Synthesis, Structure, and Deoxyribonucleic Acid Sequencing with a Universal Nucleoside: 1-(2′-Deoxy-β-D-ribofuranosyl)-3-nitropyrrole

Donald E. Bergstrom,*,† Peiming Zhang,† Pascal H. Toma,† Philip C. Andrews,‡ and Ruthann Nichols‡

Contribution from the Department of Medicinal Chemistry and Pharmacognosy, School of Pharmacy and Pharmacal Sciences, Purdue University, West Lafayette, Indiana 47907-1333, Walther Cancer Institute, Indianapolis, Indiana 46208, and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109

Received July 5, 1994[⊗]

Abstract: A nucleoside analogue, 1-(2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole (**4**) was designed to function as a universal replacement for any of the natural nucleosides in DNA sequences. Compound **4** was synthesized by the reaction of 3-nitropyrrole with sodium hydride and 1-chloro-2-deoxy-3,5-di-*O*-toluoyl-D-erythropentofuranose, and the structure was confirmed by X-ray diffraction. Nucleoside **4** was transformed to 1-(2′-deoxy-5′-dimethoxytrityl-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite) (**6**) for incorporation into oligonucleotides by conventional synthesis protocols. Analogues of the oligonucleotide, 5′-d(CGT AAT CAG AAA ACA AT)-3′ with nucleoside **4** replacing the natural nucleosides in up to 9 positions were constructed and tested as primers for dideoxy sequencing. Sequencing studies show that a substantial number of nucleotides can be replaced by **4** without loss of primer specificity. Sequencing primer 4 with substitutions of **4** at the third position in each of four codons gave a sequencing ladder comparable to primer 1, the exact match, while a 256-fold degenerate oligonucleotide mixture (primer 2) gave an unreadable sequencing ladder. Primers containing two or more mismatches gave indecipherable results. A unique property of **4** was its ability to replace long strings of contiguous nucleosides and still yield functional sequencing primers. Sequences with three (primer 8), six (primer 9), and nine (primer 10) **4** substitutions all gave readable sequencing ladders. Optical thermal profiles obtained for the oligonucleotide pairs 5′-d($C_2T_5XT_5G_2$)-3′ and 5′-d($C_2A_5YA_5G_2$)-3′ (X, Y = A, C, G, T, and **4**) fit the normal sigmoidal pattern observed for the DNA duplex to single strand transition. The melting temperatures (T_m) of the oligonucleotides containing X-**4** base pairs (X = A, C, G or T, Y = **4**) all fell within a 3 °C range of one another. However, the T_m's were significantly lower than the corresponding sequences containing only A-T and G-C base pairs. The ability of **4** to associate by stacking with a natural nucleoside was confirmed by constructing the dimer d(Ap**4**) and determining the CD spectrum.

Introduction

Base analogs designed to function as a wild card in base pairing within nucleic acid duplexes have been widely sought. A "universal base" would be of significant use for the construction of nucleic acid probes and primers where the identity of the base in a target nucleic acid is unknown. In practice, this occurs frequently because of the degeneracy of the genetic code. Extrapolation of nucleic acid sequences from protein sequences is ambiguous because amino acids may be specified by up to six different codons. At the same time, if one is designing oligonucleotide probes based on limited peptide sequence information, it may be necessary to span regions where the identity of one or more entire codons is unknown. This situation arises frequently when modified amino acids are encountered during automated Edman degradation of proteins or when proteins are sequenced at subpicomole levels. It also arises when crossing species, in which case proteins may share relatively short conserved regions separated by nonconserved regions. An ideal universal nucleoside would be one that could not only replace bases at individual isolated sites but also replace entire codons as well.

So far, the search for nucleoside analogs designed to base-pair with more than one of the four primary DNA bases has had only marginal success.[1−9] The most extensively studied nucleoside having some characteristics of a universal nucleoside is 2′-deoxyinosine (dI).[10] Structural studies on dI modified oligonucleotides show that it can base-pair to dC,[11] dA,[12−14] T,[15,16] and dG.[17] But dI has a clear preference for pairing with dC; dI-T and dI-dG base pairs are as much as 2−3 kcal/mol

† Purdue University and Walther Cancer Institute.
‡ University of Michigan.
⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1995.

(1) Millican, T. A.; Mock, G. A.; Chauncey, M. A.; Patel, T. P.; Eaton, M. A. W.; Gunning, J.; Cutbush, S. D.; Neidle, S.; Mann, J. *Nuc. Acids Res.* **1984**, *12*, 7435−7453.
(2) Seela, F.; Kaiser, K. *Nuc. Acids Res.* **1986**, *14*, 1825−1844.
(3) Lin, P. K. T.; Brown, D. M. *Nuc. Acids Res.* **1989**, *17*, 10373−10383.
(4) Brown, D. M.; Lin, P. K. T. L. *Carbohydr. Res.* **1991**, *216*, 129−139.
(5) François, P.; Perilleux, D.; Kempener, Y.; Sonveaux, E. *Tetrahedron Lett.* **1990**, *31*, 6347−6350.
(6) Habener, J. F.; Vo, C., D.; Le, D. B.; Gryan, G. P.; Ercolani, L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1735−1739.
(7) Eritja, R.; Horowitz, D. M.; Walker, P. A.; Ziehler-Martin, J. P.; Boosalis, M. S.; Goodman, M. F.; Itakura, K.; Kaplan, B. E. *Nuc. Acids Res.* **1986**, *14*, 8135−8153.
(8) Fukuda, T.; Hamana, T.; Kikuchi, K.; Marumoto, R. Z. *Naturforsch.* **1986**, *41b*, 1571−1579.
(9) Inoue, H.; Imura, A.; Ohtsuka, E. *Nuc. Acids Res.* **1985**, *13*, 7119−7128.
(10) Ohtsuka, E.; Matsuki, S.; Ikehara, M.; Takahashi, Y.; Matsubara, K. *J. Biol. Chem.* **1985**, *260*, 2605−2608.
(11) Xuan, J.-C.; Weber, I. T. *Nuc. Acids Res.* **1992**, *20*, 5457−5464.
(12) Corfield, P. W. R.; Hunter, W. N.; Brown, T.; Robinson, P.; Kennard, O. *Nuc. Acids Res.* **1987**, *15*, 7935−7949.
(13) Leonard, G. A.; Booth, E. D.; Hunter, W. N.; Brown, T. *Nuc. Acids Res.* **1992**, *20*, 4753−4759.

0002-7863/95/1517-1201$09.00/0

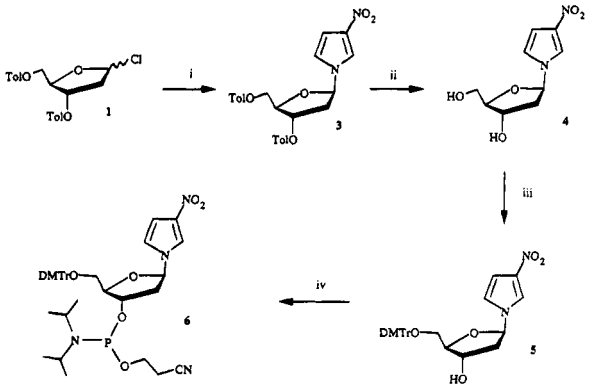
less stable than the dC-dI base pair.[18,19] More importantly, primers constructed with multiple sites of deoxyinosine substitution frequently give indecipherable results in sequencing experiments.

Our approach to the design of a universal nucleoside was to re-examine the common heterocyclic ring systems for fit into a DNA duplex opposite each of the four natural bases. The search was narrowed to five-membered ring systems when it became clear that larger rings could not be easily accommodated opposite the purines and still retain the appropriate complement of hydrogen bonds without significant distortion of the duplex. Five-membered ring heterocycles by themselves are probably too small to provide adequate stacking interactions. However, by an appropriate match of heterocyclic ring and ring substituents it appeared that it should be possible to construct a family of relatively small molecules that could provide substantial stacking energy, and in some instances both acceptor and donor sites for hydrogen bonding. The nucleoside analog described in this manuscript is one for which the potential for stacking was maximized with an anticipated sacrifice in hydrogen bonding potential. By enhancing intra- and interstrand stacking interactions, one could conceivably lessen the role hydrogen bonding plays in base pairing specificity and thereby decrease selectivity. Two functional groups that strongly polarize the electrostatic potential field of a π-aromatic system, and in doing would be expected to enhance stacking interactions, but which are not strong hydrogen bond acceptors, are nitro and cyano. Among candidate molecules, 3-nitropyrrole was considered the most attractive because of its structural and electronic resemblance to *p*-nitroaniline, derivatives of which are among the smallest known dsDNA intercalators.[20,21] The synthesis of 3-nitropyrrole deoxyribonucleoside (**4**), confirmation of its structure by X-ray crystallography, and its function as a universal nucleoside in oligonucleotides are described below. In a preliminary communication we reported the results of initial synthetic, structural, and biochemical studies with nucleoside **4** as a constituent of oligonucleotides.[22,23] In this manuscript we present structural, synthetic, and more complete data on sequencing experiments with nucleoside **4** modified oligonucleotides.

