Tetrahedron Letters 49 (2008) 7264–7267

Contents lists available at ScienceDirect

Tetrahedron Letters

journal homepage: www.elsevier.com/locate/tetlet

Synthesis and DNA duplex recognition of a triplex-forming oligonucleotide with an ureido-substituted 4-phenylimidazole nucleoside

Falk Wachowius, Michael Rettig, Gottfried Palm, Klaus Weisz *

Institut für Biochemie, Ernst-Moritz-Arndt-Universität Greifswald, Felix-Hausdorff-Str. 4, D-17487 Greifswald, Germany

ARTICLE INFO

Article history:
Received 31 August 2008
Revised 1 October 2008
Accepted 3 October 2008
Available online 8 October 2008

Keywords:
Nucleoside analog
Triple helix
UV melting
4-Phenylimidazole

ABSTRACT

A 4-(3-*n*-butylureidophenyl)imidazole nucleoside was successfully incorporated into a triplex-forming oligonucleotide (TFO). Binding affinity and base pair selectivity of the TFO containing this non-natural nucleoside were studied with various duplex targets containing all four possible Watson–Crick base pairs opposite the nucleoside analog in the third strand. Triplex thermal stabilities indicate that the synthetic nucleoside acts as a universal base in binding to all four possible Watson–Crick base pairs with moderate affinity but poor selectivity. Based on an analysis of its binding thermodynamics, this can be rationalized by the absence of strong specific interactions and more favorable entropic contributions upon triplex formation.

The use of triple-helix forming oligonucleotides in targeting defined stretches of double-helical DNA with high affinity and selectivity to generate three-stranded triplex structures has attracted much interest in past years due to their various potential applications. Thus, triplex formation may be used to block the expression of particular genes,[1,2] for site-directed mutagenesis[3,4] or for the mapping of genomic DNA.[5,6] Pyrimidine TFOs can bind parallel to the purine strand of a Watson–Crick duplex in its major groove through T·A or C^+·G Hoogsteen base pairing, yet binding suffers from the limited base pair recognition code with poor binding at inversion sites within a homopurine·homopyrimidine tract of the duplex. In order to overcome these limitations, substantial efforts aimed at a better affinity and selectivity toward a mixed target sequence have been made through base, sugar, and backbone modifications of the TFO.[7] However, despite some success in achieving a better base pair recognition, there is currently a lack of sufficient structural and thermodynamic data that would allow a more reliable prediction for the binding of newly designed nucleotides within a triple-helical complex.

Recently, we have synthesized various organic-soluble nucleoside receptors based on the 4-(3-aminophenyl)imidazole structure[8] that was initially employed by the Dervan group in a nucleoside surrogate.[9] Their ability to recognize a free CG Watson–Crick base pair primarily through specific hydrogen bond interactions was tested in aprotic organic solvents.[8] Based on ^{1}H NMR measurements in methylene chloride, strong binding was observed for the 3-*n*-butylureido-substituted phenylimidazole derivative. Such alkylated ureido substituents were first introduced in a naphth[1,2-*d*]imidazole base analog,[10] and have also been subsequently employed for other non-natural bases to provide for two potential hydrogen bonds to the guanine of a targeted CG base pair. Indeed, subsequent 2D NOE experiments clearly indicated formation of a base triple with the novel N-alkylated butylureidophenyl imidazole base associated via three hydrogen bonds to the CG base pair (Fig. 1). In contrast, we only observed a weak and less specific hydrogen bond-mediated complexation under such aprotic conditions for the 4-(3-benzamidophenyl)imidazole receptor which had previously been incorporated by Dervan and co-workers into a triplex-forming oligonucleotide and has been shown to preferentially bind opposite TA and CG base pairs, albeit through an unexpected intercalative mode of binding.[9,11] In order to shed more light on the

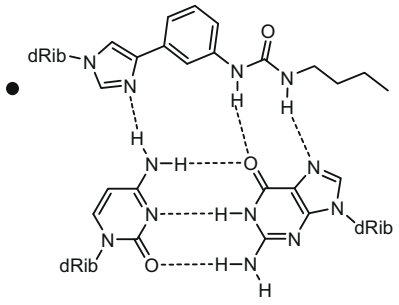


Figure 1. Hydrogen bond-mediated recognition of a CG base pair by a 4-(3-*n*-butylureidophenyl)imidazole nucleoside receptor; the filled circle indicates the mean C1′ position of protonated third strand cytosine in a non-visible superimposed C^+·GC base triad.

