2208–2215 *Nucleic Acids Research, 1998, Vol. 26, No. 9*

Exploratory studies on azole carboxamides as nucleobase analogs: thermal denaturation studies on oligodeoxyribonucleotide duplexes containing pyrrole-3-carboxamide

Peiming Zhang[1], W. Travis Johnson[2], Douglas Klewer[2], Natasha Paul[2], Geoffrey Hoops[2], V. J. Davisson[2] and Donald E. Bergstrom[1,2,*]

[1]Walther Cancer Institute, Indianapolis, IN 46208, USA and [2]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907, USA

Received November 12, 1997; Revised and Accepted March 16, 1998

ABSTRACT

In order to study base pairing properties of the amide group in DNA duplexes, a nucleoside analog, 1-(2′-deoxy-β-D-ribofuranosyl)pyrrole-3-carboxamide, was synthesized by a new route from the ester, methyl 1-(2′-deoxy-3′,5′-di-*O*-*p*-toluoyl-β-D-erythro-pentofuranosyl)pyrrole-3-carboxylate, obtained from the coupling reaction between 1-chloro-2-deoxy-3,5-di-*O*-toluoyl-D-erythropentofuranose and methyl pyrrole-3-carboxylate by treatment with dimethylaluminum amide. 1-(2′-Deoxy-β-D-ribofuranosyl)pyrrole-3-carboxamide was incorporated into a series of oligodeoxyribonucleotides by solid-phase phosphoramidite technology. The corresponding oligodeoxyribonucleotides with 3-nitropyrrole in the same position in the sequence were synthesized for UV comparison of helix–coil transitions. The thermal melting studies indicate that pyrrole-3-carboxamide, which could conceptually adopt either a dA-like or a dI-like hydrogen bond conformation, pairs with significantly higher affinity to T than to dC. Pyrrole-3-carboxamide further resembles dA in the relative order of its base pairing preferences (T > dG > dA > dC). Theoretical calculations on the model compound *N*-methylpyrrole-3-carboxamide using density functional theory show little difference in the preference for a *syn*$_\tau$ versus *anti*$_\tau$ conformation about the bond from pyrrole C3 to the amide carbonyl. The amide groups in both the minimized *anti*$_\tau$ and *syn*$_\tau$ conformations are twisted out of the plane of the pyrrole ring by 6–14°. This twist may be one source of destabilization when the amide group is placed in the helix. Another contribution to the difference in stability between the base pairs of pyrrole-3-carboxamide with T and pyrrole-3-carboxamide with C may be the presence of a hydrogen bond in the former involving an acidic proton (N3-H of T).

INTRODUCTION

3-Nitropyrrole (**1**) and its analog 5-nitroindole have been studied as universal bases which function as wild cards in base pairing within nucleic acid duplexes (1–5). Although thermal denaturation studies of 3-nitropyrrole containing oligonucleotides show that it is non-discriminating in natural base recognition, 3-nitropyrrole is highly discriminating as a template for DNA polymerase. It was found that 3-nitropyrrole functioned as both an A and a T analog. A and T were incorporated opposite nitropyrrole in a 7:3 ratio by Taq polymerase, while C or G did not appear to be incorporated at all (6). These results, along with the substantial loss in duplex stability that occurs with nitropyrrole substitution, support the need to find alternative solutions to the universal base problem. For most applications an ideal universal base should indiscriminately pair with each of the natural bases with an affinity as high as a natural Watson–Crick base pair. As discussed previously, enhancing base stacking interactions by extending the conjugated π-system (e.g. 5-nitroindole) increased oligonucleotide T_m values compared to sequences containing 3-nitropyrrole, but the increments are still not great enough to compensate the loss in duplex stability caused by nitropyrrole (7). There appears to be an inherent limit to the extent to which non-hydrogen bonding heterocycles can stabilize a duplex. As a second alternative, we have been exploring the construction of nucleobase surrogates that are configured to allow flexible hydrogen bonding patterns that mimic the natural bases. Although it is possible to design a base analog that can associate with each of the natural bases through two hydrogen bonds (8), the difficulty of building and incorporating such an analog into oligonucleotides suggests that it may be more efficient to focus on analogs that are limited to base paring with just two or three of the natural bases. Previously we showed that imidazole-4-carboxamide deoxyribonucleotide (**2**) could base pair to either T or G by assuming an A-like or a C-like hydrogen bonding configuration (9). The related analog, pyrrole-3-carboxamide deoxyribonucleotide (**3**), could exist in two different conformations accessible through rotating the amide about bond τ that would

*To whom correspondence should be addressed at: Hansen 401A, Purdue University, West Lafayette, IN 47907, USA. Tel: +1 765 494 6275; Fax: +1 765 494 9193; Email: bergstrom@pharmacy.purdue.edu

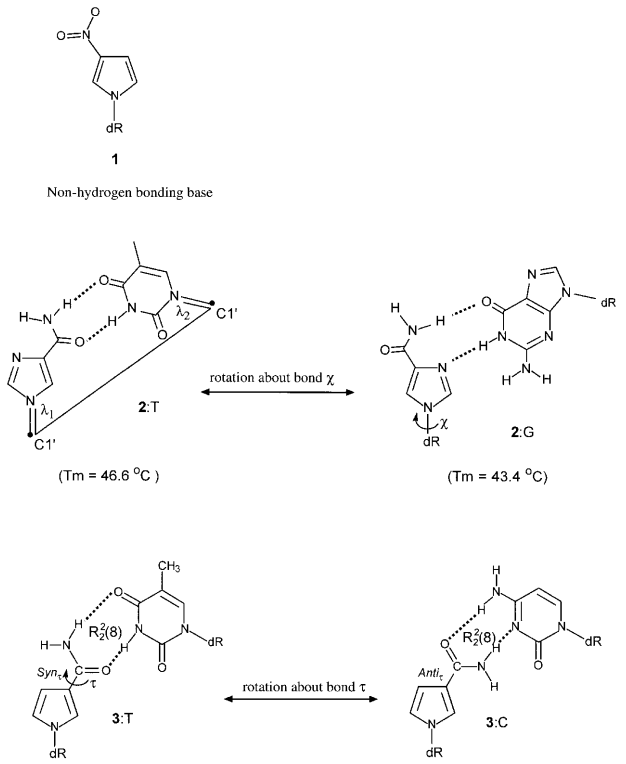


Figure 1. Rotation about bonds τ and χ would allow the amide to form hydrogen bonds without significantly disturbing the duplex backbones. The hydrogen bonding patterns assigned to the base pairs **3**:C and **3**:T are, according to graph-set notation, $R_2^2(8)$ (63–65). All hydrogen bonding patterns in the Watson–Crick base pairs are categorized as $R_2^2(8)$ (66). The T_m data was adapted from our previous studies (7,9) which used the same sequence as this study.

allow it to pair preferentially with either C or T as illustrated in Figure 1. In contrast to **2**, it was anticipated that the absence of the second nitrogen atom in the heterocyclic ring would significantly lessen the likelihood of forming a pyrimidine like hydrogen bonding pattern with **3**.