Result and Discussion

Synthesis. 1-(2′-Deoxy-β-D-ribofuranosyl)-3-nitropyrrole (**4**) was synthesized as outlined in Scheme 1 by direct attachment of the protected furanose **1** to 3-nitropyrrole (**2**). The method employed was virtually identical to the procedure previously reported by Robins and co-workers for the synthesis of pyrrole-containing nucleosides.[24,25] 3-Nitropyrrole was prepared in two steps from commercially available 1-(triisopropylsilyl)pyrrole

Scheme 1



[a] Reagents: (i) 3-nitropyrrole (**2**), NaH; (ii) ammonia, MeOH; (iii) dimethoxytrityl chloride, pyridine; (iv) 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite.

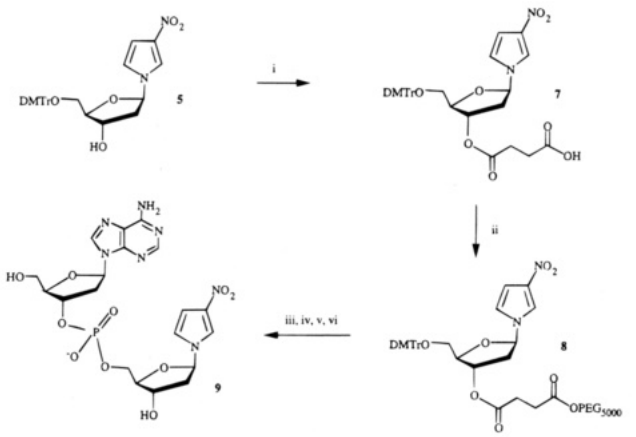
as described by Bray *et al.*[26] The sodium salt of 3-nitropyrrole, generated *in situ* by using NaH in acetonitrile, reacted rapidly with 1-chloro-2-deoxy-3,5-di-*O*-toluoyl-D-erythropentofuranose at ambient temperature. Only one product was obtained (yield: 88%). The toluoyl groups of **3** were removed by ammonia in methanol at 55 °C to yield nucleoside **4**. The configuration of **4** was assigned as β by ^{1}H NMR on the basis of a characteristic triplet for the 1′-proton observed at 5.97 ppm ($J = 6.6$ Hz) and the narrow multiplet (0.1 ppm width) observed for the two 2′-protons.[27,28] For incorporation into oligonucleotide, nucleoside **4** was converted to a nucleoside phosphoramidite. The 5′-hydroxyl was protected by the reaction of **4** with dimethoxytrityl chloride in pyridine to give nucleoside **5**. Phosphitylation of **5** gave 1-(2′-deoxy-5′-dimethoxytrityl-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-(2-cyanoethyl-*N,N*-diisopropylphosphoramidite) (**6**) in 79% yield after separation by chromatography on silica gel. Phosphoramidite **6** was used to construct modified oligonucleotides by conventional protocols in an automated DNA synthesizer. Stepwise yields, evaluated by spectrophotometric monitoring of the dimethoxytrityl release at each cycle of synthesis, were no different from those obtained using phosphoramidites of conventional bases.

In order to determine if nitropyrrole can participate in base stacking with a natural nucleic acid base the dimer, [5′-(3′-adenosylyl)-2′-deoxy-β-D-ribofuranosyl]-3-nitropyrrole d(Ap**4**), was prepared (Scheme 2) and its CD and NMR spectra determined. To obtain d(Ap**4**), 1-(5′-*O*-DMTr-2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-succinate was synthesized by the procedure of Pon et al.[29] and linked to PEG_{5000} by the procedure of Bonora et al.[30] Reaction of 5′-dimethoxytrityl-N^6-benzoyldeoxyadenosine-3′-(2-cyanoethyl-*N,N*-diisopropyl)-phosphoramidite with the PEG_{5000} linked nitropyrrole nucleoside, followed by deprotection and cleavage from the PEG_{5000}, gave [5′-(3′-adenosylyl)-2′-deoxy-β-D-ribofuranosyl]-3-nitropyrrole.

Sequencing Studies. A set of oligonucleotides containing from one to nine copies of nucleoside **4** were designed to test

(14) Uesugi, S.; Oda, Y.; Ikehara, M.; Kawase, Y.; Ohtsuka, E. *J. Biol. Chem.* **1987**, *262*, 6965–6968.
(15) Cruse, W. B. T.; Aymani, J.; Kennard, O.; Brown, T.; Jack, A. G. C.; Leonard, G. A. *Nuc. Acids Res.* **1989**, *17*, 55–72.
(16) Carbonnaux, C.; Fazakerley, G. V.; Sowers, L. C. *Nuc. Acids Res.* **1990**, *18*, 4075–4081.
(17) Oda, Y.; Uesugi, S.; Ikehara, M.; Kawase, Y.; Ohtsuka, E. *Nuc. Acids Res.* **1991**, *19*, 5263–5267.
(18) Kawase, Y.; Iwani, S.; Inoue, H.; Miura, K.; Ohtsuka, E. *Nuc. Acids Res.* **1986**, *14*, 7727–7736.
(19) Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Tinoco, I., Jr. *Nuc. Acids Res.* **1985**, *13*, 8927–8938.
(20) Gabbay, E. J. *Bioorg. Chem.* **1977**, *3*, 33–70.
(21) Gabbay, E. J. *J. Am. Chem. Soc.* **1969**, *91*, 5136–5150.
(22) Bergstrom, D. E.; Zhang, P.; Toma, P. H.; Andrews, P. C.; Nichols, R. Presented at the 206th ACS National Meeting of the American Chemical Society, Chicago, IL; American Chemical Society: Washington, DC, 1993; pp ORGN 308.
(23) Nichols, R.; Andrews, P. C.; Zhang, P.; Bergstrom, D. E. *Nature* **1994**, *369*, 492–493.
(24) Ramasamy, K.; Robins, R. K.; Revankar, G. R. *Tetrahedron* **1986**, *42*, 5869–5878.
(25) Ramasamy, K.; Robins, R. K.; Revankar, G. R. *Nucleosides and Nucleotides* **1988**, *7*, 385–392.
(26) Bray, B. L.; Mathies, P. H.; Naef, R.; Solas, D. R.; Tidwell, T. T.; Artis, D. R.; Muchowski, J. M. *J. Org. Chem.* **1990**, *55*, 6317–6328.
(27) Srivastava, P. C.; Robins, R. K. *J. Hetero. Chem.* **1981**, *18*, 1659–1662.
(28) Robins, M. J.; Robins, R. K. *J. Am. Chem. Soc.* **1965**, *87*, 4934–4940.
(29) Pon, R. T.; Usman, N.; Ogilvie, K. K. *BioTechniques* **1988**, *6*, 768–775.
(30) Bonora, G. M.; Biancotto, G.; Maffini, M.; Scremin, C. L. *Nuc. Acids Res.* **1993**, *21*, 1213–1217.

Scheme 2



[a] Reagents: (i) succinic anhydride, DMAP, pyridine; (ii) PEG_{5000}, dichloroethane, DCC; (iii) 6% TCA, dichloroethane; (iv) dA-CE phosphoramidite, tetrazole; (v) I_2, THF/H_2O/pyridine; (vi) 30% ammonium hydroxide.

Table 1. Modified Primers for Sanger Sequencing[a]

Primer No.	Sequence
1	d(CGT AAT CAG AAA ACA AT)
2	d(CGT AA*N* CA*N* AA*N* AC*N* AT)
3	d(CGT AAI CAI AAI ACI AT)
4	d(CGT AA**4** CA**4** AA**4** AC**4** AT)
5	d(CGT AAT CAG AAA ACA A**4**)
6	d(CGT AAT CAG AAA ACA **4**T)
7	d(CGT AAT CAG AAA AC**4** AT)
8	d(CGT AAT CAG AAA **444** AT)
9	d(CGT AAT CAG **444 444** AT)
10	d(CGT AAT **444 444 444** AT)
11	d(CGT AAT CAG AA<u>C</u> ACA AT)
12	d(CGT AAT CAG AAA AC<u>G</u> AT)
13	d(CGT AAT CAG AA<u>C</u> AC<u>G</u> AT)
14	d(CGT AAT CAG AAA ACA A<u>C</u>)
15	d(CGT AA<u>C</u> CA<u>A</u> AA<u>C</u> AC<u>G</u> AT)

[a] I = deoxyinosine, **4** = 3-nitropyrrole deoxyribonucleoside, N = mixture of A, C, G, and T. All underlined bases are mismatches to target sequence. Primer 2 consists of a 256-degenerate mixture.

the function of this analog as potential universal nucleoside in dideoxynucleotide mediated sequencing.[31] For this initial study, the set of modified primers was based only on the sequence 5′-d(CGT AAT CAG AAA ACA AT)-3′ (Table 1). This primer is complementary to a segment of the *Drosophila* drosulfakinin (Dsk) gene that encodes the neural peptide DSK-1, which is produced in only three cells (two MP-1 and the SP-1 cells) in the central nervous system.[32] The primers for testing dideoxy sequencing included oligonucleotides with up to four degenerate codon positions replaced by **4**, and sequences in which the bases of one to three entire codons were replaced by **4** (Table 1).

