-substituted DNA molecules in 10 mM sodium cacodylate, pH 6.6. CD spectra showing ellipticity (Θ) as a function of wavelength (λ) were collected at 10°C intervals from 5°C to 85°C (distinguished by color) for (A) d(CCAGGCCTGG) duplex (solid lines), (B) d(CCAIICCTGG) duplex (dot dashed lines), and (C) d(CCAIICCTII) duplex (long dashed lines). (D). Comparison of the decamers in the denatured state (85°C), distinguished as above.

For each decamer, an appropriate peak wavelength was chosen for more detailed analysis (Table 1). CD ellipticity was collected as a function of temperature at this wavelength (Figure 5(A)). Baselines were determined from the data by linear regression. Using

Equation 1 of Methods, the data were transformed to fraction folded (i.e. fraction annealed) as a function of temperature (Figure 5(B)). The melting temperature (T_m), which occurs at a fraction folded of 0.5, was determined using two methods. First, a sigmoid model was fitted to

Table 1. CD-derived thermal stabilities.^a

Sequence	λ (nm)	Buffer	T_m (°C) ^b	T_m (°C) ^c
CCAGGCCTGG	235	10 mM sodium cacodylate, pH 6.6	51.9	52.1
		Crystallization buffer #1 ^d	52.8	52.9
		Crystallization buffer #2 ^e	55.2	55.9
CCAIICCTGG	280	10 mM sodium cacodylate, pH 6.6	34.2	33.6
		Crystallization buffer #1 ^d	37.0	37.0
		Crystallization buffer #2 ^e	39.8	39.9
CCAIICCTII	265	10 mM sodium cacodylate, pH 6.6	9.6	9.6
		Crystallization buffer #1 ^d	16.5	16.3
		Crystallization buffer #2 ^e	21.9	22.5

^aRelative standard deviation from 3 to 4 repeats is 2% for T_m estimates (see Table S1).

^bUsing a sigmoid fit method (see Methods for details).

^cUsing the van't Hoff method (see Methods for details).

^d50 mM sodium cacodylate, pH 6.5; 25 mM magnesium acetate; 40% MPD.

^e40 mM sodium cacodylate, pH 5.5; 20 mM magnesium chloride; 10% MPD; 40 mM lithium chloride; 20 mM cobalt hexamine chloride.

the fraction folded data and the intersection point at 0.5 determined (Figure 5(B) and Table 1). Second, the fraction folded data were further transformed to equilibrium constants (see Equation 5 of Methods) and examined in the analysis interval (Figure 5(C)). Figure 5(C) shows a plot of the natural logarithm of the equilibrium constant as a function of inverse absolute temperature (in Kelvin) multiplied by 1000. In the analysis interval, linear regression analysis and the van't Hoff equation were used to determine T_m (Table 1). The relative standard deviation of either method is about 2% (see Table S1). The agreement between the two methods in Table 1 is clearly within this range.

The melting temperature results are consistent with our previous studies showing that I·C pairs have a lower melting temperature than G·C pairs because I stacks onto adjacent base pairs with less energy than G (Peters et al., 2014). Here, we have shown that this effect is more striking in short decamer duplexes, with a melting temperature penalty of $\sim 8.8^\circ\text{C}$ per inosine substitution. However, even with this consideration, these data do not suggest that the altered CD spectra collected at the lowest temperatures are attributable to partial denaturation.

3.4. Crystallization buffer effects on CD spectra and CD-derived thermodynamic parameters

We set out to use X-ray crystallography to understand the origin of the perturbed CD spectra in modified duplexes. We first compared CD spectra obtained for the duplexes

in standard buffer and in crystallization buffers. CD spectra for each duplex in the two buffers that yielded crystals are shown in Figure 6. Unusual CD spectra tend to be preserved, though increasing concentrations of the precipitant (\pm)-2-methyl-2,4-pentanediol (MPD) reduce the magnitude of some CD spectral peaks. We conclude that crystallization buffers do not significantly alter the preferred conformations of the duplexes in solution.

We next performed thermal denaturation experiments for I-substituted DNA molecules in the two crystallization buffers to determine how these conditions affected thermal stability. Examples are shown in Figure S5 (analysis in Figure S6) and the results are summarized in Table 1. In all cases, the crystallization buffers stabilized the duplexes. Stabilization was most pronounced for d(CCAIICCTII) and inversely related to the initial stability of the duplex. These experiments showed that ionic conditions of the buffers stabilized the duplexes, even in the presence of the precipitant MPD.

3.5. Features of I-substituted DNA crystal structures in the PDB

We hypothesized that evidence of I-induced changes in DNA conformation would be detectable in pairs of modified and unmodified crystal structures previously reported in the PDB. Figure 7 displays views of the central minor groove of structures of such B-DNA duplexes superimposed with their I-substituted derivatives. Additional details of these structures appear in Table 2.

Table 2. Selected structural parameters for DNA duplexes.

Duplex	PDB ID	Length (bp)	Space group	Resolution (Å)	Type	Refs.
CGCGAATTCGCG	7BNA	12	$P2_12_12_1$	1.9	B	Wing et al. (1980), Holbrook et al. (1985)
CGCIAATTCGCG	1D77	12	$P2_12_12_1$	2.4	B	Xuan and Weber (1992)
CGCGAATTAGCG	112D	12	$P2_12_12_1$	2.5	B	Brown et al. (1986)
CGCIAATTAGCG	114D	12	$P2_12_12_1$	2.5	B	Corfield et al. (1987)
CGCAAATTGGCG	111D	12	$P2_12_12_1$	2.25	B	Brown et al. (1989)
CGCAAATTIGCG	1D81	12	$P2_12_12_1$	2.5	B	Leonard et al. (1992)
CGCAAATTTGCG	1S2R	12	$P2_12_12_1$	1.53	B	Woods et al. (2004)
CCAACGTTGG	5DNB	10	$C2$	1.4	B	Privé et al. (1991)
CCAACITTGG	1D61	10	$C2$	1.3	B	Lipanov et al. (1993)
CCAAGATTGG	3DNB	10	$C2$	1.3	B	Privé et al. (1991)
CCAAIATTGG	1D62	10	$C2$	2.0	B	Heinemann, Alings, and Hahn (1994), El Hassan and Calladine (1997)
CCGGTAC[5meC]GG	1L6B	10	$C2$	1.5	Holliday	Vargason and Ho (2002)
CCIGTAC[5meC]GG	1S1L	10	$C2$	2.2	Holliday	Hays et al. (2004)
CCGGGCC[5meC]GG	323D	10	$P2_12_12_1$	2.15	A	Tippin and Sundaralingam (1997)
CCIGGCC[5meC]GG	213D	10	$P2_12_12_1$	1.6	A	Ramakrishnan and Sundaralingam (1995)
GGGGCTCC	1D92	8	$P6_1$	2.25	A	Hunter et al. (1986)
GGIGCTCC	1D90	8	$P6_1$	1.7	A	Cruse et al. (1989)
CGCGCG	392D	6	$C222_1$	3.0	Z	Malinina et al. (1998)
CGCICIG	1D53	8	$P6_5$	1.5	Z	Kumar et al. (1992)
CCAGGCCTGG	3U89	10	$C2$	0.96	B	Maehigashi et al. (2012)
CCAIICCTGG	5W20	10	$P2_12_12_1$	1.36	B	this work
CCAIICC[5BrU]II	5W1Z	10	$P1$	1.55	B	this work

The sets of structures are oriented to minimize root mean squared deviations (RMSD). Relative to the corresponding dodecamer references (with G instead of I), the RMSD of a duplex with central I·C pairs is 0.282 Å (453 atoms; Figure 7(A)), with central I·A pairs RMSD = 0.377 Å (448 atoms; Figure 7(B)), and with central A·I pairs RMSD = 0.444 Å (440 atoms; Figure 7(C)). Rather than a reference with A·G pairs (Privé et al., 1987), using the B-DNA structure of the dodecamer with A·T pairs as the reference for this latest inosine variant gives RMSD = 0.611 Å (388 atoms; Figure 7(D)). In a similar manner, relative to reference unmodified decamers, we have RMSD = 0.436 Å (370 atoms, Figure 7(E)) with central I·C pairs and RMSD = 1.18 Å (402 atoms; Figure 7(F)) with central I·A pairs. We conclude for these sets of structures with the parent DNA in B-form, that each variant structure with only one I substitution per strand closely resembles its unmodified partner.