* Corresponding author. Tel.: +49 (0) 3834 864426; fax: +49 (0) 3834 864427.
E-mail address: weisz@uni-greifswald.de (K. Weisz).

0040-4039/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetlet.2008.10.015

forces that govern base–base interactions in the two different environments and to expand our model studies in aprotic solvents to the formation of triple-helical complexes under physiological conditions, we herein report on the incorporation of the 4-(3-*n*-butylureidophenyl)imidazole-2′-deoxynucleoside **D4** into a TFO and on its binding to various duplex targets through UV melting experiments.

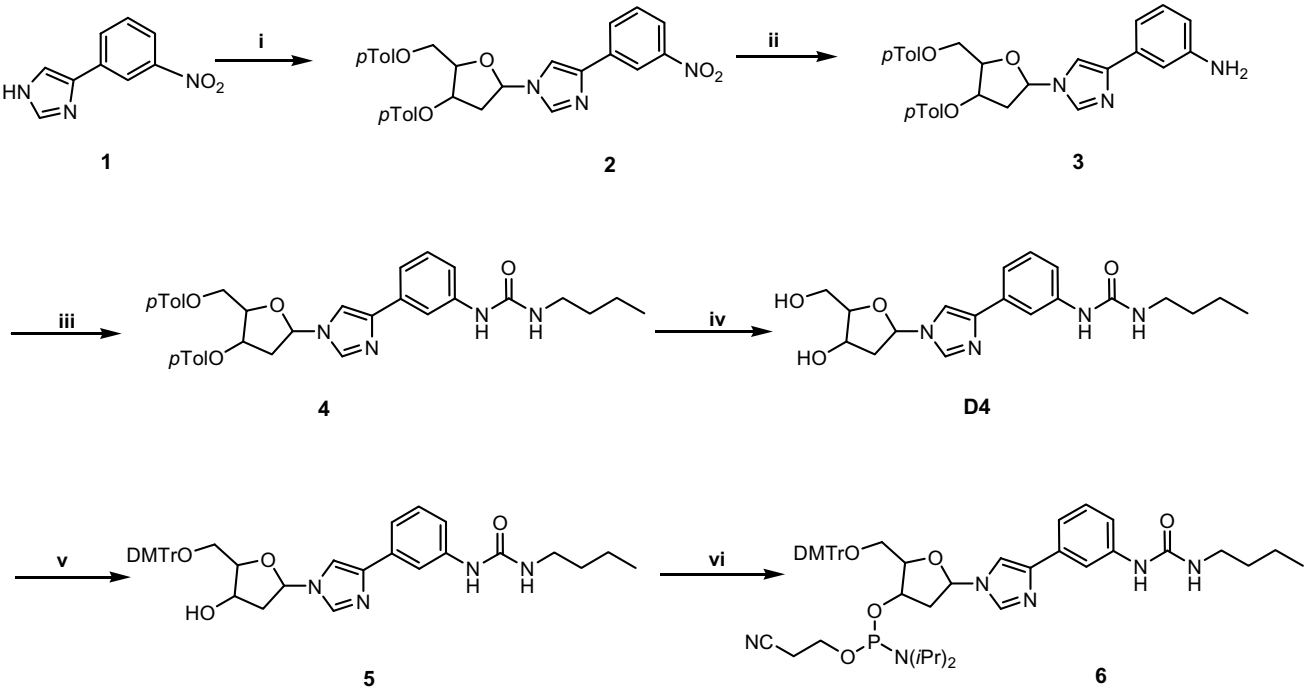
A scheme for the preparation of the 5′-tritylated and 3′-phosphitylated 2′-deoxyribonucleoside **D4**, which was used as synthon for the oligonucleotide synthesis, is depicted in Scheme 1. The sodium salt of 4-(3-nitrophenyl)imidazole, prepared from 1-nitro-3-(bromoacetyl)benzene and formamide according to the Bredereck imidazole synthesis,[8,12] was condensed with 2-deoxy-3,5-di-*O*-*p*-toluoyl-α-D-ribofuranosyl chloride[13] in acetonitrile to yield the 4-(3-nitrophenyl)imidazole nucleoside in a β:α anomeric molar ratio of 4:1.[9,14] Following separation by flash chromatography, the β-configuration was unambiguously assigned based on 2D NOE NMR experiments. The reduction of the nitro group was accomplished under mild conditions with $Fe_3(CO)_{12}$ and methanol,[15] and the resulting free amine subsequently N-substituted through reaction with *n*-butylisocyanate in THF.[8] After deprotection under alkaline conditions, the nucleoside analog **D4** was 5′-tritylated using extended reaction times of 18 h at 40 °C.[16,17] Following conversion to its phosphoramidite, **6** was directly used as synthon for the synthesis of the oligonucleotide **TFO15** using the standard β-cyanoethyl phosphoramidite method. The 15-mer oligonucleotide containing the nucleoside analog was finally characterized by MALDI-TOF mass spectrometry, giving a mass peak at *m/z* 4584.9 within experimental error of the calculated molecular weight of **TFO15** (*m/z* 4588.1).