A comparison of sequencing ladders for oligonucleotide primers 1−4 has been published.[23] In this previous manuscript it was reported that the synthetic oligonucleotide, 5′-d(CGT AA**4** CA**4** AA**4** AC**4** AT)-3′ (primer 4), containing nucleoside **4** in the third position of four codons gave an unambiguous sequencing ladder that could be as easily read as the sequencing ladder for the control sequence 5′-d(CGT AAT CAG AAA ACA AT)-3′ (primer 1). In contrast, a 256-fold degenerate mixture of primers (primer 2), containing all possible combinations of bases at positions 6 (relative to 5′-end of primer), 9, 12, and 15 gave an ambiguous and unreadable sequencing ladder. This is a

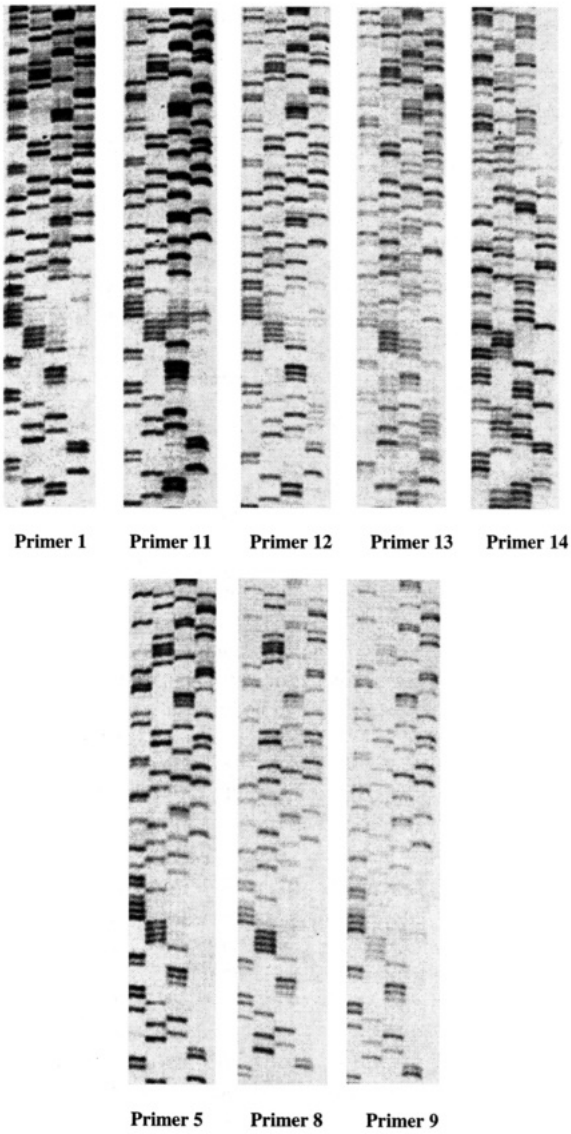


Figure 1. Dideoxy sequencing data using primers 1 and 11−14 (A, top) and primers 5, 8, and 9 (B, bottom). The primer sequences are listed in Table 1. The oligonucleotides were purified and sequencing reactions performed as described in the Experimental Section.

particularly signifcant result because the use of degenerate mixtures of primers is the usual method of choice when targeting nucleic acids for which the exact sequence is uncertain. Melting temperature studies discussed below, demonstrate that 3-nitropyrrole does not bind with high affinity to any of the natural nucleic acid bases in duplex nucleic acids, but instead appears to exhibit unusual neutrality, with very little difference in the T_m's of duplexes containing each of the four natural bases opposite 3-nitropyrrole. Primer 3, which contains a deoxyinosine at the same four locations (5′-CGT AAI CAI AAI ACI AT-3′) gave a less clear sequencing ladder than the corresponding primer containing nucleoside **4** (primer 4). It is instructive to elaborate on these results by comparing primers that contain base mismatches. Two primers (primers 11 and 12) with a single mismatch in the middle of the sequence in which either a dG or dC replaced a dA gave sequencing ladders (Figure 1A) which for the most part matched that for primer 1. However, in both there were extra bands and as a result the sequencing ladders were not completely unambiguous. With two mismatches (primer 13) the sequencing ladder deviated significantly from that of the control primer (primer 1). Primer 15 with four mismatches gave a completely unreadable sequencing ladder

(31) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5463−5467.

(32) Nichols, R. *Mol. Cellular Neurosci.* **1992**, *3*, 342−347.

Table 2. T_m Data for Hybridization of the Sequence 5′-d($C_2T_5XT_5G_2$)-3′ with 5′-d($C_2A_5YA_5G_2$)-3′ [a]

base-pair (X-Y)	T_m (°C)
C-G	59
C-A	43
C-T	49
A-G	50
A-C	45
A-T	57
4-G	43
4-C	45
4-A	46
4-T	45

[a] The melting temperatures of the complementary oligonucleotide pairs were determined in 1.0 M NaCl, 0.1 mM EDTA, 10 mM sodium phosphate, pH 7.0 on a Jasco model 710 CD spectrometer equipped with a Peltier device for heating and cooling. The oligonucleotide concentration was 50 μM.

(results not shown). A primer containing a single C−C mismatch at the 3′ terminus (primer 14) also did not give a readable sequence (Figure 1A). In contrast a sequence in which nucleoside **4** was located at the 3′ terminus of the primer (primer 5) did function as an efficient sequencing primer (Figure 1B). An unusual property of nucleoside **4** was its ability to replace long strings of contiguous nucleosides and still give perfectly functional primers. The primers 5′-d(CGT AAT CAG AAA **444** AT)-3′ (primer 8), 5′-d(CGT AAT CAG **444 444** AT)-3′ (primer 9) (Figure 1B), and 5′-d(CGT AAT **444 444 444** AT)-3′ (primer 10) (not shown in figures) gave normal sequencing ladders.

Melting Temperature Studies. In order to determine how discriminatory 3-nitropyrrole was in base-pairing to A, C, G, and T it was necessary to carry out DNA duplex melting studies. T_m data for hybridization of the sequences 5′-d($C_2T_5XT_5G_2$)-3′ and 5′-d($G_2A_5YA_5C_2$)-3′ were obtained for 3-nitropyrrole deoxyribonucleoside (Table 2). These particular sequences were chosen to match as closely as possible the sequences used by Kawase et al.[18] [d($G_2A_4XA_4G_2$) hybridized to d($C_2T_4YT_4C_2$)] and Martin et al.[19] [d(CA_3XA_3G) hybridized to d(CT_3YT_3G)], but at the same time provide a slightly higher T_m range. Oligonucleotide sequences containing long segments of A tend to have large propeller twists in which the N-6 amino group forms a three center hydrogen bond with O-4 of the opposing T as well as the T in an adjacent base pair.[33] The unusual nature of this structure suggests that the validity of using polyA containing sequences for thermal denaturation studies may be questionable. That does not appear to be the case; as noted below sequences containing a significantly greater proportion of G-C base pairs in the vicinity of the modified base gave similar results.

The helix-coil transition temperatures (T_m) of duplexes composed of the sequences 5′-d(CCT TTT T**4**T TTT TGG)-3′ and 5′-d(CCA AAA AXA AAA AGG)-3′ (X = A, C, G, or T) melted 11−14 °C lower than the corresponding sequences containing an A-T base pair (T_m = 57 °C). The T_m's, which varied from 43 to 46 °C, fell within the range observed for base mismatches. On the basis of this data one might conclude that nucleoside **4** greatly destabilizes DNA duplexes. Examination of published data on other putative universal nucleosides provides a somewhat different perspective. In one of the early attempts to find a universal base Millican et al. constructed a series of oligonucleotides, 5′-CTCAXCGTACTGGTT-3′, in which X was either a natural bases or one of the modified nucleosides 1,2-dideoxyribose or 1,2-dideoxy-1-phenyl-β-D-ribofuranose.[1] The T_m's of duplex formation with the complementary sequences 5′-AACCAGTACGNTGAG-3′ were determined. The two duplexes containing natural base pairs (N = A; X = T and N = G; X = C) had T_m's, respectively, of 45 and 42.5°. In comparison when X was 1,2-dideoxyribose and N each of the four natural nucleosides, T_m's ranged from 26−30 °C (27.5 °C average). Addition of a phenyl group at C-1 did not significantly alter the T_m's (28.9 °C average). The same sequences with mismatches at the site of modification gave intermediate T_m's (A-C, 36 °C; G-T, 38.5 °C). This is consistent with the data published by Tibanyenda et al. which established that mismatched base pairs provide significant stabilization of duplex structure through base stacking interactions.[34] Since stacking interactions are not possible for 1,2-dideoxyribose and unlikely for a nonpolar benzene ring,[35] it follows that the significantly lower T_m's found for these analogs was reasonable. Since nucleoside **4** yielded T_m's similar to those of mismatches, it must be similar to the natural bases in its ability to base stack. Most importantly, nucleoside **4** shows relatively little discrimination for base pairing to A, C, G, and T (T_m range = 3 °C), exactly what one would expect for an ideal universal base.

