# Indole as an Artificial DNA Base Incorporated via an Acyclic 2′-Deoxyriboside Substitute

Claudia Wanninger, Hans-Achim Wagenknecht*

University of Regensburg, Institute for Organic Chemistry, Universitätsstr. 31, 93053 Regensburg, Germany
Fax +49(941)9434617; E-mail: achim.wagenknecht@chemie.uni-regensburg.de

*Received 16 May 2006*

**Abstract:** Indole was synthetically incorporated into DNA as an artificial base at specific sites of duplex DNA. An acyclic substitute for the 2′-deoxyribose was applied in order to obtain a chemically stable nucleoside analogue that can be synthesized by a fast and facile procedure. Studies by methods of optical spectroscopy revealed that the indole base surrogate is intercalated into the DNA base stack and behaves as a universal base analogue with only little influence of the counterbase.

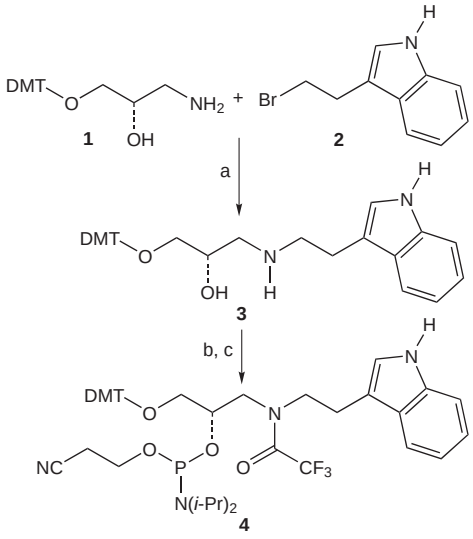
**Key words:** DNA, indole, oligonucleotide, base analogue

Indole-containing nucleosides and oligonucleotides are target molecules of significant interest in bioanalytical and medicinal applications.[1–4] Especially the synthesis of the natural-like β-nucleoside with indole as an artificial aglycone has been published several times[3,4] and used for instance in DNA base pairing and stacking experiments lacking any hydrogen bonding.[4] Moreover, indole represents a very promising charge trap for charge-transport studies in DNA due its low oxidation potential and the characteristic transient absorption of the corresponding radical.[5–7] Those experiments have been performed with indole either as part of tryptophan in DNA binding peptides[5] or proteins,[6] or as an artificial DNA base.[7] It was shown that the aromatic N–H group is required for an efficient trapping of positive charges due to the coupling with deprotonation.[5,7]

The already-mentioned, well-characterized and intensively studied indole nucleoside contains the aromatic moiety β-glycosidically linked via the nitrogen in position 1 to the anomeric center of 2′-deoxyribofuranose.[3,4] Although a significant mesomeric stabilization exists within the aromatic indole system, the glycosylamine is subject to aqueous hydrolysis. This instability complicates both the automated DNA synthesis (that takes part partially under acidic conditions) and the biochemical and biophysical experiments that are performed in aqueous buffer solutions.[8] For instance, the α-anomer and the pyranosides of the indole nucleoside were detected as contaminants during such experiments.[2] Thus, we chose a new approach in order to study the stacking interactions of indole by incorporation as an artificial DNA base at specific sites in duplex DNA. The 2′-deoxyribofuranoside moiety was replaced by an acyclic linker system which is tethered to the C-3 position of the indole heterocycle. Recently, we used a similar synthetic approach for the incorporation of ethidium[9] and perylene[10] as DNA base surrogates. Avoiding the labile glycosidic bond, the indole nucleoside analogue **3** provides the necessary chemical stability for the preparation of indole–DNA conjugates via automated phosphoramidite chemistry and for biophysical and bioanalytical applications in aqueous buffers.

Herein, we want to present the synthetic work for the indole nucleoside substitute **3**, the corresponding DNA building block **4** and the indole-modified oligonucleotides. Furthermore, we examined preliminarily the modified DNA duplexes by methods of optical spectroscopy, and characterized them by their melting temperatures, in comparison with unmodified DNA.