In this study, we have extended the theoretical calculations with density functional theory (DFT) methods and obtained complete thermal melting data as a means to determine the relative stability of base pairs between pyrrole-3-carboxamide and each of the natural bases.

MATERIALS AND METHODS

General procedures

NMR spectra were recorded using a Bruker AC 250 or a Varian VXR-500S spectrometer. ^{1}H and ^{13}C signals were internally referenced to TMS while 85% phosphoric acid was utilized as an external standard for all ^{31}P spectra. The ^{31}P spectra were not corrected for bulk susceptibility. FAB and MALDI mass spectra were recorded by the mass spectroscopy laboratories, Department of Medicinal Chemistry and Molecular Pharmacology or Department of Biochemistry, respectively, Purdue University. Elemental analysis was performed by the Microanalysis Laboratory, Department of Chemistry, Purdue University. Analytical thin layer chromatography (TLC) was carried out on pre-coated Whatman 60 F_{254} plates. Chromatotron preparative chromatography plates were prepared using silica gel 60 PF_{254} containing a gypsum binding agent manufactured by EM Science. Anhydrous solvents were freshly distilled from the appropriate drying agents or purchased from Aldrich Chemical Company. All chemicals were of reagent grade or better quality and used as received. DNA phosphoramidites and synthesis reagents made by Cruachem were purchased from Fisher Scientific.

Methyl 1-(2′-deoxy-3′,5′-di-*O*-*p*-toluoyl-β-D-erythro-pentofuranosyl)pyrrole-3-carboxylate (6)

Sodium hydride (100 mg, 4.2 mmol) was suspended in 35 ml of acetonitrile in an inert atmosphere. Methyl pyrrole-3-carboxylate (**5**) (360 mg, 2.9 mmol) was added to the suspension and the mixture was stirred for an additional 15 min. After the evolution of hydrogen gas, 1-chloro-2-deoxy-3,5-di-*O*-toluoyl-D-erythro-pentofuranose (**4**) (1.16 g, 3.0 mmol) was added to the mixture and the solution was stirred for 3 h at room temperature. The reaction mixture was filtered over Celite and the solvent removed under reduced pressure. The resulting residue was purified by silica gel chromatography eluted with a gradient of ethanol in hexane from 5% to 50%. Removal of the solvents and drying under vacuum yielded 1.1 g (80.6%) of **6**. ^{1}H NMR (250 MHz, CD_3OD) δ 2.42 (toluoyl-C$\underline{H}_3$, s, 3H), 2.43 (toluoyl-C$\underline{H}_3$, s, 3H), 2.67 (H-2′, m, 2H), 3.76 (-OC$\underline{H}_3$, s, 3H), 4.55 (H-4′, m, 1H), 4.61 (H-5′, m, 2H), 5.64 (H-3′, m, 1H), 6.03 (H-1′, t, J = 6.72 Hz, 1H), 6.60 (H-5, dd, J = 1.65, 3.04 Hz, 1H), 6.81 (H-4, m, 1H), 7.26 (toluoyl-H, m, 4H), 7.50 (H-2, t, J = 1.9 Hz, 1H), 7.93 (toluoyl-H, m, 4H). CI *m/z*: 478 (M^+H) Anal. Calcd for $C_{27}H_{27}NO_7$: C, 67.90; H, 5.70; N, 2.90. Found: C, 67.51; H, 5.66; N, 3.14.

1-(2′-Deoxy-β-D-ribofuranosyl)pyrrole-3-carboxamide (3)

To a solution of ammonia (420 mg, 24.7 mmol) in methylene chloride (10 ml) was added a 2.0 M solution of trimethyl aluminum in hexane (12.4 ml, 24.7 mmol) at –70°C under argon. After 30 min, compound **6** (100 mg, 0.21 mmol) was added to this solution. The solution was heated to reflux for 24 h. The reaction was carefully quenched with water. The solvent was removed by rotary evaporation. The residue was extracted by methanol, and the methanol solution was evaporated to dryness. The resulting residue was dissolved in water. The aqueous solution was extracted with chloroform, filtered and lyophilized. Product **3** was separated in 40% yield by chromatography on silica gel using a mixture of chloroform/methanol (4:1) as the eluent. The NMR spectra was identical to that reported (10).

1-(2′-Deoxy-5′-*O*-dimethoxytrityl-β-D-ribofuranosyl)pyrrole-3-carboxamide (7)

To a solution of 1-(2′-deoxy-β-D-ribofuranosyl)pyrrole-3-carboxamide (**3**) (268 mg, 1.18 mmol) in pyridine (3 ml) was added 4,4′-dimethoxytrityl chloride (442 mg, 1.30 mmol). The mixture was stirred at room temperature for 1 h. TLC analysis indicated the presence of a small amount of starting material. Additional 4,4′-dimethoxytrityl chloride was added to complete the reaction. The mixture was poured into water (50 ml), and extracted with methylene chloride (3 × 50 ml). The combined organic phase was washed with water and dried over anhydrous Na_2SO_4. The product was separated by silica gel chromatography on a chromatotron using a mixture of methylene chloride/methanol (95:5) as the eluent. Compound **7** was obtained as a white foam

(467 mg, 75%). ^{1}H NMR (DMSO-d6): δ 7.47 (H_5, t, J = 2.0 Hz, 1H), 7.36–7.18 (DMTr-H, m, 9H), 6.88 (H-4, t, J = 2.0 Hz, 1H), 6.86–6.84 (DMTr-H, m, 4H), 6.73 (NH, br s, 1H), 6.45 (H-2, dd, J = 1.5, 3.0 Hz, 1H), 5.92 (H-1′, t, J = 7.5 Hz, 1H), 5.29 (H-3′-OH, d, J = 4.5 Hz, 1H), 4.24–4.22 (H-3′, m, 1H), 3.89–3.86 (H-4′, m, 1H), 3.72 (OCH_3, s, 6H), 3.08 (H-5′, d, J = 5.0 Hz, 1H), 2.34–2.29 (H-2′, m, 1H), 2.24–2.19 (H-2″, m, 1H). FAB *m/z*: 303.0 ($DMTr^+$), 528.2 (M^+). Anal. Calcd for $C_{31}H_{32}N_2O_6$: C, 70.44; H, 6.10; N, 5.30. Found: C, 70.08; H,6.24; N, 5.50.