In order to examine the affinity and selectivity toward a duplex target, equimolar amounts of four AT-rich double-helical oligonucleotides were mixed with the **D4** containing oligonucleotide **TFO15**, which was designed to bind in a parallel motif to the complementary duplex. Sequences of the 15-mer double-helical targets differ in their Watson–Crick base pair XY located opposite the non-natural **D4** analog in a triplex:

TFO15:	5′-	C	T	T	C	**D4**	T	T	T	T	T	C	T	T	T	T	-3′
	5′-	G	A	A	G	**X**	A	A	A	A	A	G	A	A	A	A	-3′
	3′-	C	T	T	C	**Y**	T	T	T	T	T	C	T	T	T	T	-5′

In addition, thermal stabilities of corresponding triplexes with the internal **D4**·XY triplet substituted for canonical C^+·GC and T·AT as well as G·TA and T·CG triads were studied for comparison. Although less stable than C^+·GC and T·AT, G·TA and T·CG base triads have been found to be the best combinations for recognizing TA and CG by natural bases.[18,19] A typical UV melting experiment at pH 7.0 on a 1:1 mixture of duplex and TFO with **D4** opposite CG (**D4**·CG triplex) is shown in Figure 2. Each mixture displays a sigmoidal high-temperature transition which corresponds to the duplex to single strand transition. Except for the C^+·GC and T·AT triplexes that only exhibit one resolved transition at lower pH, additional hyperchromicity effects indicate the dissociation of the TFO from the duplex at lower temperatures.

Melting temperatures determined for the triplexes at pH 6.0, 6.5, and 7.0 are summarized in Table 1. Due to the presence of at least three protonated C^+·GC base triplets within the triplexes, thermal stabilities decrease with increasing pH as expected. Having an additional C^+·GC base triplet, the canonical C^+·GC triplex exhibits the largest drop (23 °C) in triplex melting temperature T_m upon increasing the pH from 6.0 to 7.0. Interestingly, however, triplexes carrying the **D4** analog and in particular triplexes with **D4**·GC and **D4**·AT triads are less sensitive to pH, and triplex melting temperatures only decrease by about 2 °C when going from pH 6 to 7. Obviously, the non-natural base triplet counteracts the regular pH dependence of the triplex indicating that an unprotonated imidazole base in the **D4** analog is strongly



Scheme 1. Reagents and conditions: (i) 2-deoxy-3,5-di-*O*-*p*-toluoyl-*a*-D-ribofuranosyl chloride, NaH, CH_3CN, 43%; (ii) $Fe_3(CO)_{12}$, toluene, 63%; (iii) C_4H_9NCO, THF, 79%; (iv) 1% NaOH in MeOH, 80%; (v) DMTCl, DMAP, Et_3N, pyridine, 56%; (vi) $[(CH_3)_2CH]_2NP(Cl)O(CH_2)_2CN$, *i*-$Pr_2EtN$, CH_2Cl_2.