In addition to the B-DNA sets, Figure 8 shows superimposed structures of a Holliday junction, an A-DNA decamer, an A-DNA octamer with I·T wobble base pairs, and a left-handed duplex that resembles Z-DNA. This last example (Figure 8(D)) is not a true set, because the I variant d(CGCICICG) is longer than the parent d(CGCGCG), but serves to show another type of helical conformation. RMSD = 0.898 Å (754 atoms, Figure 8(A)) for the Holliday junction, RMSD = 0.307 Å (368 atoms, Figure 8(B)) for the A-DNA decamer, and RMSD = 0.573 Å (320 atoms, Figure 8(C)) for the A-

DNA octamer. Again, each low density I-substituted structure largely resembles its partner.

Since major conformational differences were not obvious by inspection, we quantitatively analyzed main chain and glycosidic (χ) torsion angles, local base pair helical parameters, and local base pair step parameters using Web 3DNA (Anosova et al., 2016; Lu & Olson, 2003; Vargason, Eichman, & Ho, 2000). Clusters of differing helical morphology emerge when using pairs of parameters, such as χ (the glycosidic torsion angle between the sugar and base) and δ (the backbone torsion angle associated with the sugar ring) (Anosova et al., 2016), as shown by the bounding ellipses in Figure 9(A). Each open circle in Figure 9 corresponds to an individual base or base pair step from the B-DNA structures shown in Figure 7 (black symbols) or the A-DNA structures in Figure 8(B) and (C) (red symbols), whereas the larger filled circles correspond specifically to the I-containing residues. Also shown are the previously reported bounding ellipses for the given parameters pairs (Anosova et al., 2016; Lu & Olson, 2003). Figure 9(B) shows z_p (the mean projection of the two phosphorus atoms onto the z -axis of the dimer step frame) and $z_p(h)$ (the mean projection of the two phosphorus atoms onto the z -axis of the helical reference frame). Figure 9(C) shows x -displacement (the distance the midpoint of the short axis of a base pair is displaced from the helix axis) and inclination (the angle between the long axis of a base pair and a plane perpendicular to the helix axis). Figure 9(D) shows slide (the in-plane displacement of

Table 3. Selected crystal data, diffraction data collection and refinement parameters.

Crystal structure	I2G2	I4
Unit cell constants a, b, c [Å]; α, β, γ [°]	34.19, 35.04, 41.34 90.0, 90.0, 90.0	24.88, 32.80, 34.03 90.25, 107.32, 111.04
Space group	$P2_12_12_1$	$P1$
<i>Data collection</i>		
Wavelength [Å]	0.97625	0.9184
Resolution (outer shell) [Å]	50.00–1.36 (1.41–1.36)	32.24–1.55 (1.58–1.55)
Unique reflections	11,019 (1,002)	13,618 (640)
Completeness [%]	99.1 (95.0)	97.1 (90.0)
$I/\sigma(I)$	30.57 (5.20)	28.70 (2.37)
R-merge	0.073 (0.343)	0.073 (1.000)
R-pim	0.035 (0.133)	0.053 (0.434)
Redundancy	7.7 (7.6)	5.9 (4.5)
<i>Refinement</i>		
R-work	0.148	0.207 (0.316) ^a
R-free	0.181	0.234 (0.364) ^a
Number of nucleic acid atoms ^b	438	803
Avg. B-factor: nucleic acid atoms [Å ²]	10.6	22.7
Avg. B-factor: Mg ²⁺ /Na ⁺ /water [Å ²]	12.8/–/20.9	28.1/32.1/28.7
R.m.s.d. bond lengths [Å]	0.012	0.012
R.m.s.d. bond angles [°]	1.6	2.1
PDB ID	5W20	5W1Z

^aRefinement outer shell 1.57–1.55 [Å].

^bIncludes dual occupancy atoms.

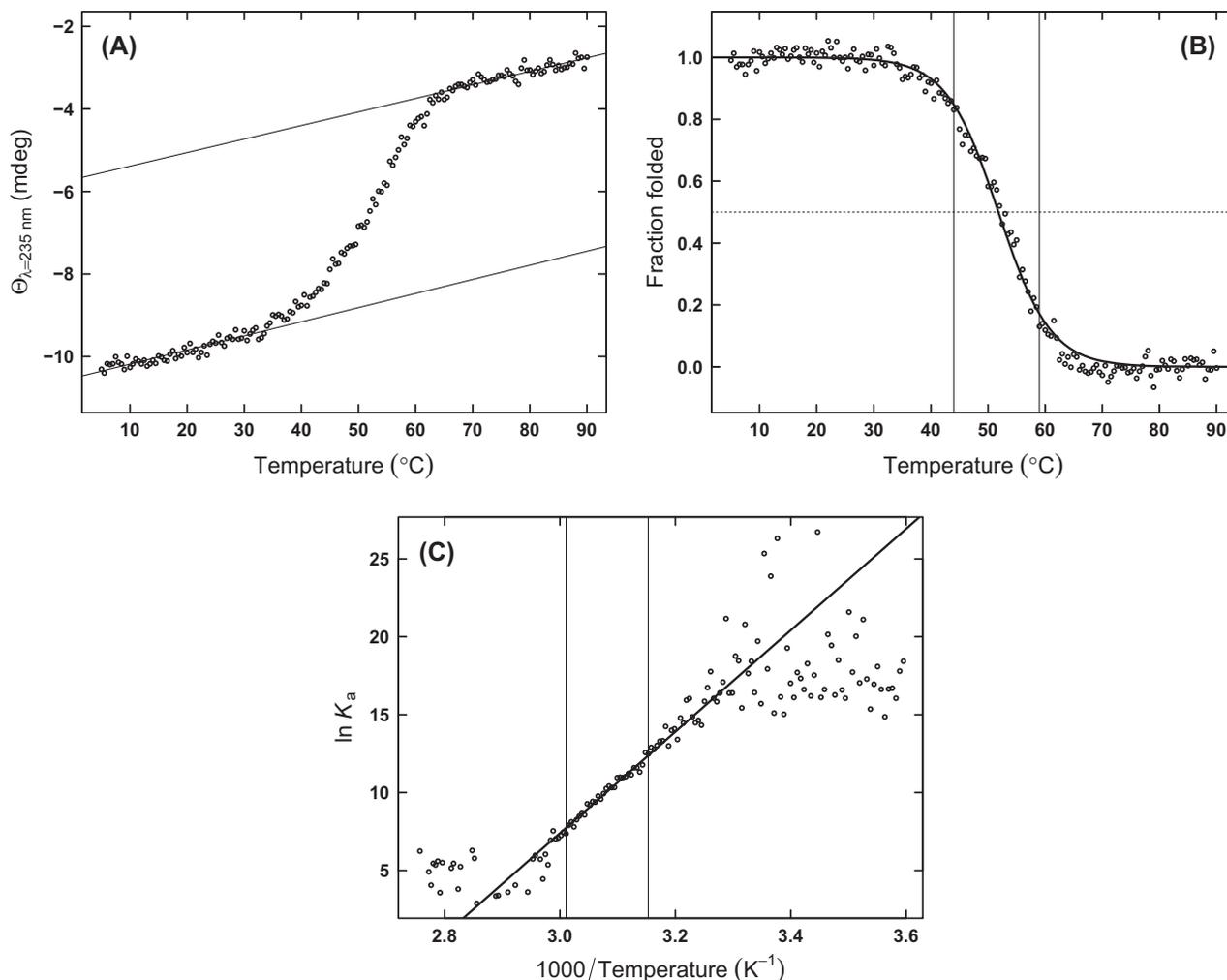


Figure 5. Processing of CD thermal denaturation data for inosine-substituted DNA molecules. (A) Representative single wavelength CD ellipticity collected as a function of temperature for the d(CCAGGCCTGG) duplex followed at 235 nm. Baselines (solid lines) were determined from the data by linear regression. (B) Using Equation 1 of Methods, data were transformed to fraction folded (i.e. fraction annealed) as a function of temperature. The melting temperature (T_m) occurs at a fraction folded of 0.5 (horizontal dashed line). Two methods were used to determine the T_m . First, a sigmoid model (bold curve) was fitted to the fraction folded data and the intersection point at 0.5 determined. Second, the fraction folded data were further transformed to equilibrium constants (Equation 5 of Methods) and examined in the analysis interval (vertical lines). (C) The natural logarithm of the equilibrium constant as a function of inverse absolute temperature (in Kelvin) multiplied by 1000. In the analysis interval (vertical lines), linear regression analysis (bold line) and the van't Hoff equation were used to determine T_m (see Methods).

one base pair relative to the previous along the short axis of the base pair step) and roll (the rotation of a base pair step about the long axis of the base pair step). In each case, there is good agreement with the previously reported discrimination between A-DNA and B-DNA (Anosova et al., 2016; Lu & Olson, 2003). We conclude from this analysis of published structures that 8%–10% I substitution does not drive altered DNA conformation as detected by crystallography. We do not know if the I-substituted DNA duplexes exhibited unusual CD spectra in solution in these cases.