1-(2′-Deoxy-5′-*O*-dimethoxytrityl-β-D-ribofuranosyl)pyrrole-3-carboxamide-3′-*O*-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite) (8)

To a solution of compound **7** (300 mg, 0.57 mmol) and diisopropylammonium tetrazolide (97 mg, 0.57 mmol) in methylene chloride (6 ml) was added 2-cyanoethyl *N*,*N*,*N*′,*N*′-tetraisopropylphosphorodiamidite (0.21 ml, 0.66 mmol). The solution was gently swirled, allowed to stand under nitrogen at room temperature for 1.5 h, and then diluted with ethyl acetate, washed with water, and dried over anhydrous Na_2SO_4. The product was separated by silica gel chromatography on a chromatotron using a mixture of ethyl acetate/triethylamine (98:2) as the eluent. Compound **8** was obtained as a white foam (316 mg, 76%). ^{31}P NMR (acetone-d6) δ 148.55 and 148.48 (3′-O-P); ^{1}H NMR (acetone-d6) δ 7.52–7.51 (H-5, m, 1H), 7.49–7.19 (DMTr-H, m, 9H), 6.94–6.91 (H-4, m, 1H), 6.89–6.85 (DMTr-H, m, 1H), 6.52–6.51 (H-2, m, 1H), 6.02 (H-1′, t, J = 7.0 Hz, 1H), 4.69–4.63 (H-3′, m, 4H), 4.22–3.99 (H-4′, m, 1H), 3.92–3.47 (OCH_2CH_2- and -CH, m, 5H), 3.34–3.27 (H-5′, m, 2H), 3.78 and 3.77 (OCH_3, s, 6H), 2.59–2.46 (H-2′ and H-2″, m, 2H), 1.26–1.09 (-CCH3, m, 12H); FAB *m/z*: 729.25 (MH^+).

Synthesis and characterization of oligodeoxyribonucleotides

Oligodeoxyribonucleotides were prepared on a Milligen/BioSearch 8700 DNA synthesizer (1 µmol scale) by standard solid-phase phosphoramidite chemistry. Stepwise yields were evaluated by monitoring the UV absorbance of the release of the dimethoxytrityl cation. Coupling efficiencies ranged from 96 to 98%. The detritylated oligonucleotides were released from the CPG and simultaneously deprotected by treatment with conc. NH_3 at 55°C for 8 h. The oligonucleotides were purified using 20% polyacrylamide–8 M urea preparative gel electrophoresis. The desired oligonucleotides were extracted from the gels and desalted with Waters C_{18} SepPaks^{TM} following the manufacturer's instructions. The purified oligomers were lyophilized to dryness and stored at –10°C. Oligodeoxyribonucleotides were characterized by MALDI mass spectrometry (Table 1).

Melting experiments and thermodynamic analysis

UV melting measurements were carried out as previously described (7). Analysis of melting curves was accomplished with the equation developed by Gralla and Crothers (11):

$$\partial\alpha/\partial(1/T) = -\alpha(1 - \alpha)\Delta H/(1 + \alpha)R$$

where α is the fraction of strands bound in a helix. The transition enthalpy was calculated from the equation:

$$\Delta H \text{ (cal/mol)} = -4.37/(1/T_{max} - 1/T_{3/4}),$$

where T_{max} is the temperature (in Kelvin) at the maximum of the differential melting curves and $T_{3/4}$ is the temperature at the upper half-height of the differential melting curves. In this analysis, ΔCp was assumed to be zero (12). Thus, enthalpy and entropy are independent of temperature. For T_{max} and $T_{3/4}$ calculations, data from the UV melting measurements was exported into Igor (Wavemetrics, Inc.) for curve fitting. Since the oligonucleotides containing azole bases had low transition temperatures, it was difficult to define the lower baselines of the melting curves for creating α curves. As a compromise, $\partial\alpha/\partial(1/T)$ was substituted with $\partial A/\partial T$ (A = absorbance) in the data process, which may introduce an error of ~3% (11). The differentiated curves were fitted with a Gaussian function following smoothing by the Savitzky–Golary method (13). The T_{max} and $T_{3/4}$ values were calculated from the Gaussian function. Calculations involving these functions and all other mathematical expressions were done in the software application Mathematica (Wolfram Research).

For self-complementary sequences, the equilibrium constant is expressed as:

$$K = \alpha/2(1 - \alpha)^2C_t,$$

where C_t is the total strand concentration (14). At $T = T_{max}$, where α = 0.414, it can be seen that:

$$K_{Tmax} = 0.603/C_t$$

The free energy at $T = T_{max}$ is calculated from the equation:

$$\Delta G_{Tmax} = -RT_{max}\ln K_{Tmax}$$

Entropy values are then calculated from the relationship:

$$\Delta S = (\Delta H - \Delta G_{Tmax})/T_{max}$$

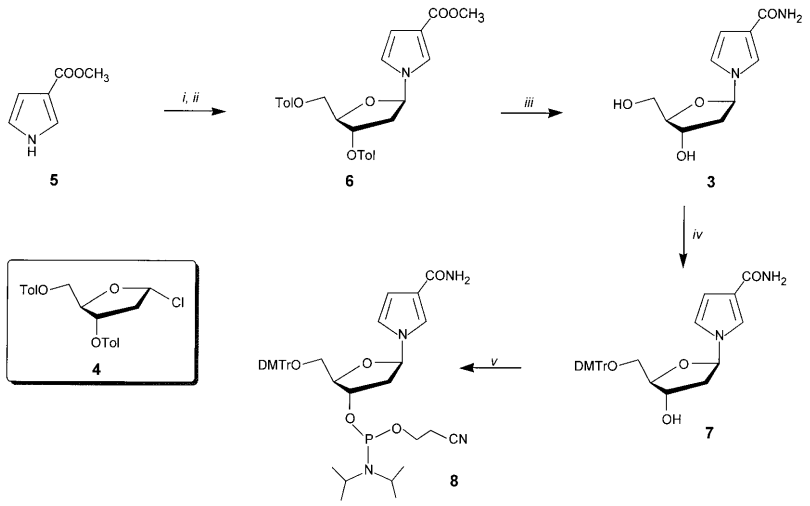
The melting temperatures were calculated from the equation:

$$T_m = \Delta H/(R\ln C_t + \Delta S)$$

Errors on the thermodynamic parameters were calculated by standard methods (15) from the estimated errors of the curve fitting coefficients. Errors estimated for ΔH were ≤1.9, ≤2.5 for ΔS, ≤1.0% for $\Delta G_{25°C}$ and ≤0.5% for T_m.

Table 1. Oligonucleotides studied in thermal denaturation experiments [oligo A, d(CGC **X**AA TT**Y** GCG); oligo B, d(CGC GA**M** **M**TC GCG)] and their MALDI-MS characterization

Oligo A **X:Y**	Oligo B **M**	Molecular weight (M-1) Calc.	Found
3:A		3629.1	3628.6
3:C		3605.2	3604.1
3:G		3645.2	3643.3
3:T		3620.2	3621.1
3:**3**		3603.9	3599.9
1:**3**		3605.1	3602.6
1:**1**		3608.8	3607.0
	1	3608.3	3610.1
	3	3603.9	3602.4



Reagents: *i.* acetonitrile, NaH; *ii.* (**4**); *iii.* $(CH_3)_2AlNH_2$, CH_2Cl_2, refluxing; *iv.*dimethoxytrityl chloride, pyridine; *v.* 2-cyanoethyl N,N,N',N'-tetraisopropylphosphorodiamidite, diisopropylammonium tetrazolide, CH_2Cl_2.