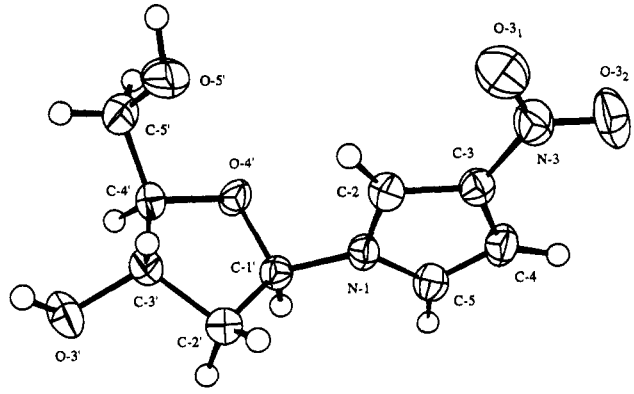


Figure 2. ORTEP II plot of the X-ray structure of 1-(2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole (**4**).

The effect of nucleoside **4** on duplex melting is not substantially altered by sequence context. The lack of discrimination of nucleoside **4** in base pairing to each of the natural bases is retained in substantially different sequences in which the modified nucleoside is located between two deoxyguanosines.[36] The T_m's of the duplexes formed from the sequences 5′-d(ACTTGGCCNCCATTTTG)-3′ and 5′-d(CAAAATGG**4**-GGCCAAGT)-3′ were lowered by 8 (**4**-dA), 11(**4**-dT), 10(**4**-dC), and 9 °C (**4**-dG).

Structure. Since nucleoside **4** was designed to function as a spacer in nucleic acid sequences, it was important to examine the structure of the molecule in detail. The crystal and molecular structures of nucleoside **4** were determined by X-ray diffraction. An ORTEPII[37] representation of nucleoside **4** is shown in Figure 2 and a packing diagram shown in Figure 3. As anticipated from the NMR data the configuration at C-1′ is β.

The bond lengths and angles are given in Tables 3 and 4. All are in the expected range. The bonds C-4′−O-4′ and O-4′−C-1′ have the same length, but the former is slightly longer than in normal nucleosides. The pyrrole ring is planar and the maximum deviation of a ring atom from the least-square plane

(33) Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, *330*, 221−226.

(34) Tibanyenda, N.; De Bruin, S., H.; Haasnoot, C., A. G.; Van der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. *Eur. J. Biochem.* **1984**, *139*, 19−27.

(35) Bugg, C. E.; Thomas, J. M.; Sundaralingam, M.; Rao, S. T. *Biopolymers* **1971**, *10*, 175−219.

(36) Loakes, D.; Brown, D. M. *Nuc. Acids Res.* **1994**, 22, 4039−4043.

(37) Johnson, C. K. *ORTEPII*; Oak Ridge National Laboratory: Oak Ridge, TN, 1976.

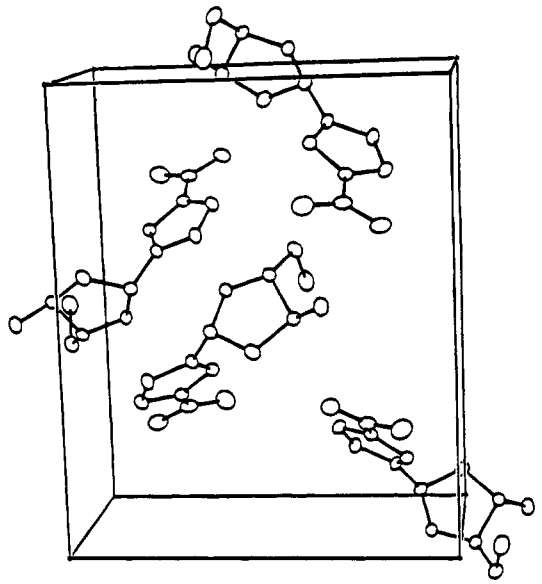


Figure 3. Packing diagram for 1-(2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole (**4**).

Table 3. Bond Distances in Å[a]

atom 1	atom 2	distance	atom 1	atom 2	distance
O-4′	C-4′	1.423(4)	N-1	C-5	1.380(4)
O-4′	C-1′	1.423(4)	N-3	C-3	1.425(5)
O-3′	C-3′	1.420(4)	C-4′	C-3′	1.533(5)
O-5′	C-5′	1.438(5)	C-4′	C-5′	1.492(5)
$O\text{-}3_2$	N-3	1.231(4)	C-3′	C-2′	1.519(5)
$O\text{-}3_1$	N-3	1.235(5)	C-2′	C-1′	1.522(5)
N-1	C-1′	1.466(4)	C-2	C-3	1.383(5)
N-1	C-2	1.394(5)	C-3	C-5	1.342(6)
			C-4	C-5	1.342(6)

[a] Numbers in parentheses are estimated standard deviations in the least significant digits.

Table 4. Bond Angles in deg[a]

atom 1	atom 2	atom 3	angle	atom 1	atom 2	atom 3	angle
C-4′	O-4′	C-1′	110.3(3)	O-3′	C-3′	C-2′	112.7(3)
C-4′	C-3′	C-2′	103.6(3)	C-3′	C-2′	C-1′	106.1(3)
C-1′	N-1	C-2	125.6(3)	O-4′	C-1′	N-1	107.0(3)
C-1′	N-1	C-5	123.8(3)	O-4′	C-1′	C-2′	107.3(3)
C-2	N-1	C-5	110.3(3)	N-1	C-1′	C-2′	115.2(3)
$O\text{-}3_2$	N-3	$O\text{-}3_1$	123.6(4)	N-1	C-2	C-3	105.3(3)
$O\text{-}3_2$	N-3	C-3	117.5(4)	N-3	C-3	C-2	123.0(3)
$O\text{-}3_1$	N-3	C-3	118.9(3)	N-3	C-3	C-4	127.0(4)
O-4′	C-4′	C-3′	106.5(3)	C-2	C-3	C-4	109.7(3)
O-4′	C-4′	C-5′	109.5(3)	C-3	C-4	C-5	105.9(3)
C-3′	C-4′	C-5′	116.4(3)	N-1	C-5	C-4	108.8(3)
O-3′	C-3′	C-4′	110.2(3)	O-5′	C-5′	C-4′	109.0(3)

[a] Numbers in parentheses are estimated standard deviations in the least significant digits.

is 0.0054 Å. The C−N bond length (1.425 Å) is typical of a conjugated nitro group, but it is bent out of the pyrrole plane by about 7.5° according to the calculation of its dihedral angle with the pyrrole plane. The deviation is probably a result of the packing forces in the crystal.

Torsion angles related to the molecular conformation of the deoxyribose are listed in Table 5. The sugar ring is puckered with C-4′ *endo* to the pyrrole ring. The displacement of C-4′ is 0.37 Å from the plane defined by O-4′, C-1′, C-2′, and C-3′. The pseudorotation parameter is **P** = 232.9° and the puckering amplitude is τ_m = −24.4°. The most common conformations of nucleosides in the solid state and solution are C-3′-*endo* (**N**) (**P** = 0−36°) and C-2′-*endo* (**S**) (**P** = 144−180°). Comparison of the values of **P** determined for other 2′-deoxyribosides show

Table 5. Selected Torsion Angles in deg[a]

atom 1	atom 2	atom 3	atom 4	angle
C-4′	O-4′	C-1′	C-2′	14.71(0.33)
O-4′	C-1′	C-2′	C-3′	1.52(0.34)
C-1′	C-2′	C-3′	C-4′	−15.46(0.33)
C-2′	C-3′	C-4′	O-4′	24.42(0.32)
C-3′	C-4′	O-4′	C-1′	−24.88(0.32)
O-4′	C-1′	N-1	C-2	45.84(0.39)
O-5′	C-5′	C-4′	O-4′	66.17(0.34)
O-5′	C-5′	C-4′	C-3′	−54.59(0.38)
C-5′	C-4′	C-3′	O-3′	92.48(0.34)

[a] Numbers in parentheses are estimated standard deviations in the least significant digits.

that most of them fall within the range 144−180° (C-2′-*endo*).[38] In A-DNA the furanose pucker is C-3′-*endo* (**N**), while in B-DNA it is C-2′-*endo*.

Other conformations including C-1′-*exo*, C-4′-*exo*, and O-4′-*endo* occur in crystal structures.[39] However the C-4′-*endo* puckering that we observe for nucleoside **4** is very unusual and does not appear to have been previously reported in the literature for simple deoxyribonucleosides.[40] There are a number of possible factors that may influence the ability of this nucleoside to assume the C-4′-*endo* conformation. The relatively small dimensions of the pyrrole ring allow the C-5′−O-5′ bond to lie over the sugar ring without steric conflict. The CH_2OH chain adopts the gauche$^+$ conformation with ϕ_{oo} = −66.17° and ϕ_{oc} = 54.59°. In addition, the data implicate the existence of a hydrogen bond between O-5′ and O-4′ (the distance between O-5′ and O-4′ is 2.846 Å). The orientation of the heterocyclic base about the glycosidic bond in nucleosides is an important conformation parameter. We define the conformation of nucleoside **4** about the glycosidic bond by torsion angle χ_{CN} = O-4′−C-1′−N-1−C-2. The corresponding value χ_{CN} = −45.85° implies that nucleoside **4** exists in the *syn* conformation with the nitro group located over the sugar ring, which may also be a factor in the stabilization of the observed conformation. The distance between O-5′ and O-3′ is 2.895 Å which implicates hydrogen bonding between them. The bonds C-4′−C-5′ and C-3′−O-3′ are almost orthogonal with the torsion angle O-3′−C-3′−C-4′−C-5′ = 92.48°, which is helpful for the formation of hydrogen bonding between O-5′ and O-3′.