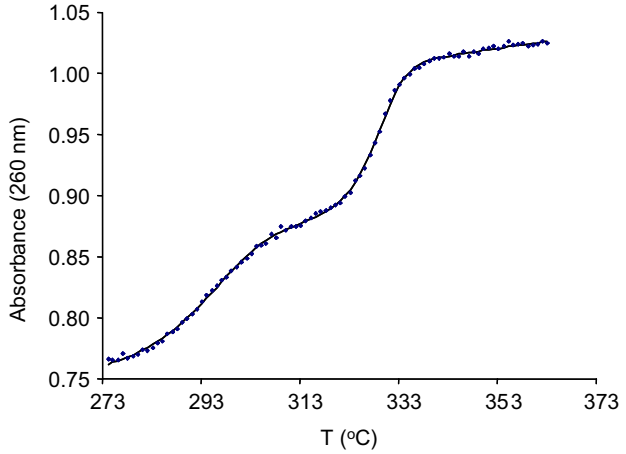


Figure 2. Temperature-dependent absorbance at 260 nm of the **D4**·CG triplex; the solid line corresponds to the fitted curve. [Triplex] = 2.5 μM, 100 mM NaCl, 20 mM sodium cacodylate, 20 mM $MgCl_2$, pH 7.

Table 1
Summary of triplex–duplex melting temperatures T_m (°C) of triplexes with different Z·XY base triplets

Z·XY base triplet	T_m (°C), pH 6.0[a]	T_m (°C), pH 6.5[a]	T_m (°C), pH 7.0[b]
C^+·GC	49.0[f]	35.7[d]	25.7[c]
T·AT	48.4[f]	49.3[f]	31.6[e]
G·TA	29.6[c]	25.7[d]	20.1[d]
T·CG	29.6[d]	21.4[d]	18.6[e]
D4·CG	29.0[c]	23.8[d]	21.8[e]
D4·TA	27.3[c]	23.0[d]	18.6[e]
D4·GC	26.2[d]	25.1[e]	24.9[e]
D4·AT	27.5[d]	27.1[e]	25.1[e]

[a] [Triplex] = 2.5 μM, 100 mM NaCl, 20 mM NaAc, 20 mM $MgCl_2$.
[b] [Triplex] = 2.5 μM, 100 mM NaCl, 20 mM sodium cacodylate, 20 mM $MgCl_2$.
[c] Uncertainty ± 0.5 °C.
[d] Uncertainty ± 1 °C.
[e] Uncertainty ± 2 °C.
[f] Only one transition from triplex to single strands is observed.

favored upon triplex formation especially when opposite a GC or AT base pair.

All **D4** triplexes fail to achieve thermal stabilities of T_m >30 °C observed for the canonical C^+·GC and T·AT triplexes under favorable conditions. Thus, the **D4**·CG triplex being the most stable non-natural triplex at low pH exhibits a T_m of 29 °C at pH 6 comparable to the melting of the G·TA and T·CG triplexes under equal conditions. On the other hand and as a result of their small pH dependence, the **D4**·GC and **D4**·AT triplexes exhibit the highest triplex melting temperature of all **D4** containing triple helices at pH 7 with a T_m of about 25 °C and compare favorably with the lower melting G·TA and T·CG triplexes as well as with the canonical C^+·GC triplex that shows about the same thermal stability at physiological pH.