Figure 2. Synthesis of 1-(2′-deoxy-5′-*O*-dimethoxytrityl-β-D-ribofuranosyl)pyrrole-3-carboxamide 3′-*O*-(2-cyanoethyl-*N*,*N*-diisopropyl)phosphoramidites.

Computational methods

All calculations were carried out with the Gaussian 94 package on an IBM RS/6000 Computer Cluster at Purdue University Computer Center (16). The molecules under study were constructed and minimized with SYBYL. The structures generated in SYBYL were used as inputs for the Gaussian computations. Molecular geometries were fully optimized by the B3LYP method with the basis set 6–31G(d,p). The Gaussian computational package contains the explanations and abbreviations for the methods and basis sets (17). Frequency calculations were performed at the 6–31G(d,p) level, and no imaginary frequencies were observed. The solvation medium effect was calculated through the self-consistent reaction field (SCRF).

RESULTS AND DISCUSSION

Synthesis

1-(2′-Deoxy-β-D-ribofuranosyl)pyrrole-3-carboxamide (**3**) has been synthesized by a four step reaction sequence starting from 3-cyanopyrrole (10,18). Although 3-cyanopyrrole has been known for many years, its preparation is still laborious (19). Alternatively, we developed a procedure that uses methyl pyrrolecarboxylate (**5**) (20–22) as the starting material, which can be made in high yield on a large scale. The sodium salt of methyl 3-pyrrolecarboxylate (**5**), generated *in situ* with NaH in acetonitrile, reacts rapidly with 1-chloro-2-deoxy-3,5-di-*O*-toluoyl-D-erythro-pentofuranose (**4**) at ambient temperature to give product **6** in 70% yield (Fig. 2).

An attempt to convert the methyl ester to the carboxamide by treatment with ammonia was unsuccessful, even at high temperature and pressure (>200°C, 150 psi) for 2 or 3 days. Weinreb and coworkers reported a method to convert esters to amides by using dimethylaluminum amides under very mild reaction conditions (23). With dimethylaluminum amide generated *in situ* by reaction of trimethylaluminum with ammonia, nucleoside **6** was converted to **3** in 40% yield. The structure of nucleoside **3** was confirmed by comparing its NMR spectra with an authentic sample prepared by the literature method (10,18). For incorporation into oligonucleotides, nucleoside **3** was converted to a nucleoside phosphoramidite. Dimethoxyltrityl chloride was used to protect the 5′-hydroxyl group of **3** to give nucleoside **7**. In the presence of diisopropylammonium tetrazolide, **7** reacted with 2-cyanoethyl-*N*,*N*,*N*′,*N*′-tetraisopropylphosphorodiamidite to produce phosphoramidite **8** in 76% yield. The synthesis of oligonucleotides containing pyrrole-3-carboxamide was carried out in an automated DNA synthesizer. The coupling efficiency of **8** was not different from those of the natural nucleoside phosphoramidites. Nine oligonucleotides containing nucleosides **1** and **3** were synthesized for this study (Table 1). The oligonucleotides in set A contained analogs **1** and **3** in the fourth position from the 5′-end, and A, C, G, T, **1** or **3** in the ninth position. Since these sequences are self complementary, the base in the fourth position pairs with the base in the ninth position upon duplex formation. The oligonucleotides in set B included sequences with two 3-nitropyrroles or pyrrole-3-carboxamides at the sixth and seventh positions.

Thermal denaturation studies

A modified version of the deoxydodecamer d(CGCGAATTCGCG) was chosen to study the base-pairing properties of nucleosides **1** and **3**. This sequence has been shown to exist in a B-form structure in solid and solution by X-ray crystallography (24–26), NMR spectroscopy (27–30) and molecular modeling (31,32). Moreover, both X-ray and NMR studies of this sequence indicate that there is an ordered water train in the minor groove of the AT-rich region termed the ‘spine of hydration’, which possibly stabilizes and increases the rigidity of the B-DNA (33–36). Placing unnatural bases in the fourth and ninth positions in the sequence should not change the structural form of the duplex since they should not disturb the ‘spine of hydration’ in the AT-rich region. The sequence d(CGCGAATTCGCG) has also been of interest for thermodynamic studies (37–39). Normally, in 1 M [Na^+] salt concentration, the melting of duplexes composed of this sequence occurs as a monophasic transition, which represents a two state helix to coil process. However, Breslauer

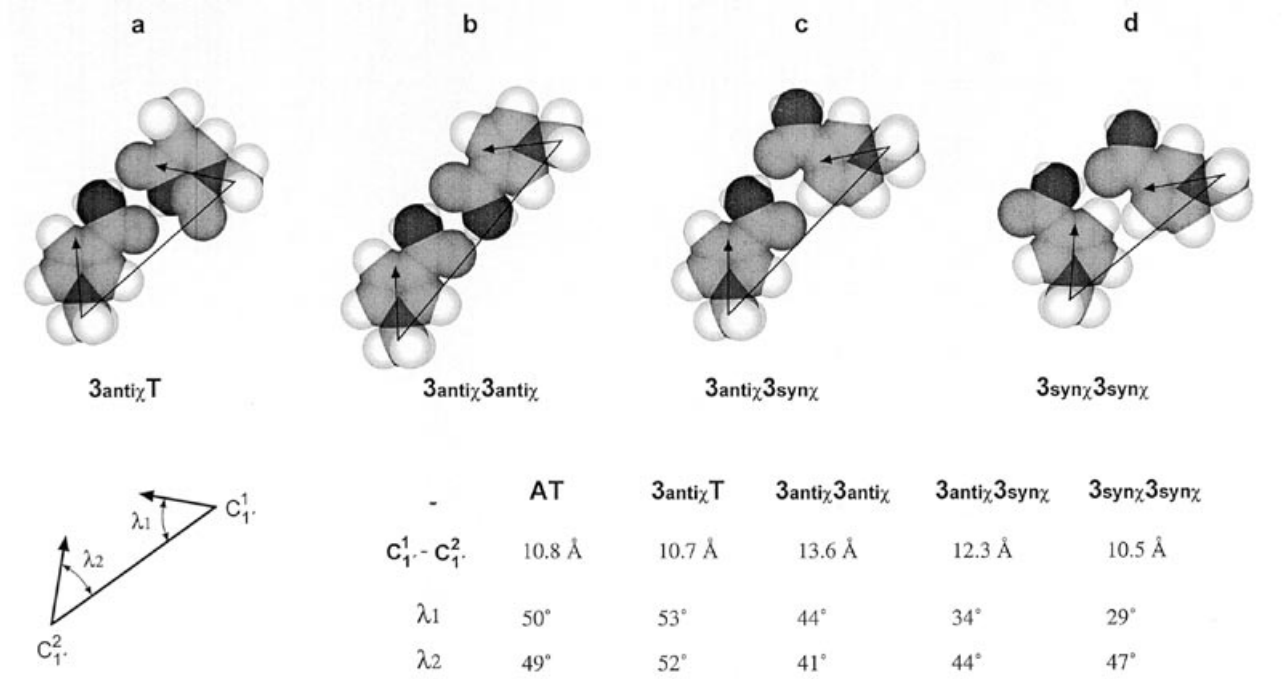


Figure 3. Space filling models and structural parameters for base pairs **3**:T and **3**:**3**. A methyl group is shown in place of C1′ for simplification

and coworkers reported that a biphasic melting profile was obtained when the buffer concentration of [Na^+] was ≤10 mM (40) which they attribute to formation of a hairpin structure.