3.6. CD spectra vs. DNA crystal structures: new I-substituted examples

We therefore, wished to analyze by X-ray crystallography our cases where I substitution clearly altered CD signatures. We first focused our attention on I2G2, the decamer d(CCAIICCTGG) whose CD signature appears as bold, dot dashed lines in Figure 3. Crystals of I2G2 diffracted to 1.36 Å and are of space group $P2_12_12_1$ with unit cell constants $34 \times 35 \times 41$ Å (Table 3). Figure 10(A) shows an overlay of I2G2 superimposed with the G4 'native' decamer (PDB code 3U89) (Maehigashi et al., 2012). The native G4 decamer is of space group $C2$.

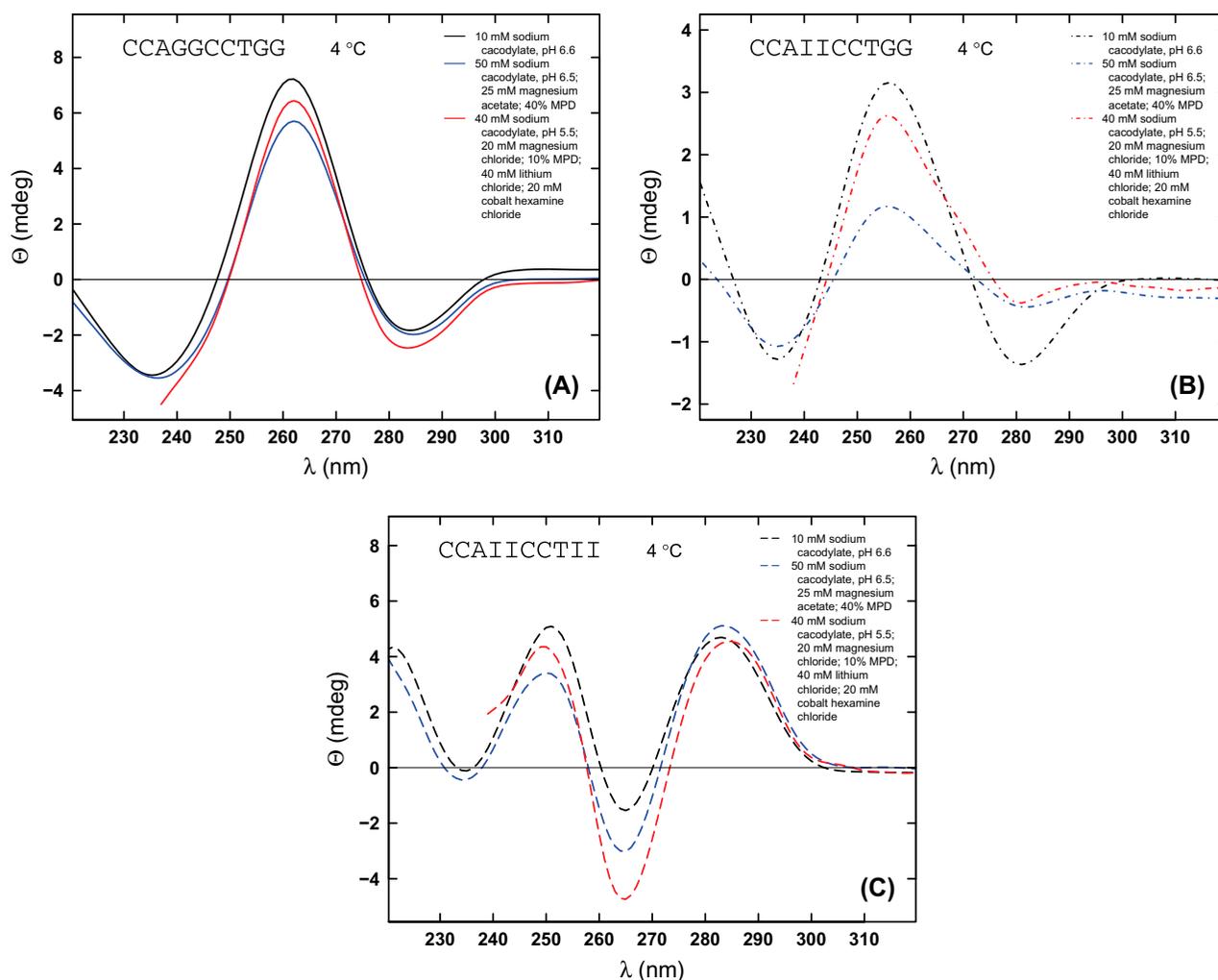


Figure 6. CD spectroscopy of inosine-substituted DNA molecules in various buffers. CD spectra showing ellipticity (Θ) as a function of wavelength (λ) were collected in 10 mM sodium cacodylate, pH 6.6 (black), 50 mM sodium cacodylate, pH 6.5; 25 mM magnesium acetate; 40% MPD (blue), or 40 mM sodium cacodylate, pH 5.5; 20 mM magnesium chloride; 10% MPD; 40 mM lithium chloride; 20 mM cobalt hexamine chloride (red) for (A) d(CCAGGCCTGG) duplex (G4, solid lines), (B) d(CCAIICCTGG) duplex (I2G2, dot-dashed lines), and (C) d(CCAIICCTII) duplex (I4, long-dashed lines). Due to high HT voltages for wavelengths less than 250 nm in crystallization buffer 2, no data were collected below 240 nm.

Therefore, incorporation of I appears to alter the packing arrangement, but not the overall structure of the duplex. The minor groove of the I2G2 decamer is only slightly narrower (Figure 10(C)). One complication arises from conformational disorder in the G4 decamer, where for many nucleotides two alternative conformations were refined (Maehigashi et al., 2012). We deleted atoms of the ‘B’ structure and only retained the ‘A’ structure for the overlay. The RMSD for 316 atom pairs is 0.878 Å. This novel structure demonstrates that for a DNA duplex with 20% base substitution and clearly distinct helical conformation from native DNA by CD in solution (Figure 3), the resulting crystal structure largely does not reveal the alternate DNA conformation.

To extend this work, we studied I4, the decamer with the highest degree of base substitution (40%) and the most distinctive helical geometry by solution CD (bold, long-dashed lines in Figure 3). We obtained the crystal structure of the I4 decamer by phasing with bromine single-wavelength anomalous dispersion (Br-SAD) (Egli, 2016; Egli, Lubini, & Pallan, 2007; Egli et al., 1998). This crystal form is of space group $P1$, contains two duplexes per asymmetric unit, and the resolution is 1.55 Å (Table 3). Unlike in the case of the G4 and I2G2 duplexes, the overlay of the G4 and I4 duplexes (Figure 10(B)) reveals a considerably different geometry for the I4 helix. The most obvious distinction relates to