The procedure for the synthesis of such indole-modified DNA is straightforward (Scheme 1): The nucleophilic substitution of the 3-(2-bromoethyl)indole (**2**) with the DMT-protected (*S*)-(–)-3-amino-1,2-propanediol **1**[9] yields the nucleoside substitute **3**.[11] According to our experience, the secondary amine of the linker system needs to be protected with a trifluoroacetyl group in order to avoid irreversible acetylation during the capping steps of

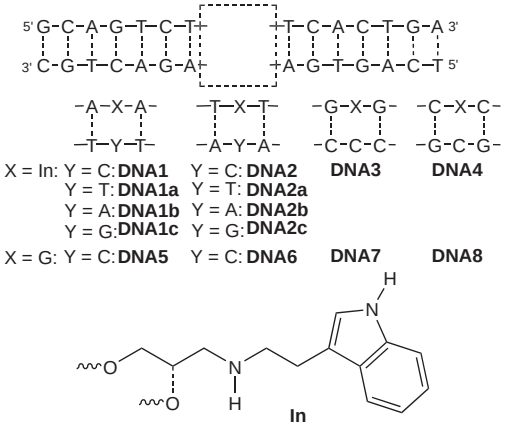



**Scheme 1** Synthesis of the DNA building block **4**: a) $EtN(i\text{-}Pr)_2$ (4 equiv), DMF, r.t., 7 d; 55%; b) $(F_3CCO)_2O$ (4.5 equiv), pyridine (abs.), $CH_2Cl_2$ (abs.), 0 °C, 30 min, r.t., 30 min, extraction with aq sat. $NaHCO_3$; 84%; c) $Et_3N$ (dry, 2.5 equiv), 2-cyanoethyl diisopropylamidochloridophosphite (1.5 equiv), $CH_2Cl_2$ (abs.), 45 min, r.t.; quant.

*SYNLETT* 2006, No. 13, pp 2051–2054
Advanced online publication: 09.08.2006
DOI: 10.1055/s-2006-948209; Art ID: G16506ST

the DNA synthesizer cycle.[9,12] The synthesis of the phosphoramidite **4** is accomplished by standard procedures and can be applied for the automated preparation of modified oligonucleotides using a modified coupling protocol[9] for an extended coupling time on the DNA synthesizer.

Using **4**, we prepared a range of four indole-modified duplexes **DNA1**–**DNA4** (Scheme 2). The sequences of the four DNA duplexes are identical except the base pairs that are placed adjacent to the indole (In) modification site. The indole-modified single-stranded (ss) oligonucleotides **ssDNA1**–**ssDNA4** were quantified by their absorbance at 260 nm[13,14] and identified by ESI mass spectrometry.[15]




**Scheme 2** Sequences of the In-modified duplexes **DNA1**–**DNA4**, **DNA1a**–**DNA1d**, **DNA2a**–**DNA2c**, and the unmodified duplexes **DNA5**–**DNA8**.

First, we measured the absorption spectra of the indole-modified **DNA1**–**DNA4** in comparison with the unmodified duplexes **DNA5**–**DNA8** (Figure 1). Although the overall absorption of **DNA1**–**DNA4** is lower, the normalized UV/Vis spectra show clearly the small contribution of the indole moiety ($\lambda_{max}$ = ca. 280 nm) compared to **DNA5**–**DNA8** in the absorption range between 260 nm and 310 nm.

In order to evaluate the influence of the indole nucleoside analogue on the thermal stability of the duplexes the DNA melting temperatures ($T_m$) of **DNA1**–**DNA4** were recorded at 260 nm (Table 1) and compared with that of the corresponding unmodified duplexes **DNA5**–**DNA8**. Obviously, the incorporation of the indole nucleoside surrogate decreases the thermal stability by values between 11 °C and 15 °C. This reduction indicates the influence of the acyclic linker as part of the nucleoside surrogate **3** instead of the 2′-deoxyribofuranoside in natural nucleosides. But with respect to the potential applicability of the indole **3** as a universal base analogue[16] it is more important to measure the influence of the counterbase to the indole modification site (In). Hence, we varied representatively in **DNA1** and **DNA2** the base that is opposite to the indole DNA base analogue. In both duplex series, **DNA1**/**DNA1a**–**DNA1c** and **DNA2**/**DNA2a**–**DNA2c**, the melting temperatures behave very similar.

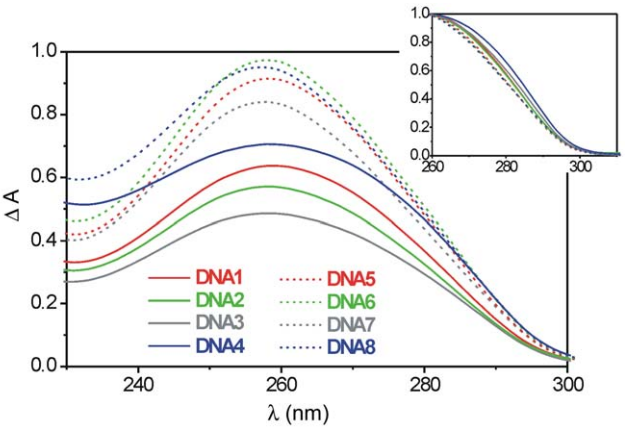



**Figure 1** Absorption spectra of the indole-modified duplexes **DNA1**–**DNA4** in comparison to the unmodified duplexes **DNA5**–**DNA8** (2.5 μM in 10 mM Na–$P_i$ buffer). The inset with the normalized spectra shows the absorption enhancement between 260 nm and 310 nm due to the presence of the indole.

From these experiments it becomes clear that the counterbase seems not to have any significant influence on the stacking interactions of the indole due to the lacking hydrogen bonding to any kind of counterbase.

To obtain more information about the structural influence of the artificial indole group we characterized the modified duplexes **DNA1**–**DNA4** by CD spectroscopy (Figure 2). The corresponding spectra exhibit clearly the shape that is typical for the B-type conformation of the DNA duplexes. As mentioned previously, the reduced melting temperatures compared to **DNA5**–**DNA8** reveal a conformational perturbation at the indole modification site in **DNA1**–**DNA4**. The results from CD spectroscopy, however, show that the structural perturbation is very local and the substitution of the 2′-deoxyribofuranose moiety in **3** does not change the global B-DNA conformation.

**Table 1** Melting Temperature ($T_m$) of **DNA1**–**DNA4**, **DNA1a**–**DNA1c**, **DNA2a**–**DNA2c**, and **DNA5**–**DNA8**[a]

| Duplex | $T_m$ (°C) up | $T_m$ (°C) down | Duplex | $T_m$ (°C) up | $T_m$ (°C) down |
|---|---|---|---|---|---|
| **DNA1** | 52.5 | 51.5 | **DNA5** | 65.3 | 64.7 |
| **DNA2** | 53.0 | 52.8 | **DNA6** | 66.1 | 65.4 |
| **DNA3** | 59.9 | 59.5 | **DNA7** | 70.9 | 70.4 |
| **DNA4** | 57.0 | 55.9 | **DNA8** | 71.7 | 71.0 |
| **DNA1a** | 55.0 | 53.5 | **DNA2a** | 52.0 | 51.5 |
| **DNA1b** | 55.0 | 53.4 | **DNA2b** | 54.1 | 52.9 |
| **DNA1c** | 54.7 | 52.8 | **DNA2c** | 53.1 | 54.9 |

[a] Conditions: λ = 260 nm, 10–90 °C, 2.5 μM DNA in 10 mM Na–$P_i$ buffer, 250 mM NaCl, pH 7.

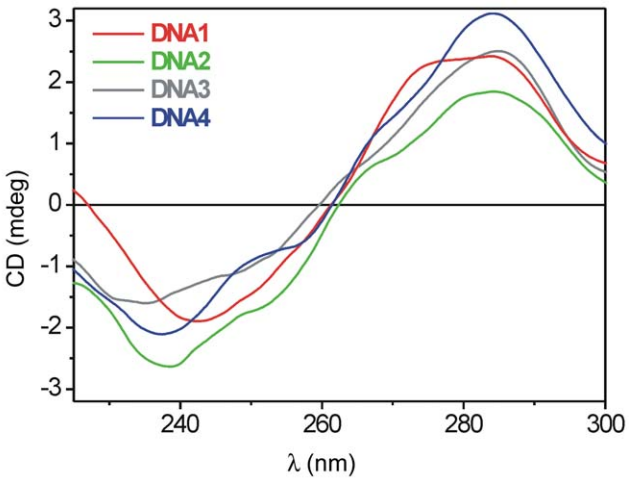



**Figure 2** CD spectra of the indole-modified duplexes **DNA1**–**DNA4** (2.5 μM in 10 mM Na–$P_i$ buffer).

Finally, we measured the steady-state emission spectra of **DNA1**–**DNA4**; 290 nm was chosen as the wavelength for excitation, which is typical for emission experiments with indole. It is important to point out that in comparison to the strong emission of the indole derivative **2**, the fluorescence of the indole in DNA is quenched by at least 90%, probably due to electron transfer processes. The maximum of the remaining emission of the indole-modified DNA duplexes depends on the flanking sequence. The typical maximum of the emission of indole heterocycle can be found at ca. 350 nm (e.g. in case of **2**), with adjacent guanines it shifts to ca. 360 nm (**DNA3**), with adenines to ca. 370 nm (**DNA1**) and with pyrimidines to ca. 405 nm (**DNA2**, **DNA4**). The latter results indicate a strong interaction of the intercalated indole chromophore with the adjacent DNA bases (Figure 3).

In conclusion, it was shown that indole can be synthetically incorporated into DNA as an artificial DNA base using an acyclic substitute for the 2′-deoxyribose of natural oligonucleotides. Although destabilizing the duplexes, the indole nucleoside analogue was found to be intercalated within the DNA base stack and to be a universal base analogue that behaves indiscriminately towards each of the four natural bases in the DNA duplexes.

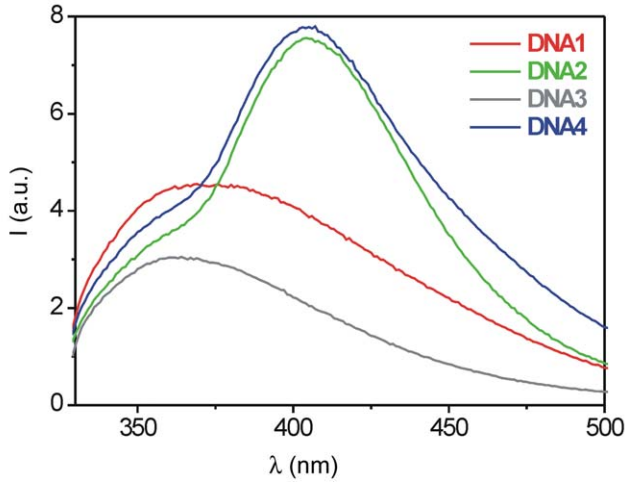



**Figure 3** Fluorescence spectra of the indole-modified duplexes **DNA1**–**DNA4** (2.5 μM in 10 mM Na–$P_i$ buffer, $\lambda_{exc}$ 290 nm).

UV/Vis spectra and melting temperatures were measured on a Cary 100 (Varian). Fluorescence spectra were recorded on a Fluoromax-3 (Jobin-Yvon) with a bandpass of 5 nm (excitation and emission) and correction for intensity and for Raman emission from the buffer solution. The CD spectroscopy was performed on a J-715 (Jasco). ESI-MS was performed on a TSY 7000 (Finnigan). C18-RP HPLC columns (300 Å) were from Supelco. The oligonucleotides were prepared on an Expedite 8909 DNA synthesizer (ABI) using CPG (1 μmol) and chemicals from ABI and Glen Research. The trityl-off oligonucleotides were cleaved and deprotected by treatment with concd $NH_4OH$ at 60 °C for 10 h, dried and purified by HPLC on RP-C18 (300 Å, Supelco) using the following conditions: A = $NH_4OAc$ buffer (50 mM), pH = 6.5; B = MeCN; gradient = 5–15% B over 60 min. Duplexes were formed by heating to 90 °C (10 min), followed by slow cooling.

## Acknowledgment

This work was supported by the the Deutsche Forschungsgemeinschaft (Wa 1386/7), the Fonds der Chemischen Industrie, and the University of Regensburg.

## References and Notes

(1) See, for example: (a) Yokoyama, M.; Nomura, M.; Togo, H.; Seki, H. *J. Chem. Soc., Perkin Trans. 1* **1996**, 2145. (b) Yokoyama, M.; Nomura, M.; Tanabe, T.; Togo, H. *Heteroat. Chem.* **1995**, *6*, 189.
(2) Zhang, X.; Lee, I.; Berdis, A. J. *Org. Biomol. Chem.* **2004**, *2*, 1703.
(3) (a) Dinh, T. H.; Bayard, M.-J.; Igolen, J. *C. R. Seances Acad. Sci., Ser. C* **1976**, *283*, 227. (b) Girgis, N. S.; Cottam, H. B.; Robins, R. K. *J. Heterocycl. Chem.* **1988**, *25*, 361. (c) Cornia, M.; Casiraghi, G.; Zetta, L. *J. Org. Chem.* **1991**, *56*, 5466.
(4) Lai, J. S.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 3040.
(5) (a) Wagenknecht, H.-A.; Stemp, E. D. A.; Barton, J. K. *J. Am. Chem. Soc.* **2000**, *122*, 1. (b) Wagenknecht, H.-A.; Stemp, E. D. A.; Barton, J. K. *Biochemistry* **2000**, *39*, 5483. (c) Mayer-Enthart, E.; Kaden, P.; Wagenknecht, H.-A. *Biochemistry* **2005**, *44*, 11749.
(6) (a) Rajski, S. R.; Kumar, S.; Roberts, R. J.; Barton, J. K. *J. Am. Chem. Soc.* **1999**, *121*, 5615. (b) Wagenknecht, H.-A.; Rajski, S. R.; Pascaly, M.; Stemp, E. D. A.; Barton, J. K. *J. Am. Chem. Soc.* **2001**, *123*, 4400.
(7) Pascaly, M.; Yoo, J.; Barton, J. K. *J. Am. Chem. Soc.* **2002**, *124*, 9083.
(8) See for ethidium nucleoside, for example: Amann, N.; Wagenknecht, H.-A. *Tetrahedron Lett.* **2003**, *44*, 1685.
(9) (a) Huber, R.; Amann, N.; Wagenknecht, H.-A. *J. Org. Chem.* **2004**, *69*, 744. (b) Amann, N.; Huber, R.; Wagenknecht, H.-A. *Angew. Chem. Int. Ed.* **2004**, *43*, 1845.
(10) Wagner, C.; Wagenknecht, H.-A. *manuscript submitted for publication.*
(11) Compound **3** was purified by flash chromatography with $CH_2Cl_2$:MeOH:$Et_3N$ = 100:3:0.1. Spectroscopic data of **3**: $^1$H NMR (300 MHz, DMSO-$d_6$): δ = 2.99 (m, 6 H, OC$H_2$, C*H*OH, C$H_2$N*H*), 3.34 (m, 2 H, $CH_2$-indole, masked under residual water in the sample), 3.73 (m, 8 H, OC$H_3$, NHC$H_2$CH2-indole), 6.87 (d, 4 H, arom. linker), 6.99 (m, 1