A charge distribution contour map of 3-nitropyrrole, generated in the software program SYBYL using the Del Re[41] and Hückel[42] methods to obtain partitioned σ- and π-charge contributions, showed an electronic distribution that resembled the average charge distribution of the DNA bases. The dipole moment of 3-nitropyrrole deoxyribonucleoside should be close to the value reported for 1-methyl-3-nitropyrrole (6.15 μ).[43] If in duplex DNA the nitro group were oriented into the helix, the direction of the dipole would be midway between that of adenine and guanine. If the nitro group were oriented out of the helix (into the major groove), the dipole moment would be in a similar direction to that of guanine. To the extent that dipole-induced forces play a role in base stacking interactions, 3-nitropyrrole should be particularly effective in this capacity. The sequence 5′-CGT AAT **444 444 444** AT-3′, which

(38) Saenger, W. In *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; pp 55−61.
(39) Birnbaum, G. I. In *Nucleic Acid Structure, Part 3*; Neidle, S., Ed.; VCH: New York, 1987; p 5.
(40) *Programming and Documentation*; Cambridge Crystallographic Data Centre: Cambridge, U.K., 1992.
(41) Del Re, G. *J. Chem. Soc.* **1958**, 4031−4040.
(42) Purcell, W. P.; Singer, J. A. *J. Chem. Eng. Data* **1967**, *12*, 235−246.
(43) Lumbroso, H.; Carpanelli, C. *Bull. Soc. Chim. Fr.* **1964**, 3198−3203.

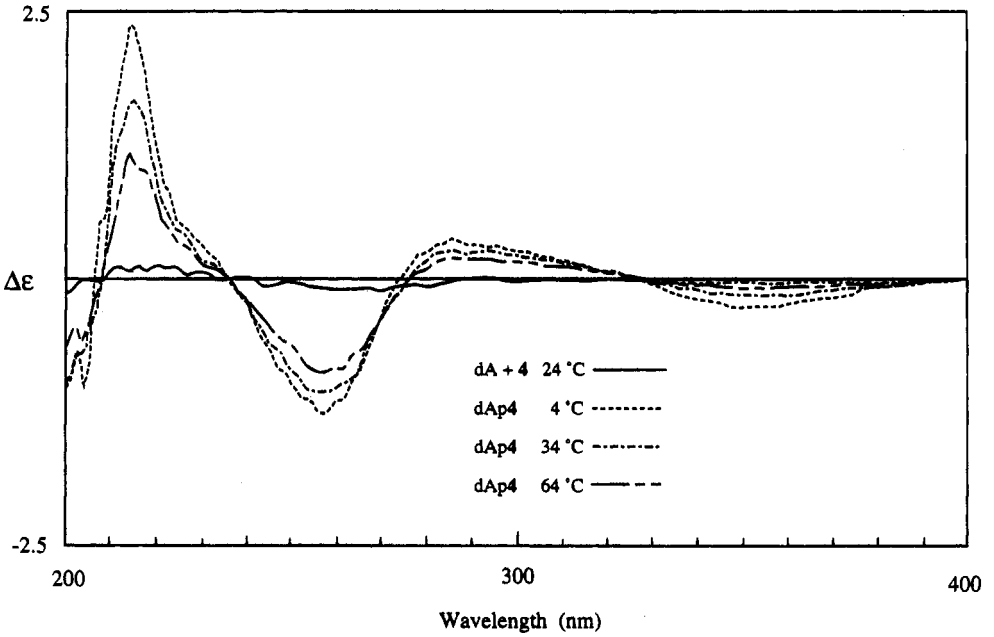


Figure 4. CD spectrum of Ap4 in 10 mM sodium phosphate buffer, pH 7.1.

functioned as a primer for T7 DNA polymerase, contains a string of 3-nitropyrroles extending nearly one full turn of a double helix. Since it is unlikely that the two bases at the 3′-end are by themselves sufficient to maintain normal duplex structure at 37 °C, stacking interactions between adjacent 3-nitropyrrole molecules may be uniquely effective in contributing to the maintenance of duplex-like geometry. As a model experiment to confirm that 3-nitropyrrole does stack with nucleic acid bases we constructed the dinucleoside analog d(Ap**4**) by reaction of the 5′-DMT protected 3′-phosphoramidite of dA with 3-nitropyrrole deoxyribonucleoside. The CD spectrum of this dimer in aqueous buffer was measured at a range of temperatures from 5 °C up to 65 °C (Figure 4). The changes that occur in the CD spectrum with change in temperature are qualitatively similar to those observed for d(ApA).[44] Interpretation of these results along with a more detailed structural analysis by NMR spectroscopy will be the subject of a later study.

As a prelude to more comprehensive structural studies we have modeled a B-form DNA duplex corresponding to the sequence 5′-d(CGCGAATTCGCG)-3′ in which nitropyrrole was placed opposite each of the natural bases. This sequence was chosen for modeling because of the availability of atomic coordinates.[45] The four modified sequences 5′-d(CGCG-**4**ATTCGCG)-3′, 5′-d(CGCGAAT**4**CGCG)-3′, 5′-d(CGC**4**AATT-CGCG)-3′, and 5′-d(CGCGAATT**4**GCG)-3′ were each placed opposite the unmodified sequence. Structures were generated on a Silicon Graphics IRIS 4D/120GTX using the program QUANTA. The lowest energy conformation was calculated using the Newton-Raphson minimization equation performed in CHARMM.[46] The base-pair parameters that are most indicative of helix deformation are (1) the distance C-1′ to C-1′ and (2) the angle λ, defined by the C-1′ to C-1′ vector and the C-1′ to N-1 (pyrimidine) or C-1′ to N-9 (purine) vector (Figure 5).[47] These parameters fall within the same relatively narrow range for both dA-T and dC-dG base pairs. For example, the X-ray structure of the Dickerson sequence, 5′-d(CGCGAAT-TCGCG)-3′ gave C-1′ to C-1′ interglycosyl distances varying from 10.23−10.46 Å for the four dA-T base pairs and 10.35−10.81 Å for the six interior dC-dG base pairs (average distance, 10.52 Å).[48] The values of the angle λ range from 53° to 60° (average angle, 54.4°). For any given base pair the values of λ_1 and λ_2 typically vary no more than 4° from each other. This result is indicative of the high degree of symmetry and the isostructural nature of dA-T and dC-dG base pairs. At least one of these two parameters (λ or C-1′−C-1′ distance) is significantly altered in mismatched base pairs. For example, in the dG-T base pair the C-1′ to C-1′ interglycosyl distance is well within normal range at 10.3 Å, but the two angles λ (dG) and λ(T) and are significantly different with values of 42° and 69°. Or alternatively, in a dA-dG mismatch (*anti-anti*) the values of λ are normal (52° and 53°), but the C-1′ to C-1′ interglycosyl distance (12.5 Å) is much longer. When dA assumes a *syn* conformation on pairing with dG, the C-1′−C-1′ distance assumes a normal value (10.7 Å), but the symmetry of the base pair is altered (λ = 58° and 40°).

The same parameters were measured in the modeling studies with the 3-nitropyrrole containing sequences to determine how well this modified base fits opposite each of the four natural nucleic acid bases. Opposite dA and dG, the best parameters are obtained when nucleoside **4** is in a *syn* conformation, while opposite dC and T, **4** fits best in an *anti* conformation. The C-1′ to C-1′ interglycosyl distances and the values of angles λ_1 (C1′, C1′, N-1(**4**)) and λ_2 measured on the model for the **4**-dA, **4**-dC, **4**-dG, and **4**-T base pairs were, respectively, 10.88 (53.8°, 53.7°), 10.94 (48.1°, 53.1°), 10.96 (54.1°, 52.2°), and 10.96 Å (49.1°, 53.5°). The C-1′ to C-1′ interglycosyl distances for base pairs containing nucleoside **4** are on the average, 0.4 Å longer than for the normal base pairs. The average value of λ (52.2°) is only slightly less than that observed for the natural base pairs. More importantly, base pair symmetry is retained.

(44) Kang, H.; Chou, P.-J.; Johnson, W. C.; Weller, D.; Huang, S.-B.; Summerton, J. E. *Biopolymers* **1992**, *32*, 1351−1363.
(45) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Aca. Sci. U.S.A.* **1981**, *78*, 2179−2183.
(46) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(47) Kennard, O. In *Nucleic Acids and Molecular Biology*; Eckstein, F., Lilley, D. M. J., Eds.; Springer-Verlag: Berlin, 1987; Vol. 1, pp 25−52.
(48) Drew, H. R.; Samson, S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4040−4044.