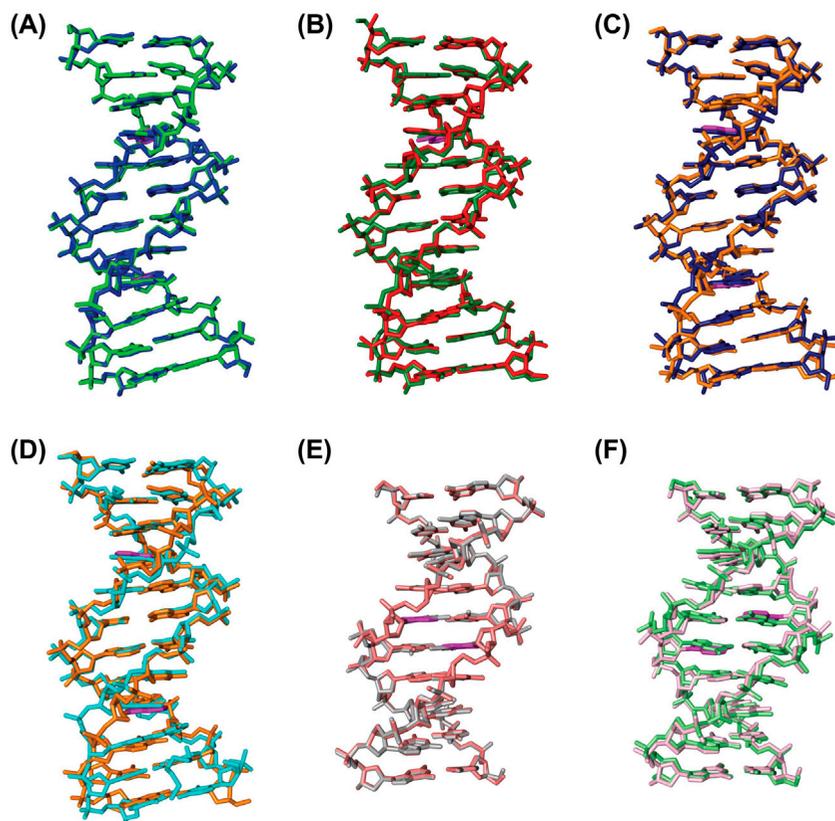


Figure 7. Crystal structures of B-DNA duplexes with their inosine-substituted derivatives (see Table 2). I nucleobases are highlighted in magenta. (A) Superimposition of d(CGCGAATTCGCG) (blue; PDB code 7BNA) (Holbrook, Dickerson, & Kim, 1985; Wing et al., 1980) and d(CGCIAATTCGCG) (green; PDB code 1D77) (Xuan & Weber, 1992) duplexes, corresponding to the relative orientation with the smallest root mean squared deviation. (B) Superimposition of d(CGCGAATTAGCG) (forest; PDB code 112D) (Brown et al., 1986) and d(CGCIAATTAGCG) (red; PDB code 114D) (Corfield et al., 1987). (C) Superimposition of d(CGCAAATTGGCG) (dark blue; PDB code 111D) (Brown et al., 1989) and d(CGCAAATTIGCG) (orange; PDB code 1D81) (Leonard et al., 1992). (D) Superimposition of d(CGCAAATTTGCG) (cyan; PDB code 111D) (Woods et al., 2004) and d(CGCAAATTIGCG) (orange; PDB code 1D81) (Leonard et al., 1992). (E) Superimposition of d(CCAACGTTGG) (gray; PDB code 5DNB) (Privé et al., 1991) and d(CCAACITTGG) (salmon; PDB code 1D61) (Lipmanov et al., 1993). (F) Superimposition of d(CCAAATTGG) (lime green; PDB code 3DNB) (Privé et al., 1991) and d(CCAAIATTGG) (light pink; PDB code 1D62) (Lipmanov et al., 1993).

minor groove width, which remains very narrow along the entire length of the I4 decamer (Figure 10(C)).

Symmetric DNA A-tracts are generally defined as DNA sequences of at least four consecutive A-T base pairs without an intervening T-A step, whereas asymmetric A-tracts are generally A_nT_m , although two or more As followed by pyrimidines, A_nY_m , are referred to as an A-tract in addition to homo-A-homo-T runs. When embedded in a general sequence, A-tracts centrally curve the DNA double helix towards the minor groove, which progressively narrows in the 5' to 3' direction of the adenine strand (Drsata et al., 2014). A major finding of previous A-tract crystal structures is that the minor groove is exceptionally narrow (Shatzky-Schwartz et al., 1997). As the minor groove of I4 is contracted, it could be referred to as a DNA 'AI'-tract in analogy to A-tract DNA. In fact, the minor groove remains narrow up to

the very 3' terminus (II) of the strand (Figure 10(B)). This cooperative behavior by sequential I residues may also partly explain the two different CD spectra observed for two I substitutions per strand in Figure 3, whereby the CD spectra were dependent on whether the two modifications were spaced (two dashed line) or adjacent (bold, dot dashed line) to one another.

Besides a narrow minor groove, the hallmark of an A-tract in solution, and especially phased A-tracts, is the induced curvature of the helix axis. Gel electrophoresis experiments demonstrate that apparent curvature is largely retained when I substitutions are introduced singly into A-tracts but decreases abruptly for pure I-tracts. One I substitution does not cause any detectable change in curvature at the 5' terminus of an A-tract (e.g. comparing AAATTT with IAATTC), it reduces curvature by 10%–12% within an A-tract (e.g. AAIAA or AIATCT) (Diek-

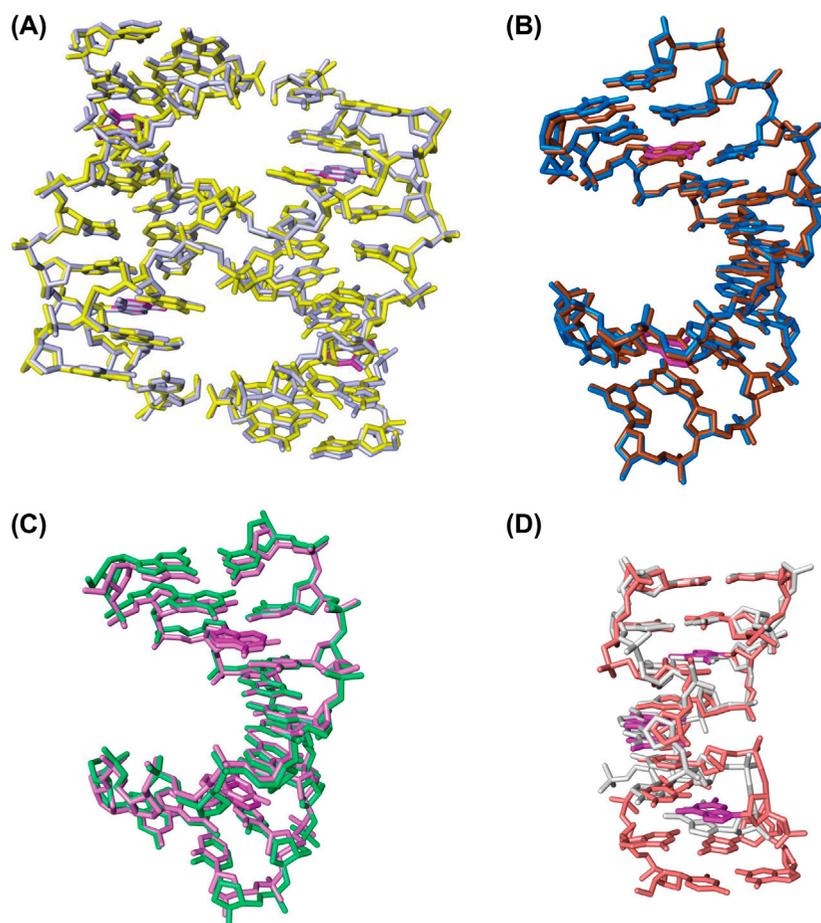


Figure 8. Crystal structures of non-B form duplexes with their inosine-substituted derivatives (see Table 2). I nucleobases are highlighted in magenta. (A) Superimposition of four-way junction formed from d(CCGGTAC[5meC]GG) (light blue; PDB code 1L6B) (Vargason & Ho, 2002) and d(CCIGTAC[5meC]GG) (yellow; PDB code 1S1L) (Hays et al., 2004). (B) Superimposition of the A-DNA decamers d(CCGGGCC[5meC]GG) (brown, PDB code 323D) (Tippin & Sundaralingam, 1997) and d(CCIGGCC[5meC]GG) (marine blue, PDB code 213D) (Ramakrishnan & Sundaralingam, 1995). (C) Superimposition of the A-DNA octamers d(GGGGCTCC) (violet; PDB code 1D92) (Hunter et al., 1986) and d(GGIGCTCC) (limegreen; PDB code 1D90) (Cruse et al., 1989). (D) Superimposition of the Z-DNA d(CGCGCG) (white; PDB code 392D) (Malinina et al., 1998) and d(CGCICIG) (bright orange; PDB code 1D53) (Kumar et al., 1992).

mann, von Kitzing, McLaughlin, Ott, & Eckstein, 1987; Koo & Crothers, 1987; Shatzky-Schwartz et al., 1997), and exhibits the greatest effect at the 3'-terminus of the A-tract (Mollegaard, Bailly, Waring, & Nielsen, 1997). Global axis curvature strongly differs between G4 (0.5°) and I4 (10.4°). This and other differences in helical parameters that distinguish the G4 and I4 duplexes were calculated with the program CURVES (Lavery & Sklenar, 1996, 1988). For instance, in the G4 duplex the average helical rise (global inter-base pair parameter) is 3.44 Å (max. 3.75 Å, min. 3.14 Å), but it is reduced in the I4 duplex to 3.29 Å (max. 3.55 Å, min. 3.07 Å).