H, arom. indole), 7.08 (m, 1 H, arom. indole), 7.24 (m, 7 H, arom. linker), 7.39 (m, 2 H, arom. linker, 1 H arom. indole), 7.53 (m, 1 H, arom. indole), 10.95 (br s, 1 H, N*H*-indole). $^{13}$C NMR (75,4 MHz, DMSO-$d_6$): δ = 45.03 (*C*$H_2$NH), 55.25 (O*C*$H_3$), 66.06 (*C*$H_2$O), 73.47 (*C*HOH), 111.68, 113.37, 118.43, 118.56, 121.29, 123.20, 124.13, 127.12, 127.92, 128.03, 129.95, 135.78, 135.83, 136.36, 145.13, 149.82, 158.26. HRMS (ESI): *m/z* calcd: 537.2753 [$MH^+$]; found: 537.2761 [$MH^+$].

(12) Interestingly, the trifluoroacetylated derivative of **3** shows two sets of NMR signals (due to *cis*/*trans* isomers of the immonium structure): $^1$H NMR (300 MHz, DMSO-$d_6$): δ = 2.90 (m, 3 H, OC$H_2$, C*H*OH), 3.35 (m, 2 H, NC$H_2$ linker), 3.64 (m, 2 H, NC$H_2$C$H_2$-indole), 3.71 (m, 6 H, OC$H_3$), 4.01 (m, 2 H, NC$H_2$C$H_2$-indole), 6.85 (m, 4 H, arom. linker), 6.93–7.17 (m, 2 H, arom. indole), 7.22 (m, 7 H, arom. linker), 7.45 (m, 3 H, arom.), 7.55 (d, 1 H, arom. indol), 10.86 (br s, 0.5 H, N*H*-indole, *cis*-isomer), 10.91 (br s, 0.5 H, N*H*-indole, *trans*-isomer).

(13) For the extinction coefficients of the oligonucleotides, see: Puglisi, J. D.; Tinoco, I. *Methods Enzymol.* **1989**, *180*, 304.

(14) Indole extinction coefficient: $\varepsilon_{260}$ = 4000 $M^{-1}cm^{-1}$ in MeOH.

(15) ESI ms data (negative mode): **ssDNA1**: *m/z* calcd: 5157.9 $[M - H^+]^-$; found: 1719.6 $[M - 3\ H^+]^{3-}$; **ssDNA2**: *m/z* calcd: 5139.9 $[M - H^+]^-$; found: 1713.6 $[M - 3\ H^+]^{3-}$; **ssDNA3**: *m/z* calcd: 5189.9 $[M - H^+]^-$; found: 1730,3 $[M - 3\ H^+]^{3-}$; **ssDNA4**: *m/z* calcd: 5009.9 $[M - H^+]^-$; found: 1703.7 $[M - 3\ H^+]^{3-}$.

(16) A universal base analogue does not discriminate between the four natural DNA bases as the counterpart in the complementary strand; see for example: (a) Loakes, D.; Brown, D. M. *Nucleic Acids Res.* **1994**, *22*, 4039. (b) Loakes, D.; Hill, F.; Brown, D. M.; Salisbury, S. A. *J. Mol. Biol.* **1997**, *270*, 426.