It has to be noted that there is no single most stable **D4** triplex within a pH range of $6 \leqslant pH \leqslant 7$. Rather, relative stabilities of **D4** triplexes vary according to their different pH dependence, but selectivities toward the four target base pairs are rather poor at any given pH value. Thus, only a maximum difference in T_m of 6.5 °C is observed between the most stable **D4**·AT and the least stable **D4**·TA triplex at pH 7. Consequently, the **D4** analog must be regarded as a universal base binding opposite all four base pairs with moderately high affinity at physiological pH.

Thermodynamic parameters for triplex formation at pH 7 were determined by fitting the melting curves with a non-linear least-squares program. The analysis is based on a two-state model and assumes that extinction coefficients of the individual DNA species are linear functions of temperature and enthalpy ΔH as well as entropy ΔS of triplex formation are constant over the temperature range of triplex melting.[20] As a consequence of the two-state approximation employed, a reliable fitting is restricted to UV melting curves with two clearly separated transitions for triplex and duplex melting. Even with a rather conservative evaluation of the fitted parameters to possibly avoid their overinterpretation considering the various simplifications of the analysis, thermodynamic data from the curve fitting as summarized in Table 2 provide for obvious trends in the thermodynamics of third strand binding with the non-natural TFOs. With a ΔH of 320–340 kJ/mol, that is, with 21–23 kJ/mol per base, the van't Hoff enthalpy change of triplex formation for the two canonical C^+·GC and T·AT triplexes falls within the range determined from previous UV melting experiments on DNA triplexes.[21] Although still driven by a negative enthalpy change, third strand association with a **D4** analog is accompanied by less exothermic binding amounting to only −158 kJ/mol for the **D4**·AT triplex up to −191 kJ/mol for the **D4**·TA triplex. However, as seen from Table 2, such a less favorable enthalpic contribution to triplex formation is partially balanced by a significantly less unfavorable negative entropy change determined for all **D4** triplexes. Such thermodynamic profiles indicate that the **D4** analog only possesses a limited number of strong and specific interactions with the duplex target. Also, binding through the hydrophilic ureido functionality may be enthalpically compromised by the need of its initial desolvation. On the other hand, less specific interactions are expected to result in an enhanced conformational flexibility associated with a less negative entropy change. Also, a major driving force for the **D4** association when compared to natural bases may arise from the release of water from the hydrophobic surface of the analog upon binding and possibly from additional release of structural water upon accommodating the *N*-butyl substituent within the DNA major groove. The poor base pair discrimination exhibited previously by a corresponding *N*-alkylureido-substituted isoindolin-1-one receptor within a triple-helical complex, being in contrast to its ability for effective complexation of a free CG base pair in organic solvents, may likewise originate from similar enthalpic and entropic contributions to binding in an aqueous medium.[16]

Table 2
Thermodynamic parameters for the triplex–duplex transition from the fitting of UV melting curves obtained from triplexes with different Z·XY base triplets at pH 7

Z·XY base triplet	T_m[a] (°C)	ΔH (kJ/mol)	ΔS (J/mol K)	ΔG[b] (kJ/mol)
C^+·GC	26.8	−339	−1018	−35.6
T·AT	31.3	−322	−947	−39.8
D4·CG	22.3	−169	−460	−31.9
D4·TA	19.4	−191	−541	−29.8
D4·GC	23.9	−188	−521	−32.7
D4·AT	24.2	−158	−420	−32.8

[a] From curve fitting.
[b] At 25 °C.

In summary, our thermodynamic data on triplex formation with the **D4** containing TFO in aqueous solution indicate less specific binding of the analog with a major driving force arising from hydrophobic effects. As a consequence, moderately strong binding affinities toward all possible base pairs with a poor base pair discrimination are expected and observed based on triplex thermal stabilities. This is in contrast to our previous model system of free nucleosides in aprotic solvents showing the formation of a well-defined hydrogen-bonded **D4**·CG base triplet closely isomorphous to the canonical base triads, and highlights the importance of additional contributions to binding within an aqueous environment. For the future design of novel nucleoside surrogates,

thermodynamic data together with structural information will be indispensible in achieving a more predictable base pair recognition accompanied by efficient complexation.

Acknowledgment

We thank the Deutsche Forschungsgemeinschaft, Bonn, for financial support.

References and notes

1. Duval-Valentin, G.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 504–508.
2. Kim, H.-G.; Miller, D. M. *Biochemistry* **1998**, *37*, 2666–2672.
3. Havre, P. A.; Gunther, E. J.; Gasparro, F. P.; Glazer, P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7879–7883.
4. Jones, W. R.; Stone, M. P. *Nucleic Acids Res.* **1998**, *26*, 1070–1075.
5. Dervan, P. B. *Nature* **1992**, *359*, 87–88.
6. Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639–1642.
7. (a) Guianvarc'h, D.; Fourrey, J. L.; Maurisse, R.; Sun, H. S.; Benhida, R. *Bioorg. Med. Chem.* **2003**, *11*, 2751–2759; (b) Buchini, S.; Leumann, C. J. *Angew. Chem., Int. Ed.* **2004**, *43*, 3925–3928; (c) Eldrup, A. B.; Nielson, P. E. *J. Am. Chem. Soc.* **1997**, *119*, 11116–11117; (d) Rusling, D. A.; Powers, V. E. C.; Ranasinghe, R. T.; Wang, Y.; Osborne, S. D.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2005**, *33*, 3025–3032; for reviews, see: (e) Gowers, D. M.; Fox, K. R. *Nucleic Acids Res.* **1999**, *27*, 1569–1577; (f) Doronina, S. O.; Behr, J.-P. *Chem. Soc. Rev.* **1997**, 63–71; (g) Purwanto, M. G. M.; Weisz, K. *Curr. Org. Chem.* **2003**, *7*, 427–446.
8. Wang, W.; Purwanto, M. G. M.; Weisz, K. *Org. Biomol. Chem.* **2004**, *2*, 1194–1198.
9. Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976–7982.
10. Zimmerman, S. C.; Schmitt, P. *J. Am. Chem. Soc.* **1995**, *117*, 10769–10770.
11. Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908–7909.
12. Bredereck, H.; Theilig, G. *Chem. Ber.* **1953**, *86*, 88–96.
13. (a) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777–2781; (b) Rolland, V.; Kotera, M.; Lhomme, J. *Synth. Commun.* **1997**, *27*, 3505–3511.
14. Kazimierczuk, Z.; Cottam, H. B.; Revankar, G. R.; Robins, R. K. *J. Am. Chem. Soc.* **1984**, *106*, 6379–6382.
15. Landesberg, J. M.; Katz, L.; Olsen, C. *J. Org. Chem.* **1972**, *37*, 930–936.
16. Mertz, E.; Mattei, S.; Zimmerman, S. C. *Bioorg. Med. Chem.* **2004**, *12*, 1517–1526.
17. *Spectroscopic data of* **5**: ^{1}H NMR (600 MHz, 293 K, $CDCl_3$): δ (ppm) 0.94 (t, 3H, CH_3), 1.37 (m, 2H, CH_2), 1.50 (m, 2H, CH_2), 2.45 (m, 2H, H2′, H2″), 3.25 (m, 2H, CH_2), 3.37 (m, 2H, H5′, H5″), 3.79 (s, 6H, OCH3), 4.18 (m, 1H, H4′), 4.59 (m, 1H, H3′), 5.98 (t, J = 6.5 Hz, 1H, H1′), 6.85 (dd, J_1 = 8.7, J_2 = 8.9 Hz, 4H, ArH), 7.20–7.40 (m, 10H, ArH), 7.46 (m, 2H, ArH), 7.59 (m, 2H, ArH), 7.68 (s, 1H, ArH).
18. Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967–971.
19. Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weiss, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3840–3844.
20. Longfellow, C. E.; Kierzek, R.; Turner, D. H. *Biochemistry* **1990**, *29*, 278–285.
21. Kamiya, M.; Torigoe, H.; Shindo, H.; Sarai, A. *J. Am. Chem. Soc.* **1996**, *118*, 4532–4538.