Systematic differences in the sugar puckers are also noteworthy. More than half of the 2'-deoxyribose sugars (11 of 20) in the I4 duplex display a southeastern pucker mode (C1'-*exo*). By comparison, only five sugars in the

G4 reference duplex exhibit such a pucker. It is interesting that half of the eight A-tract residues (underlined) in the central portion of the DDD duplex d(CGCGAAATTCCGCG) also adopt the C1'-*exo* pucker (PDB ID 355D) (Shui, McFail-Isom, Hu, & Williams, 1998). However, unlike in the DDD and G4 duplexes, the minor groove in I4 remains narrow throughout and becomes even narrower at the duplex ends (Figure 10(C)). With respect to d(CCAAICC[5BrU]II), the I4 decamer features a central AI-tract (underlined) that apparently zips up the minor groove, and this extremely narrow minor groove is among the hallmarks of an I-rich duplex.

Interestingly, waters form a partial spine in the center of the minor groove of I4. The metal ion coordination of both I2G2 and I4 can be seen together with the quality of the electron density in Figure 11. Figure 11(A) shows the

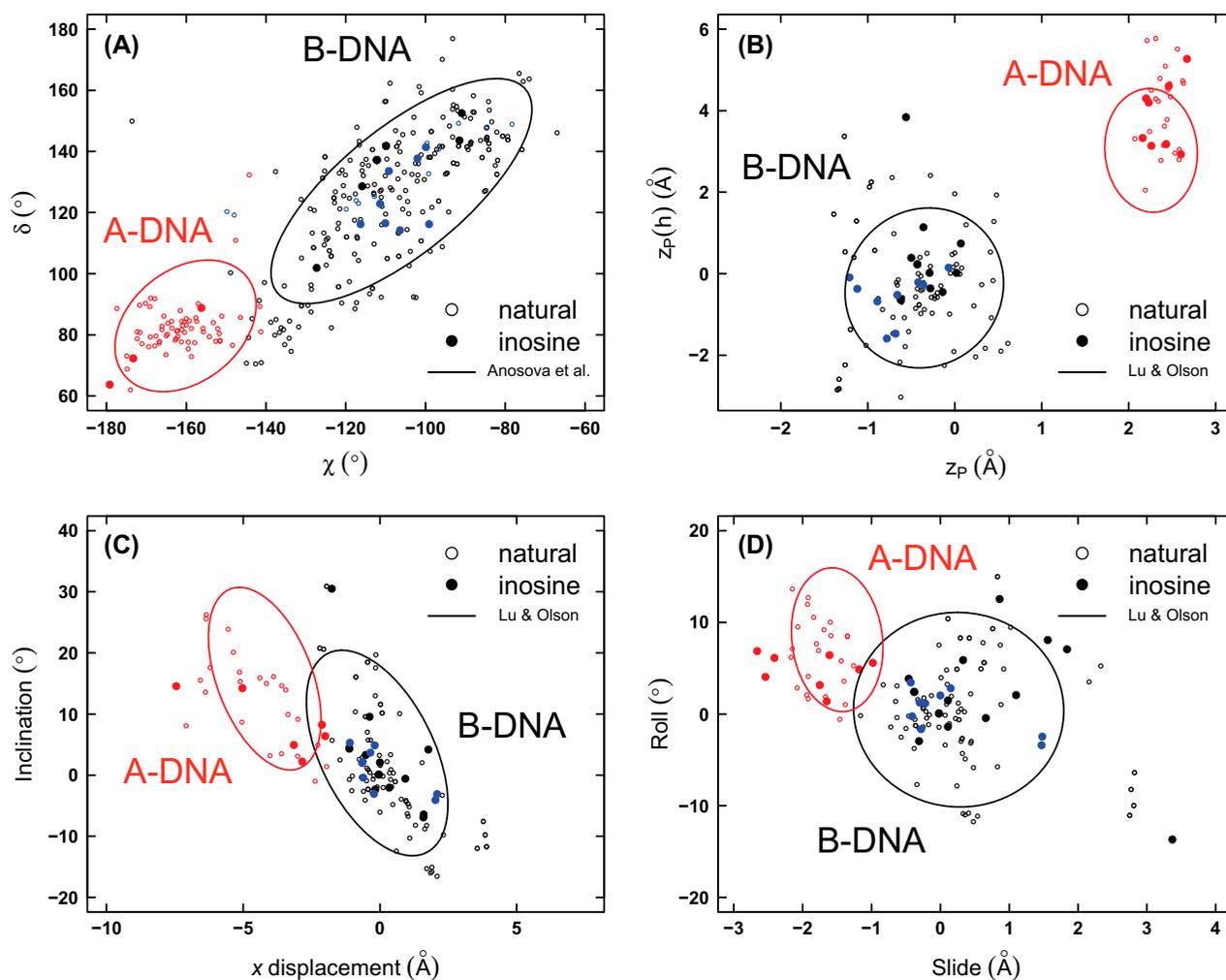


Figure 9. Plots of pairs of selected conformational parameters showing the differences between A-DNA (red) and B-DNA (black). For comparison, the parameters for the new I4 structure are shown in blue. Parameters for natural base pairs or base pair steps are shown with open circles, whereas the larger filled circles correspond specifically to the I-containing residues. The bounding ellipses were previously reported (Anosova et al., 2016; Lu & Olson, 2003). (A) Torsion angles δ and χ . (B) Projected phosphorous positions, z_P and $z_P(h)$. (C) Helical inclination and x -displacement. (D) Dimer step roll and slide.

density for I2G2 around Mg^{2+} ions in the major and minor groove on opposite sides of the same I-C pair, and Figure 11(B) shows the density for I4 around the Na^+ coordinated to O2 oxygens of C residues on opposite strands at the central base pair step. Figure 11(C) depicts the metal ions bound in the major and minor grooves of the three duplexes. The view is into the major groove, G4 is at the left, I2G2 is in the middle and I4 is at the right. Filled circles in red indicate Mg^{2+} hexahydrate ions in the major groove and open circles indicate Mg^{2+} hexahydrate ions in the minor groove. G4 occupies a crystallographic dyad and the three open circles next to T and C in the minor groove indicate three partially occupied ions (A, B, and C in PDB 3U89). The coordinations of Mg^{2+} ions change in I2G2 so that they are grouped around central I-C pairs in both grooves, indicating that the minor groove is still wide

enough to accommodate a Mg^{2+} hexahydrate in that duplex (the asymmetric unit in the I2G2 structure contains one duplex in a general location). Finally, the I4 duplex shows Mg^{2+} hexahydrates as well as Na^+ (open cyan circles indicate minor groove coordination). However, the central minor groove in I4 is too narrow to accommodate a Mg^{2+} ion. Instead, Mg^{2+} ions are positioned at the ends of the duplex that leave more space for the hexahydrate ion. Na^+ is coordinated at the center of the minor groove and bridges O2 oxygens of C residues from opposite strands. Alkali metal ions make comparable contacts at central T residues in the minor groove of A-tracts, e.g. AATT (Tereshko, Minasov & Egli 1999b). This observation further underscores the similarity between A- and AI-tracts. Na^+ are smaller and they are not fully hydrated compared to Mg^{2+} , i.e. they form inner-sphere complexes. In