It is of interest to compare how well deoxyinosine fits opposite the natural bases in duplex DNA. Crystallographic studies show that the dI-T base pair is relatively unsymmetrical in comparison to the natural base-pairs.[15] In an A-DNA duplex formed from the self-complementary sequence 5′-d(GGIGCTCC)-3′ the two C-1′ to C-1′ interglycosyl distances for the dI-T base pairs were both 10.3 Å. For one of the dI-T pairs λ(dI) and λ (T) were, respectively, 40° and 70°, and for the other, 48° and 72°. These parameters are very close to those found for the T-G mispair which implies similar duplex distortion. In a B-DNA duplex formed by the deoxydodecamer 5′-d(CG-CIAATTAGCG)-3′, dA assumed a *syn* conformation in order to pair with dI in an *anti* conformation.[12] One dA-dI base pair gave a C-1′ to C-1′ interglycosyl distance of 11.2 Å and the other 10.5 Å. For one pair, λ(dI) and λ(dA) were 49° and 37°, and for the other 59° and 40°. The distortion of the helix is slightly less for the dA-dI base pair than it is for the dI-T base pair.

In the models, with 3-nitropyrrole deoxyribonucleoside in an *anti* conformation (nitro group facing toward the opposing base), the cytosine and thymine each show a single hydrogen bond from N-4 and N-3, respectively, to the nitro group. When in the *syn* conformation the nitro group instead projects into the major groove where it is not in position to hydrogen bond either to dA or dG. Although the C-1′ to C-1′ interglycosyl distance is significantly greater when dA or dG base pair to nucleoside **4** in an *anti* conformation, the models indicate that hydrogen bonding is possible. Where the nitro group did approach within hydrogen bonding distance, values of 2.97−3.10 Å to the NH of the opposing base were measured on the model. In comparison, the NH−O distances of A-T and G-C base pairs fall in the range 2.86−2.93 Å.[49] There is a sizable body of experimental data which indicates that, at least in the solid state, nitro groups interact with N−H protons. According to Etter's analysis of hydrogen-bond patterns of nitroanilines, the N−H proton typically associates with the nitro group via a three-center hydrogen bond, $R_1^2[4]$.[50,51] Crystal structures of nitroanilines typically show NH−O distances in the range 3.0−3.2 Å. Direct conjugation of the nitro group with the exocyclic amino group significantly enhances the electron density of the nitro group oxygens and increases their ability to function as hydrogen bond acceptors. In the same way, conjugation between the ring nitrogen of pyrrole and the 3-nitro group must significantly enhance electron density on the nitro group oxygens. Nevertheless, predicting whether in the double helix environment the nitro group can compete with effective hydrogen bond acceptors, such as water, would be difficult. The results obtained by Loakes and Brown with a series of nitroindole deoxyribonucleosides would appear to support a model in which the nitro group is not involved in hydrogen bonding.[36] Despite a significant increase in size of the heterocycle, duplex structure is maintained with relatively high helix to coil transition temperatures, and lack of discrimination in pairing to each of the natural nucleosides is maintained. The T_m's of sequences containing 5-nitroindole are higher than the corresponding sequences containing 3-nitropyrrole. This clearly implicates stacking forces as the major contributing factor to the stabilization of nucleic acid duplexes by nitro substituted heterocycles.

In order to establish the relative role of base stacking and hydrogen bonding interactions a series of substituted five-membered ring heterocycle deoxyribonucleosides have been constructed and are being evaluated in both physicochemical and biochemical experiments. It is clear from these experiments that there will be other five-membered ring heterocycles which share many of the properties described herein for nitropyrrole. For example pyrrole 3-carboxamide also functions effectively as a universal spacer in dideoxynucleotide mediated sequencing. Moreover, the T_m's of oligonucleotides containing pyrrole 3-carboxamide are higher than those for 3-nitropyrrole.

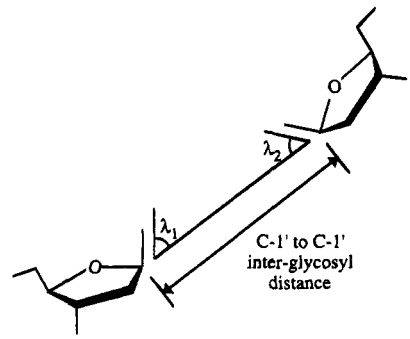


Figure 5. Angles λ1 and λ2 and the interglycosyl distance C-1′ to C-1′.

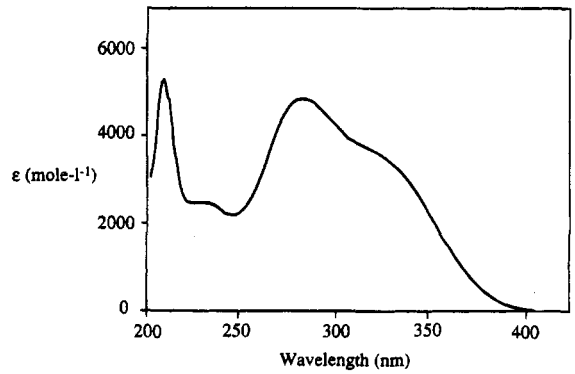


Figure 6. UV−visible spectrum 1-(2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole (**4**) in 0.1 M NH_4OAc, pH 6.96.

Perhaps the most significant outcome of this research is the unveiling of a new paradigm for modified base design. For the past decade most attempts to construct a universal base have focused on molecules which closely resemble either purines or pyrimidines and which were designed to hydrogen bond to each of the natural bases. 3-Nitropyrrole represents a unique departure from the conventional approach because it does not closely resemble either a purine or a pyrimidine nor does it contain a functional group of significant hydrogen bonding capacity. Instead it was designed on the basis of its potential to participate in stacking interactions. The ability of 3-nitropyrrole and structurally related compounds like nitroindoles to maintain duplex structure with relatively modest decreases in T_m supports the concept of the significance of base stacking interactions in maintaining helix structure. Nitroheterocycle containing oligonucleotides are likely to have other useful properties beyond simple function as universal spacers. Nishiwaki and co-workers reported that 3-nitropyrrole containing polyamides modeled after the DNA binding agents netropsin and distamycin cleave DNA on irradiation at 360 nm.[52] Nitroheterocycle containing oligonucleotides may mediate similar reactions.

Experimental Section

General Information. NMR spectra were obtained on a Varian VXR-500S spectrometer. ^1H and ^{13}C were referenced to TMS; and

(49) Seeman, N. C. In *Nucleic Acid Geometry and Dynamics*; Sarma, R. H., Ed.; Pergamon Press: New York, 1980; pp 133−135.
(50) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120−126.
(51) Panunto, T. W.; Urganczyk-Lipkowska, Z.; Johnson, R.; Etter, M. C. *J. Am. Chem. Soc.* **1987**, *109*, 7786−7797.
(52) Nishiwaki, E.; Lee, H.; Matsumoto, T.; Toyooka, K.; Sakurai, H.; Shibuya, M. *Tetrahedron Lett.* **1990**, *31*, 1299−1302.

85% phosphoric acid was used as an external standard for ^{31}P. Melting points were determined using a Buchi 510 apparatus and are uncorrected. FAB mass spectra were recorded by the Mass Spectroscopy Laboratory, Department of Medicinal Chemistry and Pharmacognosy, Purdue University. CD measurements were performed on a JASCO Model J6000 CD instrument, which was calibrated with (+)-camphorsulfonic acid [$\Delta\epsilon = 2.37$ (19.00 mdeg) at 290.5 nm]. 2′-Deoxyribose was purchased from Crystal Chem. 4,4′-Dimethoxytrityl chloride, β-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite, 1-(triisopropylsilyl)pyrrole, and all anhydrous solvents were purchased from Aldrich. TLC plates (Kieselgel 60F-254) and silica gel 60 PF254 containing gypsum for chromatotron purifications are products of Merck. All the reagents were used as received. Oligonucleotides were synthesized on an ABI Model 391 synthesizer using standard protocols. The oligonucleotides were purified using HPLC on a 25 cm × 4 mm PolyLC polyhydroxyethylaspartamide column or by 20% acrylamide-8 M urea electrophoretic gels and size exclusion chromatography (Sephadex G-25) using triethylammonium carbonate as buffer. The synthetic oligonucleotides listed in Table 1 were tested as primers for sequencing single-stranded DNA by the Sanger method. Sanger dideoxy sequencing was performed using the United States Biochemical (USB) Sequenase version 2.0 sequencing kit. The DNA sequenced was a Hind III-Bluescript SK+ subclone of a *Drosophila* neural peptide gene described previously by Nichols et al.[53] The template for the sequencing shown in Figure 1 was single-stranded DNA containing of the order of 4kb. Approximately 1 μg DNA was sequenced with 0.1 μg oligonucleotide according to conditions provided by the supplier using ^{35}S-dATP (Amersham). Aliquots of the sequencing reactions were electrophoresed on 6% acrylamide-8 M urea sequencing gels after which the gel was dried and exposed to Kodak XAR X-ray film. The four lanes for each primer are, from left to right, GATC.

Modeled B-DNA duplexes consisted of the sequence 5′-d(CGCGAATTCGCG)-3′ hybridized to 5′-d(CGCGAAT**4**CGCG)-3′ (**4** opposite A), 5′-d(CGC**4**AATTCGCG)-3′ (**4** opposite C), 5′-d(CGCGAATT**4**GCG)-3′ (**4** opposite G), and 5′-d(CGCG**4**ATTCGCG)-3′ (**4** opposite T). Structures were generated on a Silicon Graphics IRIS 4D/120GTX using the program QUANTA. The lowest energy conformation was obtained using the Newton-Raphson minimization equation performed in CHARMM.