All of the UV thermal melting experiments in our laboratory were carried out in buffer containing 10 mM phosphate and 1 M NaCl at pH 7. The T_m values and thermodynamic parameters of the duplexes containing nitropyrrole, pyrrole-3-carboxamide, and two sequences containing only natural bases are included in Table 2 for comparison. Duplexes containing either analog have significantly lower T_m and ΔG values than the corresponding natural DNA. As noted previously, 3-nitropyrrole base pairs indiscriminately (7). The difference in ΔG (ΔΔG) between the most and least stable duplexes is only 0.4 kcal/mol. Since each duplex contains two modified base pairs, the ΔΔG value per modified base pair is ~0.2 kcal/mol. On the basis of melting temperature and ΔG, pyrrole-3-carboxamide shows a preference for pairing to the natural bases in the order **3**:T > **3**:G > **3**:A > **3**:C, the same order of preference observed for imidazole 4-carboxamide (9). But, the binding affinity of pyrrole-3-carboxamide for all of the natural bases except C is substantially lower than that of imidazole-4-carboxamide. The highest T_m is 32.2°C for the duplex with **3**:T base pairs, which is ~13°C higher than that for the duplex with **1**:T base pairs, but 14.4°C lower than the T_m for the duplex containing **2**:T base pairs. The melting temperature of the duplex containing **3**:C base pairs is only 13.3°C, which is nearly the same as the T_m observed for the duplex containing **2**:C base pairs. On the other hand, the average value of ΔG for the four sequences containing **3** is only slightly higher than that for sequences containing **1**. The duplex with **1**:C base pairs has a substantially higher T_m (23.2°C) and appears to be the most stable of the 3-nitropyrrole base pairs. The ΔΔG for the duplexes containing pyrrole-3-carboxamides is 2.5 kcal/mol (1.3 kcal/mol per modified base pair). Because the ΔS value for the **3**:T base pair is significantly higher than that for any of the other base pairs, the **3**:T base pair must be held more rigidly within the duplex. One interpretation is that the more restricted motion is a result of increased hydrogen bonding. The hydrogen bonding pattern of the **3**:T base pair would be isomorphous with a natural A:T base pair as shown in Figure 1. Models show that the C1′–C1′ distance falls within 0.1 Å and λ_1 and λ_2 within 3° of the A:T base pair (Fig. 3).

Table 2. Melting temperature and thermodynamic parameters for helix–coil transition of the sequence d(CGC**X**AATT**Y**GCG) containing nucleoside **1** and/or **3**

X:Y	T_m (°C)	–ΔH (kcal/mol)	–ΔS (cal/kmol)	$-\Delta G_{25°C}$ (kcal/mol)	$\Delta\Delta G_{25°C}$ (kcal/mol)
A:T	64.4	58.1	150.0	13.4	
G:C	68.9	60.0	153.3	14.3	
1:A	17.8	25.7	66.4	6.0	0.4
1:C	23.2	34.5	94.4	6.4	
1:G	18.9	26.7	69.4	6.0	
1:T	19.3	26.7	69.4	6.1	
3:A	16.5	31.0	85.0	5.7	2.5
3:C	13.3	34.7	99.1	5.2	
3:G	21.6	27.1	69.8	6.2	
3:T	32.2	45.6	127.2	7.7	
1:**1**	30.9	34.0	89.6	7.2	
1:**3**	28.2	41.5	115.8	7.0	
3:**3**	4.4	29.1	82.7	4.4	

Absorbance versus temperature profiles of the sequences were determined at 260 nm. Measurements were made in 10 mM phosphate containing 1 M NaCl and 0.1 mM EDTA at pH 7.0 with an oligonucleotide concentration of ~15 μM.

Table 3. Total energies (Hartrees), relative energies (kcal/mol) and dipole moments (μ, D) for *syn*$_\tau$-**9** and *anti*$_\tau$-**9**

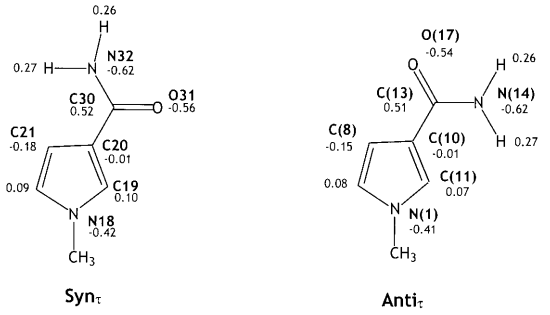
	6–31G(d,p), ε = 1			6–31G(d,p), ε = 40		
	$E_{SCF(rel)}$	E_{SCF}	Dipole moment	$E_{SCF(rel)}$	E_{SCF}	Dipole moment
Syn$_\tau$-**9**	0.00	–418.200 954 7	4.22	0.57	–418.204 319 8	5.20
Anti$_\tau$-**9**	1.06	–418.199 267 4	5.66	0.00	–418.205 224 7	6.98

We also explored how these nucleosides behave in a DNA duplex by placing them opposite one another. The results of melting experiments were very striking. The duplex with 3-nitropyrrole opposite 3-nitropyrrole (**1**:**1**) is more stable than the duplexes with 3-nitropyrrole opposite any of the natural bases. On the average, the T_m value for the **1**:**1** duplex is ~11°C higher than T_m value for the duplexes with 3-nitropyrrole opposite each of the four natural bases. Although these data support our original hypothesis that 3-nitropyrrole contributes to helix stability by stacking interactions, it also reveals the importance of a solvophobic effect. This effect was studied by Schweitzer and Kool for the hydrophobic, non-hydrogen-bonding base pairs difluorotoluene and trimethylbenzene (41). Each of these hydrophobic bases pairs with itself more effectively than with a natural base. As pointed out by these authors, a desolvation effect destabilizes base pairs involving one natural base and one hydrophobic base. The hydrogen bonding groups of a natural base are desolvated on insertion into the helix. If there is no concomitant compensation through hydrogen bonding to the base pair partner, then there is a net energy cost. For hydrophobic base pairs, there is little price to pay for desolvation and in fact a gain due to the solvophobic effect.

Interestingly, the T_m value for the duplex in which pyrrole-3-carboxamide was paired opposite itself was only 4.4°C. One might expect formation of a cyclic dimer between two amides. However, this cannot occur without significantly warping the double helix. To be isomorphous with A:T or G:C base pairs, the C1′–C1′ distance must be in the range of 10.8–11.0 Å, and λ_1 and λ_2 around 50° (42,43). As illustrated in Figure 3b, the paired amide configuration would require a significant increase in the C1′–C1′ distance between the two paired nucleosides. Space filling models arranged to maximize hydrogen bonding between two amide groups gives a C1′–C1′ distance 13.6 Å, λ_1 and λ_2 44° and 41°, respectively. Furthermore, both amide groups in this paired configuration are buried within the center of helix which would require significant desolvation. Models show that there are no pairing configurations in which the C1′–C1′ distance and λ_1 and λ_2 are optimized. When one amide group faces inward towards the helix axis while the other faces outward into the major groove (Fig. 3c), the C1′–C1′ distance is still too great (12.3 Å) and λ_1 and λ_2 are far too small (34° and 44°). In this **3***anti*$_\chi$:**3***syn*$_\chi$ arrangement, only one weak hydrogen bonding interaction would be possible, and desolvation of the carboxamide group would still occur. In the **3***syn*$_\chi$:**3***syn*$_\chi$ arrangement (Fig. 3d), because of the geometry it is difficult to achieve a hydrogen bonding interaction between the two amides. If the two pyrroles are brought within van der Waals contact, the C1′–C1′ distance is too short and λ_1 and λ_2 are too small. In contrast to the base pair **3**:**3**, the duplex with the **1**:**3** base pair has a T_m value which is almost identical to that for the **1**:**1** base pair. One interpretation is that the nitro group, unlike the amide group, is not highly solvated by water and hence can more favorably assume the conformation in which the nitro group faces inward towards the helix axis, then the amide could assume a conformation in which the carbonyl projects outward into the major groove in order to maintain amide solvent interactions.

Theoretical studies of 1-methyl pyrrole-3-carboxamide

In earlier studies, the energies of two conformations of methyl pyrrole-3-carboxamide (**9**) shown in Figure 4 were calculated by AM1 and found to differ by only 1.3 kcal/mol (18). This led us to predict that pyrrole-3-carboxamide would behave as a universal purine that pairs with C and T as illustrated in Figure 1. Since our melting experiments have shown that pyrrole-3-carboxamide pairs to C with significantly less affinity than to T, we have carried out more extensive theoretical calculations on the model compound, methyl pyrrole-3-carboxamide, in order to determine if the experimental results would match a more accurate method for predicting molecular properties. *Ab initio* calculations have been utilized successfully for theoretical elucidation of molecular structures and properties. However, these calculations often combine extensive electron correlation methods with an extended basis set for high-level computations. Possible alternatives to *ab initio* calculations are the DFT methods (44). Recent progress in the DFT method has made it more efficient than classical Hartree–Fock (HF) and Moller–Plesset perturbation (MPn) methods. One of the most extensively used DFT methods, B3LYP (45–48), was chosen for our studies. As illustrated in Figure 1, the amide group on the pyrrole ring would be expected to preferentially adopt two different conformations, *syn*$_\tau$ or *anti*$_\tau$, that should allow it to mimic either A or G. The results of the DFT calculation are summarized in Table 3 and Figure 4 for the optimized *syn* and *anti* conformations of **9**. These two conformations have almost identical charge distribution on the amide groups. In the gas phase (ε = 1), *syn*$_\tau$-**9** (A-mimic) is more stable than *anti*$_\tau$-**9** (G-mimic) by 1.06 kcal/mol. The stability of duplex DNA containing **9** depends on the stability of its base pairs in the interior of the duplex. Since the dipole moments of the two conformations are relatively large, but differ by 1.44 D, the dielectric constant of the surrounding medium could affect the relative stability of the two conformations and hence the relative stability of the base pairs with T and C. A self-consistent reaction field (SCRF) method was chosen to study the solvent medium effects (49,50), which has been applied to study the guanine–cytosine and isoguanine–isocytosine base pairs (51,52). The Onsager's reaction field model was used in the calculation (53). As shown in Table 3, proceeding from gas phase to a polar medium resulted in an increase in stability of the two conformations. The dielectric constant within the base pairing region of the helix is unknown. Typically ε values in the range 2–4 are used (54,55–58); more recently ε = 40 was assumed by Florian and Leszczynski and Roberts *et al.* (51,52). Experimental values of 20 and 51 have been determined for the minor and major grooves of dsDNA



Selected Torsion Angles in deg

N(18)-C(19)-C(20)-C(30)	-179.6	C(13)-C(10)-C(11)-N(1)	-180.0
C(21)-C(20)-C(30)-N(32)	6.0	C(8)-C(10)-C(13)-N(14)	-168.0
C(21)-C(20)-C(30)-O(31)	-172.7	C(8)-C(10)-C(13)-O(17)	14.0

Figure 4. Charge distributions in the two conformations of **9** and their torsion angles from the DFT calculation.

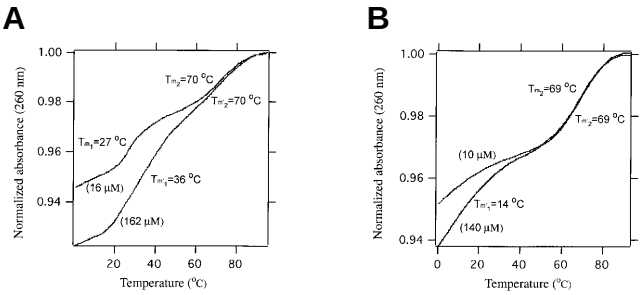


Figure 5. Hairpin formation in d(CGCGA**MM**TCGCG). Measurements were made in 10 mM phosphate containing 1 M NaCl and 0.1 mM EDTA at pH 7.0; absorbance versus temperature profiles of the sequences were determined at 260 nm. (**A**) **M** = **1**, at two different concentrations (16 and 162 μM); (**B**) **M** = **3**, at two different concentrations (10 and 140 μM).

respectively (59,60). Recent calculations using Poisson–Boltzmann methods agree with the experimental results and further show that dielectric constant falls rapidly on moving closer to the center of the duplex (61). Our results show that the dielectric constant does make a difference, but the difference is not large. By carrying out the calculation at both ε =1 and ε = 40 we are able to at least provide a plausible range for the energy differences in amide conformation. When ε was set equal to 40 the *anti*$_\tau$ conformation (G-mimic) is preferable to the *syn*$_\tau$ by 0.57 kcal/mol. Since the duplexes containing the **3**·C and **3**·T base pairs differ in energy by 11 kcal/mol, it appears that the conformational energy difference (1.06 kcal/mol or less between *syn*$_\tau$ and *anti*$_\tau$) is not a major contributor. Figure 4 shows the optimized structures for the *syn*$_\tau$ and *anti*$_\tau$ conformations of **9**. The plane of the amide deviates from the plane defined by the pyrrole ring in both conformations. For *syn*$_\tau$-**9**, the amide oxygen is out of the plane ~0.16 Å and the amide nitrogen ~0.14 Å. For *anti*$_\tau$-**9**, the amide oxygen is out of the plane ~0.30 Å and the amide nitrogen ~0.29 Å. This means that in an optimal conformation the pyrrole-3-carboxamide would deviate from planarity more significantly than either of the natural bases adenine or guanine. This twist may be a source of destabilization when the amide group is placed in the helix, but it is not necessarily the only explanation for the difference in stability between the **3**:T and **3**:C base pairs.