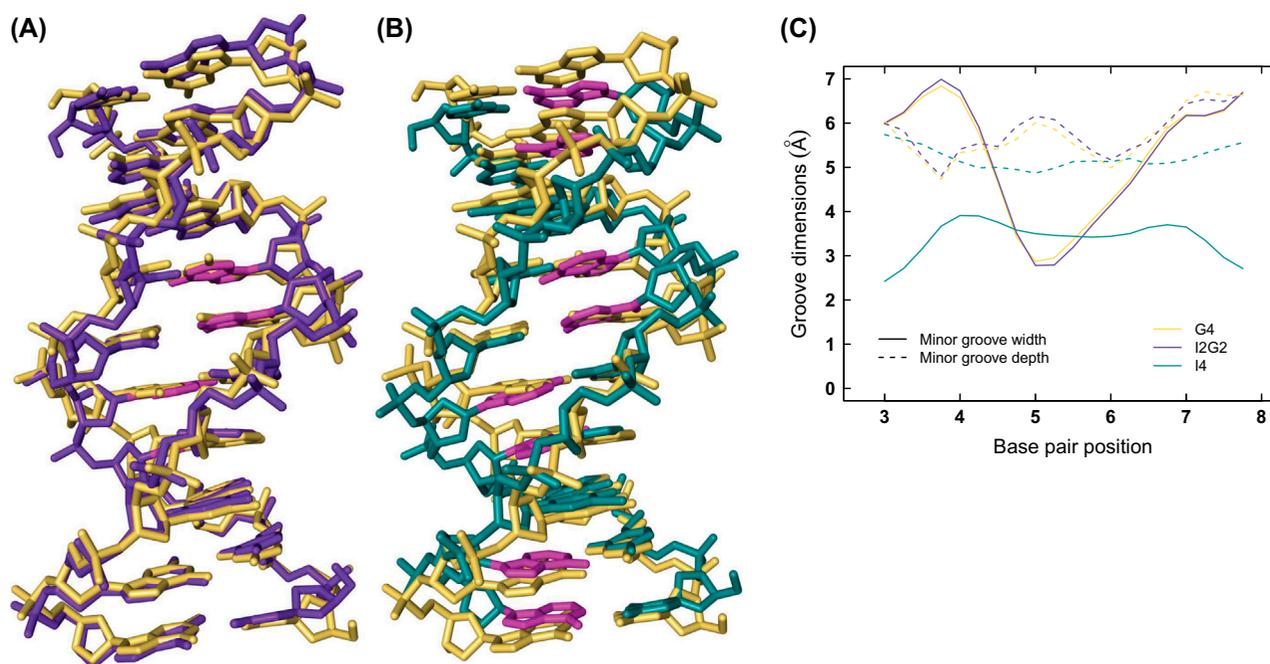


Figure 10. Crystal structures of current inosine-substituted derivatives (see Table 2). I nucleobases are highlighted in magenta. (A) Superimposition of d(CCAGGCCTGG) (G4, beige; PDB ID 3U89) (Maehigashi et al., 2012) and d(CCAIICCTGG) (I2G2, purple) duplexes, corresponding to the relative orientation with the smallest root mean squared deviation. (B) Superimposition of d(CCAGGCCTGG) (G4, beige) and d(CCAIICC[5BrU]II) (I4, deep teal) demonstrates the decreased overall length of the latter duplex. (C) Minor groove width (solid lines) and depth (dashed lines) in Å for the G4 (beige), I2G2 (purple) and I4 (deep teal) duplexes, illustrating the significantly contracted minor groove in the I4 duplex. The two parameters were calculated with the program CURVES (Lavery & Sklenar, 1996, 1989).

addition a Na^+ ion lies adjacent to BrU in the minor groove. Because I4 features two independent duplexes, Figure 11(C) depicts ion position information from the two minor grooves in one duplex.

The unusual conformation of I4 is intriguing considering the unique pattern of metal ion coordination. Metal ion interactions may play a key role in stabilizing this DNA conformation. It is possible that disruption of groove hydration (e.g. by dehydrating agents) could drive a transition to a more conventional DNA conformation, reducing AI-tract-dependent curvature.

4. Summary and conclusions

Motivated by our previous finding based on CD spectroscopy that base analogs can drive new DNA conformations with non-canonical physical properties, we wished to characterize these non-B conformations (Figure 1). Specifically, we hypothesized that X-ray crystallography would reveal inosine-driven changes in DNA duplex helical conformation. A literature review showed that single-I or double-I substitutions typically preserve conventional A-DNA and B-DNA in crystal structures. From CD experiments it is apparent that complete G to I substitution induced unusual DNA conformations in longer constructs, and this was observed as well in deca-

meric DNA duplexes, with CD changes becoming more pronounced for duplexes with higher numbers of I substitutions (e.g. I4 duplex vs. I2G2 duplex). Our crystallography experiments demonstrate that the I2G2 decamer with two I substitutions per strand and four I-C pairs per duplex adopts a conformation that is similar to that of the G4 reference duplex without inosine. However, the crystal structure of the I4 decamer with four I substitutions per strand and eight I-C pairs per duplex displays striking changes relative to the conformation of the G4 duplex. Distinguishing features include base pair rise, sugar conformation, axis curvature, groove topology, and groove occupancy by water and ions. To the extent that the duplex with almost all base pairs replaced by I-C shows a marked change in conformation relative to canonical B-form DNA (e.g. G4 duplex), the crystallographic data are consistent with results from solution studies wherein increasing levels of substitution trigger more pronounced deviations in the CD spectra. However, the X-ray structure still lies within the B-DNA family of conformations (c.f. the blue symbols in Figure 9). Based on the I4 crystal structure, we can conclude that hallmarks of an I-rich duplex include an extremely narrow minor groove, a contraction in length as a consequence of overall more negative inclination, and a modest curvature of the calculated helical axis in the decamer. How