1-(2′-Deoxy-3′,5′-di-*O*-toluoyl-β-D-ribofuranosyl)-3-nitropyrrole (3). To a stirred solution of 3-nitropyrrole (1.10 g, 9.81 mmol) in acetonitrile (145 mL) was added sodium hydride (0.33 g, 13.75 mmol) under nitrogen. When the solution became clear 1-chloro-2-deoxy-3,5-di-*O*-toluoyl-α-D-erythropentofuranose[54] (3.81 g, 9.81 mmol) was added. The mixture was stirred for 1 h and filtered, and the solid was washed with acetonitrile. The filtrate was evaporated to dryness, and the residue was purified by column chromatography on silica gel (hexane−acetone as the eluent solvent). Compound **3** was obtained as a white solid (4.0 g, 88%): mp 121−122 °C; ^1H NMR (acetone-d_6) 8.02 (dd, H-2, $J_{2,4} = J_{2,5} = 2.0$ Hz, 1H), 7.99−7.91 (m, aromatic H, 4H), 7.37−7.29 (m, aromatic H, 4H), 7.12 (dd, H-4, $J_{4,5} = 3.3$ Hz, $J_{2,4} = 2.0$ Hz, 1H), 6.69 (dd, H-5, $J_{4,5} = 3.3$ Hz, $J_{2,5} = 2.0$ Hz, 1H), 6.31 (dd, H-1′, $J = 8.0$, 6.0 Hz, 1H), 5.77−5.73 (m, H-3′, 1H), 4.68−4.63 (m, H-4′, 5′, 3H), 2.92−2.82 (m, H-2′, 2H), 2.41 (s, methyl H, 3H), 2.39 (s, methyl H, 3H); ^{13}C NMR (acetone, δ) 166.37 and 166.10 (*C* = O), 145.02 and 144.76 (CO-aromatic C), 130.37, 130.24, 129.98, 129.96 (aromatic C), 127.84 and 127.75 (CH_3-aromatic C), 120.95 (C-4), 120.70 (C-2), 105.98 (C-5), 89.84 (C-1′), 83.55 (C-4′), 75.87 (C-5′), 64.85 (C-3′), 39.55 (C-2′), 21.56 and 21.52 (methyl C); FAB mass 465.0 (MH^+), 353 ($C_{21}H_{21}O_5^+$). Anal. Calcd for $C_{25}H_{24}N_2O_7$: C, 64.65; H, 5.21; N, 6.03(%). Found: C, 64.83; H, 5.17; N, 6.01.

1-(2′-Deoxy-β-D-ribofuranosyl)-3-nitropyrrole (4). Compound **3** was poured into a solution of concentrated ammonia in methanol and heated overnight at 55 °C in a pressure vessel. The ammonia and methanol were removed by evaporation *in vacuo*, and the residue was separated by chromatography on silica gel (hexane−acetone as an eluent solvent). Compound **4** was obtained as a white solid (1.55 g, 83%): mp 95−96 °C; ^1H NMR (methanol-d_4) 7.98 (dd, H-2, $J_{2,4} = J_{2,5} = 2.0$ Hz, 1H), 7.01 (dd, H-4, $J_{4,5} = 3.5$ Hz, $J_{2,4} = 2.0$ Hz, 1H), 6.68 (dd, H-5, $J_{4,5} = 3.5$ Hz, $J_{2,5} = 2.0$ Hz, 1H), 5.97 (t, H-1′, $J = 6.6$ Hz), 4.44−4.40 (m, H-3′, 1H), 3.95 (m, H-4′, $J_{4',5''}$ and $J_{4',5'} = 4.0$ Hz, $J_{4',3'} = 7.5$ Hz), 3.73 (q, H-5′, $J_{4',5'} = 4.0$ Hz, $J_{5',5''} = 12$ Hz), 3.67 (q, H-5″, $J_{4,5''} = 4.0$ Hz, $J_{5',5''} = 12$ Hz), 2.44−2.34 (m, H-2′, 2H); ^{13}C NMR (methanol-d_4, δ) 121.51 (C-3), 121.01 (C-2), 106.13 (C-4), 90.26 (C-5), 89.13 (C-1′), 72.34 (C-4′), 63.11 (C-5′), 48.98 (C-3′), 42.83 (C-2′); high resolution FAB mass MH^+ (calcd 229.0824, found 229.0819); UV−vis spectrum (0.1 M NH_4OAc, pH 6.96) λ_{max} 284 nm ($\epsilon = 4.85 \times 10^3$) (Figure 6). Anal. Calcd for $C_9H_{12}O_5N_2$: C, 47.38; H, 5.37; N, 12.28. Found: C, 47.46; H, 5.37; N, 12.29.

1-(2′-Deoxy-5′-dimethoxytrityl-β-D-ribofuranosyl)-3-nitropyrrole (5). Compound **4** (330 mg, 1.45 mmol) and dimethoxytrityl chloride (500 mg, 1.48 mmol) were dissolved in pyridine (3 mL) and allowed to stand at room temperature for 1 h. Additional small portions of dimethoxytrityl chloride were added until the starting material disappeared completely by TLC analysis. The solution was poured into water. The products were extracted with ether, washed three times with water, dried over sodium sulfate, and separated by chromatography on silica gel (hexane−acetone as an eluent solvent). Compound **5** was obtained as a white foam (674 mg, 88%): ^1H NMR (methanol-d_4) 7.94 (t, H-2, $J = 2.0$ Hz, 1H), 7.46−7.19 (m, aromatic H, 9H), 7.04 (br t, H-4, 1H), 6.87−6.85 (br d, aromatic H, 4H), 6.67 (q, H-5, $J = 1.5$, 3.5 Hz, 1H), 6.08 (t, H-1′, $J = 6.5$ Hz, 1H), 4.56 (br s, H-3′, 1H), 4.11 (ddd, H-4′, $J = 4.0$ Hz, 1H), 3.77 (s, methyl H, 6H), 3.32−3.27 (m, H-5′, 2H), 2.56−2.45 (m, H-2′, 2H).

1-(2′-Deoxy-5′-dimethoxytrityl-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite) (6). To a solution of **5** (413 mg, 0.78 mmol) and diisopropylethylamine (0.80 mL, 4.60 mmol) in methylene chloride (4.0 mL) was added 2-cyanoethyl-*N*,*N*-diisopropylamino-chlorophosphoramidite (0.43 mL, 1.93 mmol). The reaction was completed in 10 min. The solution was diluted with ethyl acetate, washed with saturated $NaHCO_3$, dried over Na_2SO_4, and evaporated to dryness. The residue was purified by a rapid chromatography on silica gel on a chromatotron, using hexane−acetone−triethylamine (6:3:1) as an eluent solvent. Compound **6** was obtained as a white foam (450 mg, 79%): ^{31}P NMR (acetone-d_6) 149.94, 148.85 (phosphoramidite stereoisomers); ^1H NMR (acetone-d_6) 8.06 (t, H-2, 0.5H), 8.04 (t, H-2, 0.5H), 7.57−7.30 (m, aromatic H, 9H), 7.15 (q, H-4, 0.5H), 7.14 (q, H-4, 0.5H), 6.98−6.94 (m, aromatic H, 4H), 6.77−6.76 (m, H-5, 1H), 6.22−6.20 (m, H-1′, 1H).

1-(5′-*O*-DMTr-2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-succinate (7). A solution of 1-(5′-*O*-DMTr-2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole (250 mg, 0.47 mmol), succinic anhydride (142 mg, 1.42 mmol), and (dimethylamino)pyridine (57.5 mg, 0.47 mmol) in pyridine (2 mL) was stirred at room temperature until TLC showed that the reaction was complete (2 days). The solution was coevaporated with toluene. The residue was dissolved in $CHCl_3$, and the solution washed first with saturated NaCl solution and then water and dried over anhydrous Na_2SO_4, and then the solvent evaporated to yield a foam (>95% yield). The crude product was used for the following reaction without further purification: ^1H NMR ($CDCl_3$) 7.70−7.69 (m, H-2, 1H), 7.39−7.26 (m, aromatic H, 9H), 6.84−6.81 (m, aromatic H, 4H), 6.76−6.75 (m, H-4, 1H), 6.70−6.69 (m, H-5, 1H), 5.86 (q, H-1′, $J_{1',2'} = 8.3$ Hz, $J_{1',2''} = 6.0$ Hz, 1H), 5.41−5.39 (m, H-3′, 1H), 4.23−4.22 (m, H-4′, 1H), 3.79 (s, OCH_3, 6H), 3.39 (q, H-5′, $J_{5',5''} = 10.5$ Hz, $J_{4',5'} = 4.0$ Hz, 1H), 3.35 (q, H-5″, $J_{5',5''} = 10.5$ Hz, $J_{4',5''} = 4.0$ Hz, 1H), 2.73−2.49 (m, H-2′ and CH_2CH_2, 6H).