Hairpin structures from azole analog modified nucleosides

The sequences d(CGCGA**11**TCGCG) and d(CGCGA**33**TCGCG) were constructed to study the effect of tandem azole bases on oligonucleotide stability. However, it was found that the two sequences formed hairpin structures in buffer containing 1 M NaCl. As shown in Figure 5, the melting curves for these sequences have two transitions. The lower temperature transition, *transition 1*, is assigned to duplex dissociation with concomitant rearrangement to hairpin structure and the higher temperature transition, *transition 2*, is assigned to the melting of the hairpins. The corresponding T_m values are T_m1 and T_m2, respectively. T_m1 for *transition 1* increases with oligomer concentration, while T_m2 for *transition 2* is independent of oligomer concentration. Hairpin structures in the parent sequence d(CGCGAATTCGCG) consisting of a four base pair stem and a four base loop have only been observed in buffers in which the NaCl concentration was <10 mM. With **1** or **3** in the middle of the sequence, the hairpin structure is favored even at high sodium ion and oligomer concentration. The stabilities of the hairpin structures are similar for both sequences. However, the duplex containing two **3**:**3** base pairs (T_m = 14°C at 140 μM) was significantly less stable than the duplex containing two **1**:**1** base pairs (T_m = 36°C at 162 μM). This is in accord with our results for the **1**:**1** and **3**:**3** base pairs in the sequence d(CGC**X**AATT**Y**GCG) as discussed above. For self complementary sequences, duplex formation competes with hairpin formation in solution. It seems reasonable that any change in base structure which destabilizes base pairing in position 6 and 7 would have a more significant effect on duplex formation than on hairpin formation since the bases in the hairpin loop are not involved in base pairing.

CONCLUSION

There are a number of important results from this study. First, it is clear that replacing the nitro group of nitropyrrole with a carboxamide group reintroduces base pairing selectivity. Nucleoside **3** pairs to T better than it does to any of the other natural bases. However, it is also readily apparent that the T:**3** base pair is much less stable than a natural T:A base pair despite the fact that modeling suggests the possibility for two hydrogen bonds. There are at least three possible reasons for this difference in stability between AT and **3**:T base pairs. First, pyrrole-3-carboxamide would be expected to stack less effectively than adenine since it contains only a single five-membered ring. Second, the position of the amide group does not precisely match the position of N1 and N6 in adenine, which means that the **3**:T base pair is not strictly isomorphous with the A:T base pair (Fig. 3). Third, the rotational freedom about bond τ between C3 and the amide is lost on base pair formation, which is entropically unfavorable. Finally, we predicted that pyrrole-3-carboxamide should form equally stable base pairs with C and T. Yet, the thermal denaturation studies revealed that pyrrole-3-carboxamide formed the most stable base pair with T and the least stable one with C. One explanation is that the *anti*$_\tau$ conformation of pyrrole-3-carboxamide distorts base stacking in the DNA duplex. However, we believe that differences in the hydrogen bonding abilities of bases play the most significant role in base pairing stability. The common element in each of the base pairs in which the modified

azole is strongly paired to a natural base (T_m > 35°C) is the potential for formation of one hydrogen bond involving an acidic (p*K*as 9–10) base proton. It has been well documented that the more closely matched the pKas of the donor and the acceptor, the stronger a hydrogen bond. As Rich and coworkers pointed out 30 years ago (62), the acidity of the N-3 proton of 5-substituted uracil derivatives critically influences the stability of A-U base pairs. A difference of ~2 pKa units can lead to an order of magnitude difference in association constant. The pKa of an amide proton of **3** would be expected to be 6 pKa units less acidic than that of N3-H of T or N1-H of G. If the hydrogen bond between N1-H of G and the imidazole nitrogen of **2** or N3-H of T and the amide carbonyl of **2** and **3** is the major contributor to hydrogen bond mediated base pair stabilization, then the **3**:C base pair which lacks an acidic hydrogen may simply be less stabilized by hydrogen bonding interactions. Further discussion of this topic will follow in future publications.

ACKNOWLEDGEMENTS

We thank Dr Jian Li (The Scripps Research Institute) for his advice on theoretical calculations. The National Institutes of Health are gratefully acknowledged for support of this research.

REFERENCES

1 Bergstrom,D.E., Zhang,P., Toma,P.H., Andrews,C.A. and Nichols,R. (1995) *J. Am. Chem. Soc*., **117**, 1201–1209.
2 Loakes,D. and Brown,D.M. (1994) *Nucleic Acids Res*., **22**, 4039–4043.
3 Aerschot,A.V., Rozenski,J., Loakes,D., Pillet,N., Schepers,G. and Herdewijn,P. (1995) *Nucleic Acids Res*., **23**, 4363–1370.
4 Nichols,R., Andrews,P.C., Zhang,P. and Bergstrom,D.E. (1994) *Nature*, **369**, 492–493.
5 Loakes,D., Brown,D.M., Linde,S. and Hill,F. (1995) *Nucleic Acids Res*., **23**, 2361–2366.
6 Hoops,G.C., Zhang,P., Johnson,W.T., Paul,N., Bergstrom,D.E. and Davisson,V.J. (1997) *Nucleic Acids Res*., **25**, 4866–4871.
7 Bergstrom,D.E., Zhang,P. and Johnson,W.T. (1997) *Nucleic Acids Res*., **25**, 1935–1942.
8 Bergstrom,D.E., Zhang,P. and Zhou,J. (1994) *J. Chem. Soc. Perkin Transactions I*, 3029–3034.
9 Johnson,W.T., Zhang,P. and Bergstrom,D.E. (1997) *Nucleic Acids Res*., **25**, 559–567.
10 Ramasamy,K., Robins,R.K. and Revankar,G.R. (1986) *Tetrahedron*, **42**, 5869–5878.
11 Gralla,J. and Crothers,D.M. (1973) *J. Mol. Biol*., **78**, 301–319.
12 Law,S.M., Eritja,R., Goodman,M.F. and Breslauer,K.J. (1996) *Biochemistry*, **35**, 12329–12337.
13 Savitzky,A. and Golay,M.J.E. (1964) *Anal. Chem*., **36**, 1627.
14 Marky,L.A. and Breslauer,K.J. (1987) *Biopolymers*, **26**, 1601–1620.
15 Bevington,P.R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill Book Company, New York.
16 Frisch,M.J., Trucks,G.W., Schlegel,H.B., Gill,P.M.W., Johnson,B.G., Robb,M.A., Cheeseman,J.R., Keith,T., Petersson,G.A., Montgomery,J.A. *et al.* (1995) Gaussian 94, Revision D.2, Gaussian, Inc., Pittsburgh PA.
17 Frisch,M.J., Frisch,Æ. and Foresman,J.B. (1995) Gaussian 94 User's Reference.
18 Bergstrom,D.E., Zhang,P. and Johnson,W.T. (1996) *Nucleosides Nucleotides*, **15**, 59–68.
19 Loader,C.E. and Anderson,H.J. (1981) *Can. J. Chem*., **59**, 2673–2676.
20 Rapoport,H. and Willson,C.D. (1961) *J. Org. Chem*., **26**, 1102–1104.
21 Rokach,J., Hamel,P. and Kakushima,M. (1981) *Tetrahedron Lett.*, **22**, 4901–4904.
22 Cativiela,C. and Garcia,J.I. (1986) *Organic Preparations Procedures Int.*, **18**, 283–285.
23 Basha,A., Lipton,M. and Weinreb,S.M. (1977) *Tetrahedron Lett*., 4171–4174.
24 Wing,R., Drew,H., Takano,T., Broca,C., Itakura,K. and Dickerson,R.E. (1980) *Nature*, **287**, 755–757.
25 Dickerson,R.E. and Drew,H.R. (1981) *J. Mol. Biol*., **149**, 761–786.
26 Dickerson,R.E. (1990) In Sarma,R.H. and Sarma,M.H. (eds), *Structure and Methods, DNA and RNA Proceeding of the Sixth Conversation in Biomolecular Stereodynamics*. Adenine Press, Schenectady, NY, Vol. 3, pp. 1–38.
27 Bax,A. and Lerner,J. (1988) *J. Magn. Reson*., **79**, 429–438.
28 Nerdal,W., Hare,D.R. and Ried,B.R. (1989) *Biochemistry*, **28**, 10008–10021.
29 Lane,A., Jenkins,J.C., Brown,T. and Neidle,S. (1991) *Biochemistry*, **30**, 1372–1385.
30 Moe,J.G. and Russu,I.M. (1990) *Nucleic Acids Res*., **18**, 821–827.
31 Berveridge,D.L. and Ravishanker,G. (1994) *Curr. Opin. Struct. Biol*., **4**, 246–255.
32 Guarnieri,F. and Mezei,M. (1996) *J. Am. Chem. Soc*., **118**, 8493–8494.
33 Drew,H.R. and Dickerson,R.E. (1981) *J. Mol. Biol*., **151**, 535–556.
34 Kopka,M.L., Fratini,A.V., Drew,H.R. and Dickerson,R.E. (1983) *J. Mol. Biol*., **163**, 129–146.
35 Kubinec,M.G. and Wemmer,D.E. (1992) *J. Am. Chem. Soc*., **114**, 8739–8740.
36 Liepinsh,E., Otting,G. and Wuthrich,K. (1992) *Nucleic Acids Res*., **20**, 6549–6553.
37 Patel,D.J., Kozlowski,S.A., Marky,L.A., Broka,C., Rice,J.A., Itakura,K. and Breslauer,K.J. (1982) *Biochemistry*, **21**, 428–436.
38 Patel,D.J., Kozlowski,S.A., Marky,L.A., Rice,J.A., Broka,C., Itakura,K. and Breslauer,K.J. (1982) *Biochemistry*, **21**, 445–451.
39 Leonard,G.A., Booth,E.D. and Brown,T. (1990) *Nucleic Acids Res*., **18**, 5617–5623.
40 Marky,L.A., Blumenfeld,K.S., Kozlowski,S. and Breslauer,K.J. (1983) *Biopolymers*, **22**, 1247–1257.
41 Schweitzer,B.A. and Kool,E. (1995) *J. Am. Chem. Soc*., **117**, 1864–1872.
42 Kennard,O. (1987) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Vol. 1, pp. 25–52.
43 Hunter,W.N. (1992) In Lilley,D.M.J., and Dahlberg,J.E. (eds), *Methods in Enzymology: DNA Structures Part A: Synthesis and Physical Analysis of DNA*. Academic Press, New York, Vol. 211, pp. 221–231.
44 Parr,R.G. and Yang,W. (1989) *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York.
45 Becke,A.D. (1993) *J. Chem. Phys*, **98**, 5648–5652.
46 Becke,A.D. (1988) *Phys. Rev*., **A38**, 3098–3100.
47 Lee,C., Yang,W. and Parr,R.G. (1988) *Phys. Rev*., **B37**, 785–789.
48 Vosko,S.H., Wilk,L. and Nusair,M. (1980) *Can. J. Phys*., **58**, 1200.
49 Wong,M.W., Frisch,M.J. and Wiberg,K.B. (1991) *J. Am. Chem. Soc*., **113**, 4776–4782.
50 Tapia,O. and Goscinski,O. (1975) *Mol. Phys*., **29**, 1653–1661.
51 Roberts,C., Bandaru,R. and Switzer,C. (1997) *J. Am. Chem. Soc*., **119**, 4640–4649.
52 Florian,J. and Leszczynski,J. (1996) *J. Am. Chem. Soc*., **118**, 3010–3017.
53 Onsager,L. (1936) *J. Am. Chem. Soc*., **58**, 1486–1493.
54 Olson,W.K. (1982) In Neidle,S. (ed.), *Topics in Nucleic Acid Structure Part 2*. The Macmillan Press Ltd., London, pp. 12.
55 Hingerty,B.E., Ritchie,R.H., Ferrell,T.L. and Turner,J.E. (1985) *Biopolymers*, **24**, 427–439.
56 Sarai,A., Mazur,J., Nussinov,R. and Jernigan,R.L. (1988) *Biochemistry*, **27**, 8498–8502.
57 Friedman,R.A. and Honig,B. (1992) *Biopolymers*, **32**, 145–159.
58 Kollman,P.A., Weiner,P.K. and Dearing,A. (1981) *Biopolymers*, **20**, 2583–2621.
59 Jin,R. and Breslauer,K.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8939–8942.
60 Barawkar,D.A. and Ganesh,K.N. (1995) *Nucleic Acids Res*., **23**, 159–164.
61 Lamm,G. and Pack,G.R. (1997) *J. Phys. Chem*., **101**, 959–965.
62 Kyogoku,Y., Lord,R.C. and Rich,A. (1967) *Proc. Natl. Acad. Sci. USA*, **57**, 250–257.
63 Etter,M.C. (1990) *Acc. Chem. Res*., **23**, 120–126.
64 Etter,M.C., MacDonald,J.C. and Bernstein,J. (1990) *Acta Cryst*., **B46**, 256–262.
65 Bernstein,J., Davis,R.E., Shimoni,L. and Chang,N. (1995) *Angew. Chem. Int. Ed. Engl*., **34**, 1555–1573.
66 Shimoni,L. and Glusker,J.P. (1995) *Protein Sci.*, **4**, 65–74.