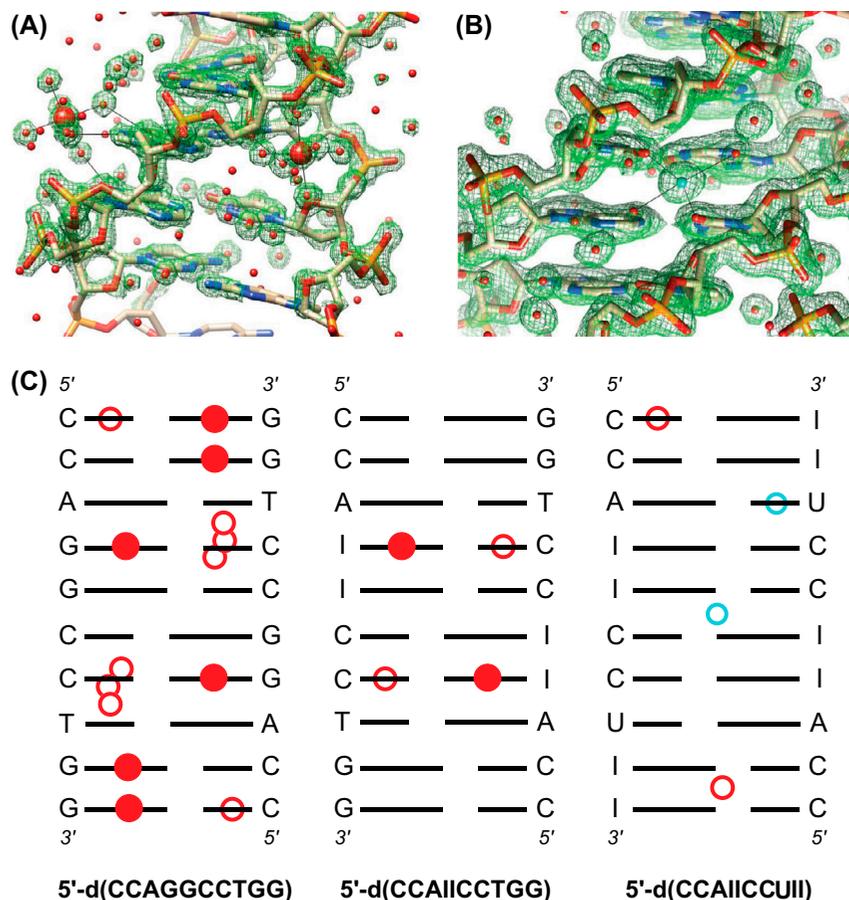


Figure 11. Quality of the electron density in the crystal structures of the I2G2 and I4 duplexes and comparison of the metal ion coordination in the grooves of the G4, I2G2, and I4 duplexes. Final Fourier 2Fo-Fc sum electron density (1.2σ level) in the (A) I2G2 and (B) I4 crystal structures. The I2G2 duplex is viewed across the major (left) and minor (right) grooves, with Mg^{2+} (red spheres of larger radius) hexahydrate complexes coordinating on both sides of an I-C pair in the central portion. The I4 duplex is viewed into the minor groove with a Na^+ ion (cyan sphere) invading the water spine and coordinating to O2 atoms of cytosine residues from opposite strands at the central d(IpC)-d(IpC) base pair step. (C) Schematic illustration of metal ion coordination inside the grooves of the native G4 duplex and the I-substituted I2G2 and I4 duplexes (from left to right). The view is into the major groove, filled circles in red indicate Mg^{2+} hexahydrate ions inside the major groove, and open red and cyan circles indicate Mg^{2+} hexahydrate and Na^+ , respectively, inside the minor groove. G4 sits on a crystallographic dyad and the three open circles next to T and C in the minor groove indicate three partially occupied ions (labeled (A), (B), and (C) in PDB 3U89). The asymmetric unit of the I2G2 structure contains one duplex in a general location. The asymmetric unit of the I4 structure contains two independent duplexes and the schematic at the right depicts a composite of the ion coordination to these two duplexes, i.e. ions coordinating to them were combined inside the minor groove of a single duplex.

these properties might contribute to the observed changes in the CD spectra of I-substituted DNA duplexes in solution remains an open question.

This project was motivated by an analysis of DNA mechanics where we assume that DNA conformation is homogenous and without anomalous junctions between DNA segments of altered conformation (Schurr, Delrow, Fujimoto, & Benight, 1997). It is formally possible that unusual CD spectra for base-substituted DNAs are due to isolated segments of altered conformation, with intervening junctions that might explain the observed changes in mechanical properties. However, we reasoned that any such junctions would manifest through substantial reduc-

tion of melting temperature and sites of enhanced bend flexibility through kinking. This was not observed (Peters et al., 2013, 2014), nor was there evidence of spontaneous kinking in base-substituted DNA molecules ~750 bp long visualized using atomic force microscopy (Peters et al., 2014).

The necessarily static picture of DNA derived from X-ray crystallography studies is contrary to the many experimental and theoretical studies that reveal that DNA is highly polymorphic and dynamic. Indeed, divalent metal ions and packing forces play a significant role in the structure of short DNA duplexes determined by crystallography (Minasov, Tereshko, & Egli, 1999). For instance,

when the structure of the A-DNA decamer d (GCGGGCCCGC) was studied in two different crystal forms (Ramakrishnan & Sundaralingam, 1993a), the authors concluded that the crystal environment dominated base sequence effects on DNA conformation (Malinina et al., 1998; Ramakrishnan & Sundaralingam, 1993b). However, study of the same DNA sequence in different crystal packing arrangements can provide useful insight into structural variability. This conundrum is analogous to the challenges of X-ray crystallography applied to understand apparent DNA curvature induced by A-tract sequences. These crystal structures showed more straight conformations, whereas solution nuclear magnetic resonance (NMR) structures of A-tract DNA showed more pronounced helix curvature, consistent with the extensive evidence from anomalous gel mobility experiments of phased A-tracts (Brahms & Brahms, 1990; Barbič, Zimmer, & Crothers, 2003; MacDonald, Herbert, Zhang, Polgruto, & Lu, 2001; Steff, Wu, Ravindranathan, Sklenár, & Feigon, 2004; Young et al., 1995). Despite this limitation that crystal-packing or lattice effects tend to suppress differences in helical geometry detected by NMR or CD in solution, X-ray analysis has demonstrated that alternating purine regions, e.g. AIA, are structurally similar to pure A-tracts (Shatzky-Schwartz et al., 1997), consistent with our findings for the I-rich duplex. Moreover, differences in the behavior of I·C vs. A·T base pairs became more pronounced as additional I substitutions were made, consistent with our CD and crystal structure data. Additionally, X-ray analysis of the I-tract-containing decamer d (CCIIICCCGG), showed that runs of inosines share the structural feature of a narrow minor groove with A-tracts (Shatzky-Schwartz et al., 1997).

Here we have shown that high-resolution X-ray structures of short DNA duplexes with moderate I substitution preserve B-form DNA features in crystals, despite the fact that in solution these molecules exhibit non-B form character by CD spectroscopy, even under crystallization conditions. Moreover, highly I-substituted oligonucleotides with pronounced non-canonical helical conformations apparent by CD displayed altered helical conformations that were less pronounced by crystallography. High-resolution structural methods other than X-ray crystallography may be necessary to fully resolve this ongoing mystery.

Supplementary material

The supplementary material for this paper is available online at <https://doi.org/10.1080/07391102.2017.1369164>.

Acknowledgments

We thank Marina Ramirez-Alvarado for instrument access and Zdzislaw Wawrzak for help with X-ray diffraction data collection and processing. The authors also thank Loren Williams for advice during the early stages of this work.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Mayo Foundation, Vanderbilt University, Mayo Edward C. Kendall Fellowship (JPP), and the National Institutes of Health [grant number GM75965 to LJM and GM071461 to ME].

ORCID

L. James Maher III  <http://orcid.org/0000-0002-5043-6422>

References

- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., ... Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, 66, 213–221.
- Anosova, I., Kowal, E. A., Dunn, M. R., Chaput, J. C., Van Horn, W. D., & Egli, M. (2016). The structural diversity of artificial genetic polymers. *Nucleic Acids Research*, 44, 1007–1021.
- Balaceanu, A., Pasi, M., Dans, P. D., Hospital, A., Lavery, R., & Orozco, M. (2017). The role of unconventional hydrogen bonds in determining BII propensities in B-DNA. *The Journal of Physical Chemistry Letters*, 8, 21–28.
- Barbič, A., Zimmer, D. P., & Crothers, D. M. (2003). Structural origins of adenine-tract bending. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 2369–2373.
- Berger, I., Tereshko, V., Ikeda, H., Marquez, V. E., & Egli, M. (1998). Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymine: Implications of conformational preorganization for duplex stability. *Nucleic Acids Research*, 26, 2473–2480.
- Brahms, S., & Brahms, J. G. (1990). DNA with adenine tracts contains poly(dA).poly(dT) conformational features in solution. *Nucleic Acids Research*, 18, 1559–1564.
- Brown, T., Hunter, W. N., Kneale, G., & Kennard, O. (1986). Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 83, 2402–2406.
- Brown, T., Leonard, G. A., Booth, E. D., & Chambers, J. (1989). Crystal structure and stability of a DNA duplex containing A(anti).G(syn) base-pairs. *Journal of Molecular Biology*, 207, 455–457.
- Case-Green, S. C., & Southern, E. M. (1994). Studies on the base pairing properties of deoxyinosine by solid phase hybridisation to oligonucleotides. *Nucleic Acids Research*, 22, 131–136.
- Chen, X., Mitra, S. N., Rao, S. T., Sekar, K., & Sundaralingam, M. (1998). A novel end-to-end binding of two netropsins to the DNA decamers d(CCCCCIII)2, d(CCCBr 5CCIII)2 and d(CBr5CCCCIII)2. *Nucleic Acids Research*, 26, 5464–5471.
- Collaborative Computational Project, N. (1994). The CCP4 suite: Programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography*, 50, 760–763.

- Corfield, P. W., Hunter, W. N., Brown, T., Robinson, P., & Kennard, O. (1987). Inosine.adenine base pairs in a B-DNA duplex. *Nucleic Acids Research*, *15*, 7935–7949.
- Cruse, W. B., Aymani, J., Kennard, O., Brown, T., Jack, A. G., & Leonard, G. A. (1989). Refined crystal structure of an octanucleotide duplex with I.T. mismatched base pairs. *Nucleic Acids Research*, *17*, 55–72.
- Dans, P. D., Danilane, L., Ivani, I., Drsata, T., Lankas, F., Hospital, A., ... Orozco, M. (2016). Long-timescale dynamics of the Drew-Dickerson dodecamer. *Nucleic Acids Research*, *44*, 4052–4066.
- Diekmann, S., von Kitzing, E., McLaughlin, L., Ott, J., & Eckstein, F. (1987). The influence of exocyclic substituents of purine bases on DNA curvature. *Proceedings of the National Academy of Sciences of the United States of America*, *84*, 8257–8261.
- Ding, D., Gryaznov, S. M., & Wilson, W. D. (1998). NMR solution structure of the N3' → P5' phosphoramidate duplex d(CGCGAATTCGCG)2 by the iterative relaxation matrix approach. *Biochemistry*, *37*, 12082–12093.
- Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981). Structure of a B-DNA dodecamer: Conformation and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *78*, 2179–2183.
- Drsata, T., Spackova, N., Jurecka, P., Zgarbova, M., Sponer, J., & Lankas, F. (2014). Mechanical properties of symmetric and asymmetric DNA A-tracts: Implications for looping and nucleosome positioning. *Nucleic Acids Research*, *42*, 7383–7394.
- Egli, M. (2016). Diffraction techniques in structural biology. *Current Protocols in Nucleic Acid Chemistry*, *65*, 7.13.1–7.13.41.
- Egli, M., Lubini, P., & Pallan, P. S. (2007). The long and winding road to the structure of homo-DNA. *Chemical Society Reviews*, *36*, 31–45.
- Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., Sanishvili, R., ... Manoharan, M. (1998). X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers*, *48*, 234–252.
- El Hassan, M. A., & Calladine, C. R. (1996). Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *Journal of Molecular Biology*, *259*, 95–103.
- El Hassan, M. A., & Calladine, C. R. (1997). Conformational characteristics of DNA: Empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *355*, 43–100.
- Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, *60*, 2126–2132.
- Fishel, R., Anziano, P., & Rich, A. (1990). [35] Z-DNA affinity chromatography. *Methods in Enzymology*, *184*, 328–340.
- Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, *171*, 740–741.
- Friedman, J., & Tibshirani, R. (1984). The monotone smoothing of scatterplots. *Technometrics*, *26*, 243–250.
- Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A., & Wang, A. H. (1989). The molecular structure of the left-handed Z-DNA double helix at 1.0-Å atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *Journal of Biological Chemistry*, *264*, 7921–7935.
- Harteis, S., & Schneider, S. (2014). Making the bend: DNA tertiary structure and protein-DNA interactions. *International Journal of Molecular Sciences*, *15*, 12335–12363.
- Hartmann, B., Piazzola, D., & Lavery, R. (1993). B I - B II transitions in B-DNA. *Nucleic Acids Research*, *21*, 561–568.
- Hays, F. A., Jones, Z. J., & Ho, P. S. (2004). Influence of minor groove substituents on the structure of DNA Holliday junctions. *Biochemistry*, *43*, 9813–9822.
- Heath, P. J., Clendenning, J. B., Fujimoto, B. S., & Schurr, J. M. (1996). Effect of bending strain on the torsion elastic constant of DNA. *Journal of Molecular Biology*, *260*, 718–730.
- Heinemann, U., Alings, C., & Hahn, M. (1994). Crystallographic studies of DNA helix structure. *Biophysical Chemistry*, *50*, 157–167.
- Holbrook, S. R., Dickerson, R. E., & Kim, S. H. (1985). Anisotropic thermal-parameter refinement of the DNA dodecamer CGCGAATTCGCG by the segmented rigid-body method. *Acta Crystallographica Section B-Structural Science*, *41*, 255–262.
- Hud, N. V., & Plavec, J. (2003). A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, *69*, 144–158.
- Hud, N. V., & Polak, M. (2001). DNA-cation interactions: The major and minor grooves are flexible ionophores. *Current Opinion in Structural Biology*, *11*, 293–301.
- Hunter, W. N., Brown, T., & Kennard, O. (1986). Structural features and hydration of d(C-G-C-G-A-A-T-T-A-G-C-G); a double helix containing two G.A mispairs. *Journal of Biomolecular Structure and Dynamics*, *4*, 173–191.
- Imeddourene, A. B., Xu, X., Zargarian, L., Oguey, C., Foloppe, N., Mauffret, O., & Hartmann, B. (2016). The intrinsic mechanics of B-DNA in solution characterized by NMR. *Nucleic Acids Research*, *44*, 3432–3447.
- Janke, E. M., Riechert-Krause, F., & Weisz, K. (2011). Low-temperature NMR studies on inosine wobble base pairs. *The Journal of Physical Chemistry B*, *115*, 8569–8574.
- Koo, H. S., & Crothers, D. M. (1987). Chemical determinants of DNA bending at adenine-thymine tracts. *Biochemistry*, *26*, 3745–3748.
- Kumar, V. D., Harrison, R. W., Andrews, L. C., & Weber, I. T. (1992). Crystal structure at 1.5-Å resolution of d(CGCI-CICG), an octanucleotide containing inosine, and its comparison with d(CGCG) and d(CGCGCG) structures. *Biochemistry*, *31*, 1541–1550.
- Kypr, J., Kejnovska, I., Renciuik, D., & Vorlickova, M. (2009). Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research*, *37*, 1713–1725.
- Lavery, R., & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *Journal of Biomolecular Structure and Dynamics*, *6*, 63–91.
- Lavery, R., & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: Conventions and principles. *Journal of Biomolecular Structure and Dynamics*, *6*, 655–667.
- Lavery, R., & Sklenar, H. (1996). Curves 5.1: Helical analysis of irregular nucleic acids. *Institut de Biologie PhysicoChimique*.
- Leonard, G. A., Booth, E. D., Hunter, W. N., & Brown, T. (1992). The conformational variability of an adenosine.inosine base-pair in a synthetic DNA dodecamer. *Nucleic Acids Research*, *20*, 4753–4759.
- Leslie, A. G., Arnott, S., Chandrasekaran, R., & Ratliff, R. L. (1980). Polymorphism of DNA double helices. *Journal of Molecular Biology*, *143*, 49–72.

- Vagin, A., & Teplyakov, A. (1997). MOLREP : An automated program for molecular replacement. *Journal of Applied Crystallography*, *30*, 1022–1025.
- van Dam, L., & Levitt, M. H. (2000). BII nucleotides in the B and C forms of natural-sequence polymeric DNA: A new model for the C form of DNA. *Journal of Molecular Biology*, *304*, 541–561.
- Vargason, J. M., Eichman, B. F., & Ho, P. S. (2000). The extended and eccentric E-DNA structure induced by cytosine methylation or bromination. *Nature Structural Biology*, *7*, 758–761.
- Vargason, J. M., & Ho, P. S. (2002). The effect of cytosine methylation on the structure and geometry of the Holliday junction: The structure of d(CCGGTACm5CGG) at 1.5 Å resolution. *Journal of Biological Chemistry*, *277*, 21041–21049.
- Vorlickova, M., & Sagi, J. (1991). Transitions of poly(dI-dC), poly(dI-methyl5dC) and poly(dI-bromo5dC) among and within the B-, Z-, A- and X-DNA families of conformations. *Nucleic Acids Research*, *19*, 2343–2347.
- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., & Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, *282*, 680–686.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *171*, 737–738.
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature*, *287*, 755–758.
- Woods, K. K., Maehigashi, T., Howerton, S. B., Sines, C. C., Tannenbaum, S., & Williams, L. D. (2004). High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]₂. *Journal of the American Chemical Society*, *126*, 15330–15331.
- Xuan, J. C., & Weber, I. T. (1992). Crystal structure of a B-DNA dodecamer containing inosine, d(CGCAATTCGCG), at 2.4 Å resolution and its comparison with other B-DNA dodecamers. *Nucleic Acids Research*, *20*, 5457–5464.
- Young, M. A., Srinivasan, J., Goljer, I., Kumar, S., Beveridge, D. L., & Bolton, P. H. (1995). [5] Structure determination and analysis of local bending in an A-tract DNA duplex: Comparison of results from crystallography, nuclear magnetic resonance, and molecular dynamics simulation on d(CGCAAAAATGCG). *Methods in Enzymology*, *261*, 121–144.
- Zheng, G., Lu, X.-J., & Olson, W. K. (2009). Web 3DNA – a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Research*, *37*, W240–W246.