[5′-(3′-Adenosylyl)-2′-deoxy-β-D-ribofuranosyl]-3-nitropyrrole (9). The dimer was synthesized by the procedure published by Bonora et al.[30] PEG_{5000} (1.0 g, 0.20 mmol of free OH groups) was dehydrated by co-evaporation with anhydrous pyridine and dissolved in 1,2-dichloroethane (DCE). To a solution of 1-(5′-*O*-DMTr-2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-succinate (213 mg, 0.34 mmol) dissolved in DCE (5 mL) was added dicyclohexylcarbodiimide (38 mg, 0.18 mmol) with stirring at 0 °C. The solution was filtered after 15 min. The filtrate was poured into the PEG solution, followed by addition of 4-(dimethylamino)pyridine (50 mg, 0.41 mmol). The combined solution was stirred at room temperature for 1 day. After filtering the modified PEG was precipitated with diethyl ether. The precipitate was recrystallized from absolute ethanol and dried *in vacuo*

(53) Nichols, R.; Schneuwly, S. A.; Dixon, J. E. *J. Biol. Chem.* **1988**, *263*, 12167−12170.

(54) Bhat, C. C. In *Synthetic Procedures in Nucleic Acid Chemistry*; Zorbach, W., Tipson, R. S., Eds.; Wiley Interscience: New York, 1968; Vol. 1, pp 521−522.

Table 6. Positional Parameters and Estimated Standard Deviation of Non-Hydrogen Atoms[a]

atom	X	Y	Z	B (Å²)
O-4′	0.2289(3)	0.4103(2)	0.4496(2)	2.65(4)
O-3′	−0.0675(4)	0.6733(2)	0.4627(2)	3.49(5)
O-5′	0.5128(4)	0.6014(3)	0.4560(2)	4.16(6)
$O\text{-}3_2$	0.6992(5)	0.1685(3)	0.7884(2)	4.73(6)
$O\text{-}3_1$	0.7507(4)	0.3530(3)	0.7438(2)	5.15(7)
N-1	0.2121(4)	0.3160(2)	0.6016(2)	2.37(5)
N-3	0.6477(5)	0.2608(3)	0.7457(2)	3.44(6)
C-4′	0.1855(5)	0.5277(3)	0.4150(2)	2.38(6)
C-3′	0.0808(5)	0.5934(3)	0.4999(2)	2.35(5)
C-2′	−0.0097(5)	0.4927(3)	0.5607(3)	3.13(7)
C-1′	0.0934(5)	0.3783(3)	0.5264(2)	2.46(6)
C-2	0.3922(5)	0.3522(3)	0.6370(2)	2.60(6)
C-3	0.4556(5)	0.2610(3)	0.6985(2)	2.61(6)
C-4	0.3119(6)	0.1685(3)	0.6989(3)	3.20(7)
C-5	0.1641(5)	0.2043(3)	0.6382(3)	2.95(6)
C-5′	0.3738(6)	0.5836(3)	0.3762(3)	3.46(7)

[a] Numbers in parentheses are estimated standard deviations in the least significant digits.

(loading amount: 55 μmol/g (determined by the trityl analysis)). The unreacted OH groups of the PEG were capped by an acetonitrile solution containing 10% acetic anhydride, 10% 2,6-lutidine, and 10% *N*-methylimidazole (10 mL) in a reaction carried out at room temperature for 3 min. The modified PEG (**8**) was precipitated from an ice-cooled solution with ether, filtered, washed with ether, and dried *in vacuo*.

1-(5′-*O*-DMTr-2′-deoxy-β-D-ribofuranosyl)-3-nitropyrrole-3′-*O*-succinate-PEG (8) (1.0 g) was dissolved in DCE (10 mL), to which was added dropwise a 6% soln of trichloroacetic acid in DCE (10 mL) with stirring at 0 °C. After 15 min the polymer support was precipitated with diethyl ether, filtered and washed with diethyl ether. This polymer support was further purified by dissolving it in DCE and reprecipitating with diethyl ether, following which it was dried *in vacuo*. For condensation the polymer support was dissolved in acetonitrile (3 mL) under nitrogen. To this solution was added 5′-dimethoxytrityl-N^6-benzoyldeoxyadenosine-3′-(2-cyanoethyl-*N*,*N*-diisopropyl)phosphoramidite (350 mg, 0.4 mmol), followed by a solution of tetrazole (70 mg, 1.0 mmol) in acetonitrile (2 mL). The solution was stirred for 5 min. The polymer was precipitated with diethyl ether, filtered, washed with diethyl ether, and dried *in vacuo*. The polymer was dissolved in acetonitrile (20 mL), to which a 0.05 M iodine solution in THF/H_2O/pyridine (7:2:1, 10 mL) was added. After 10 min, the PEG-phosphate derivative was precipitated with diethyl ether, filtered, and washed with diethyl ether. The product was dissolved in a 30% ammonium hydroxide, heated at 55 °C for 1 day, and lyophilized. The residue was dissolved in water (10 mL), treated by a solution of 80% acetic acid (10 mL) for 30 min. The acetic acid was removed by evaporation. The residue was redissolved in water and extracted with chloroform. The dimer (**9**) was separated from PEG by chromatography using SEPHADEX G-25. The final crude product was purified by a preparative reverse phase HPLC (C_{18} column; buffer A: 0.1 M triethylammonium acetate, pH = 7.0; buffer B: acetonitrile). Dimer **9** was characterized by 1D NMR, COSY, and NOESY: ^{1}H NMR (D_2O), 3-nitropyrrole deoxyribonucleoside, 7.63 (t, H-2, 1H), 6.74 (t, H-4, 1H), 6.29 (t, H-5, 1H), 5.83 (t, H-1′, 1H), 4.56 (m, H-3′, 1H), 3.98 (m, H-4′, 1H), 3.59 (d, H-5′, 2H), 2.45−2.30 (m, H-2′, 2H); deoxyadenosyl, 8.07 (s, H-8, 1H), 8.04 (s, H-2, 1H), 6.17 (t, H-1′, 1H), 4.76 (m, H-3′, 1H), 4.05 (m, H-4′, 1H), 3.70 (m, H-5′, 2H), 2.58−2.46 (m, H-2′, 2H).

X-ray Analysis of 1-(2′-Deoxy-β-D-ribofuranosyl)-3-nitropyrrole (4). Crystals were grown in a mixture of chloroform and methanol. A colorless cube of **4** ($C_9H_{12}N_2O_5$) having approximate dimension of 0.22 × 0.21 × 0.19 mm was mounted on a glass fiber in a random orientation. Preliminary examination and data collection were performed with a rotating anode generator, CuKα radiation (λ = 1.54184 Å) on an Enraf-Nonius CAD4 computer controlled κ axis diffractometer equipped with a graphite crystal, incident beam monochromator. Cell constants and an orientation matrix for data collection were obtained from least-squares refinement, using the setting angles of 25 reflections in the range $43 < \theta < 48°$, measured by the computer controlled diagonal slit method of centering. The following data were obtained: orthorhombic, $P2_12_12_1$, a = 6.6248 (6), b = 11.1648 (9), c = 13.6041 (9) Å, V = 1006.2 (2) Å^3, Z = 4, M_r = 228.21, ϱ_{calc} = 1.506 g cm^{-3}, μ = 10.19 cm^{-1}, $F(000)$ = 480.0.

The data were collected at a temperature of 293 ± 1 K using the ω-2θ scan technique ($2\theta_{max}$ = 130.0°). A total of 1030 unique reflections were collected. Lorentz and polarization were applied to the data, and an empirical absorption correction based on the method of Walker and Stuart[55] was applied. Relative transmission coefficients ranged from 0.440 to 1.000 with an average of 0.593.

The structure was solved using the structure solution program SHELX-86.[56] The remaining atoms were located in succeeding difference Fourier syntheses. Hydrogen atoms were located and added to the structure factor calculation but not refined. The structure was refined in full-matrix least-square where the function minimized was $\sum w(|F_o| - |F_c|)^2$. Only the 957 reflections having intensities 3.0 times greater than their standard deviation were used in the refinements. The final cycle of refinement included 153 variable parameters and converged (largest parameter shift was 0.07 times its esd) with unweighted and weighted agreement factors of R = 0.042 and R_w = 0.063. All calculations were performed on a VAX computer. Refinement was done using MolEN.[57] Final atomic parameters are given in Table 6.[58]

Acknowledgment. The National Institutes of Health is gratefully acknowledged for support of this research through NIH Grant GM45551-03. We thank M. F. Tibbetts, Biology Department, University of Michigan for his contribution to the sequencing, S. E. Olsen and T. A. Poley, Biological Chemistry, University of Michigan for T_m studies, and N. Gerry and D. Klewer, Department of Medicinal Chemistry and Pharmacognosy for the molecular modeling studies. We thank P. E. Fanwick, Chemistry Department, Purdue University, for his assistance and the use of the diffractometer.

JA942133O

(55) Walker, N.; Stuart, D. *Acta Crystallogr.* **1983**, *A39*, 158−166.
(56) Sheldrick, G. M. *SHELX 86*; University of Gottingen: Gottingen, Germany, 1986.
(57) Fair, K. *MolEN*; Enraf-Nonius: Delft, The Netherlands, 1990.
(58) The authors have deposited atomic coordinates for this structure with the Cambridge Crystallographic Data Centre. The coordinates can be obtained, o. r., from the Director, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK.