Pergamon

Tetrahedron Letters 40 (1999) 8085–8088

TETRAHEDRON LETTERS

# NMR recognition studies of C·G base pairs by new easily accessible heterobicyclic systems

Florence Lecubin,[a] Rachid Benhida,[a,*] Jean-Louis Fourrey [a] and Jian-Sheng Sun [b]

[a] *Institut de Chimie des Substances Naturelles, CNRS, 91198 Gif-sur-Yvette, France*

[b] *Laboratoire de Biophysique, MNHN, 43 rue Cuvier, 75231 Paris Cedex 05, France*

Received 25 June 1999; accepted 31 August 1999

**Abstract**

Recognition of DNA duplexes by triplex forming oligonucleotides (TFO) is limited to DNA homopurine sequences. As a first step to overcome this limitation we report here NMR recognition studies of the C·G base pair by new heterocyclic systems, derived from benzimidazole and benzoxazole units bearing an urea donor moiety, designed to bound to the 4-amino group of the cytosine and the O4- and N7-atoms of the guanosine bases, respectively. 

*Keywords:* heterocyclic base; hydrogen bond; Watson–Crick base pair; triple helix.

As a result of their selective binding within the major groove of DNA duplexes (ds-DNA), triplex forming oligonucleotides (TFOs) form stable local triple-helical structures.[1] Interestingly, in a biological context, TFOs can compete with proteins for ds-DNA recognition. Accordingly, such a property gives them the capacity to interfere with gene expression. These observations prompted the development of the antigene strategy which proposes to apply TFOs as new tools for artificial gene repression in gene related diseases.

Unfortunately, this potential therapeutic approach has a major intrinsic limitation since triplex formation is restricted to oligopurine·oligopyrimidine DNA targets. Accordingly, in the past few years many efforts have been undertaken to extend the recognition pattern of TFOs to polypurine sequences which are interrupted by a pyrimidine.[2] In continuation of our studies directed at overcoming the physicochemical limitations of TFO applications,[3] we proposed to design new heterocyclic motifs which are able: (i) to make hydrogen bonds with a C·G Watson–Crick base pair; and (ii) to be easily incorporated into TFOs (Fig. 1). For the selection of these heterocyclic systems the NMR method which revealed hydrogen bonding contacts, as recently exploited by Zimmerman in the case of 3-butyl-2-methyl-8-(*N'*-*n*-butylureido)naphth[1,2-*d*]imidazole (ZB), proved to be well adapted.[4]

Herein, we show, on the basis of such NMR studies in an aprotic solvent, that the three easily chemically accessible heterocyclic bases designed HBi (with i=1, 2 and 3) manifest C·G base pair

* Corresponding author. Fax: 01 69 07 72 47; e-mail: benhida@icsn.cnrs-gif.fr

0040-4039/99/$ - see front matter 
*PII:* S0040-4039(99)01668-8

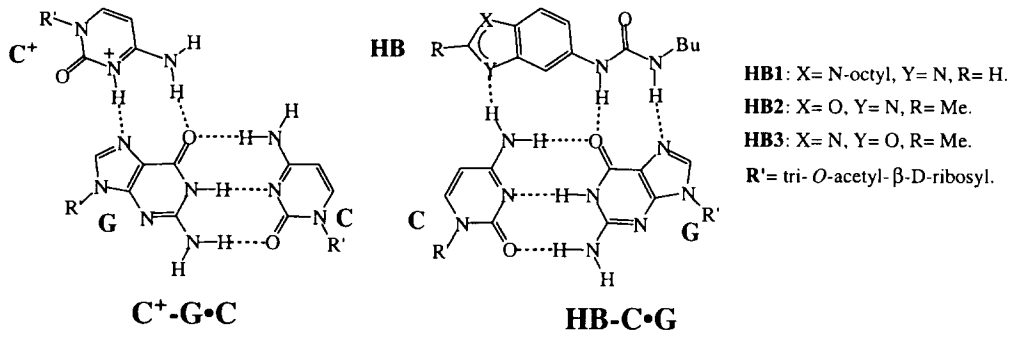
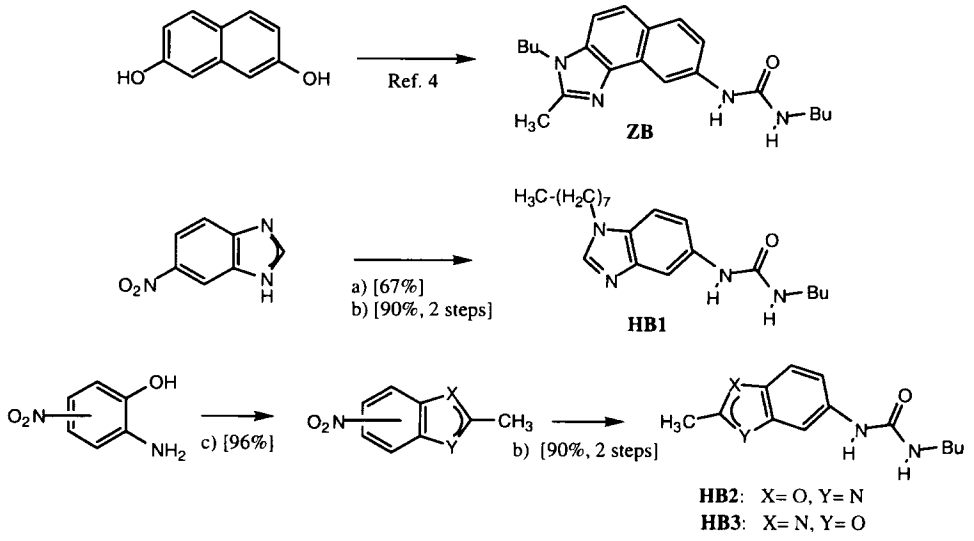


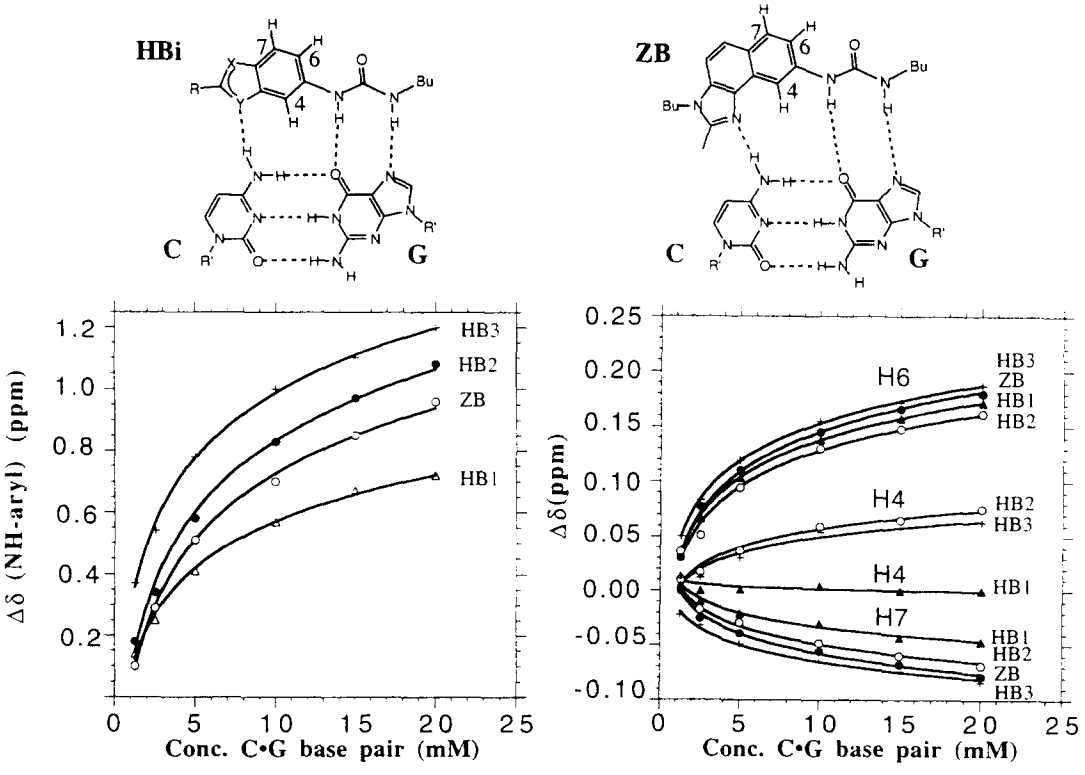

Figure 1. Hydrogen bonding within $C^+$-G·C and HB-C·G triplexes

recognition (Scheme 1). Namely, the benzimidazole and benzoxazole derivatives HB1, HB2 and HB3 have been synthesized as outlined in Scheme 1 together with the Zimmerman base (ZB)[4] for comparison. Tertiary complex formation studies between HBi and the C·G base pair[5] were performed in $CDCl_3$ by: (i) direct titration of a 1:1 mixture of the C·G base pair (2.5 mM) with heterocyclic bases HBi (1 to 20 mM); (ii) reverse titration of HBi (2.5 mM) with equimolecular mixtures of the C·G base pair (1 to 20 mM); and (iii) equimolar titration of the complex HBi–C·G with increasing concentrations.




Scheme 1. Synthesis of the benzimidazolyl- and benzoxazolylurea systems (HBi). Conditions: (a) $I(CH_2)_7CH_3$, NaH, THF; (b) i. $H_2$–Pd/C, EtOH, ii. *N*-butylisocyanate, NaH, DMF; (c) $Ac_2O$, then *p*TSA (0. 2 equiv.), toluene, 100°C

Compared to their respective controls (Fig. 2), during the titration of each heterocyclic base (HBi) (2.5 mM) with increasing concentrations of C·G base pair (1 to 20 mM), we have observed significant low field displacements of the urea protons. The corresponding chemical shift variations were $\Delta\delta$=0.72 ppm for HB1, 1.08 ppm for HB2 and 1.20 ppm for HB3, respectively (20 mM). This suggests that the urea motif undergoes hydrogen bonding to guanine O6 and N7 positions. It is noteworthy that the affinity is higher in the cases of the benzoxazole derivatives (HB2 and HB3) compared to the benzimidazole (HB1). This observation clearly indicates that the five membered ring is implicated in the complexation. Upon direct titration of the C·G base pair (2.5 mM) by increasing the concentration of HBi (1.25 to 20 mM), very little variations were noticed for the displacements of the urea exchangeable protons. This finding supports the formation of the ternary complex even at low concentration. We have also observed the same




**a.** Curves of the chemical shift variations (Δδ) of NH aryl urea protons in ternary complexes (HB: 2.5 mM) as a function of CG base pair concentration.

**b.** Curves of the chemical shift variations (Δδ) of aromatic protons H6 and H7 (HB: 2.5 mM) as a function of CG base pair concentration.

Figure 2. NMR complexation studies of ternary complexes HBi–C·G

significant downfield shift for the urea protons during the equimolar titration. Finally, confirmation of the triad existence was brought by observing the chemical shift variations of the aromatic protons H4, H6 and H7 of HBi (i=1–3) upon reverse and equimolar titrations.[6] In the three cases the corresponding H6 proton was deshielded by $\delta\Delta$=0.18 ppm while H7 underwent a weak upfield shift. For H6, the significant deshielding is probably due to a carbonyl anisotropic effect which is suggestive of a blocked conformation of the aromatic system by means of hydrogen bonding as indicated in Fig. 2. Since, the data reported by Zimmerman[4] suggested a stronger interaction of ZB (Scheme 1) with the C·G base pair, we decided to evaluate its behaviour under our experimental conditions. We noticed that the NMR titration experiments with ZB gave a $\delta\Delta$=0.98 ppm for the urea protons and approximatively the same values for the aromatic proton H6 as observed with the three HB compounds. These data which are in line with our observations with HBi favour a very similar mode of recognition.

In conclusion, we have designed three novel and very accessible heterobicyclic systems for recognition of C·G base pairs. As shown by NMR studies, performed in an aprotic solvent, their mode of association with the corresponding base pair involves the formation of three hydrogen bonds. This behaviour is in agreement with the one previously reported by Zimmerman[4] with a different heterocycle. It is noteworthy that compared to the bicyclic HB systems, ZB is tricyclic and, accordingly, its use in a triplex context could be problematic since it might prefer to intercalate than to exchange hydrogen bonds with a C·G base pair.[2d] To complete this work, it remains for us to determine the best way to incorporate these

new heterocyclic bases within TFOs. To solve this problem, various oligonucleotides containing bases HBi are currently being synthesized in this laboratory. They will be annealed to DNA duplexes having a polypurine strand interrupted by one or more deoxycytidine(s).[7]

## Acknowledgements

We thank the Ligue Nationale de Recherche sur le Cancer for financial support (F.L.) and Dr. Eric Guittet for stimulating discussions.

## References

1. Reviews: (a) Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666–690. (b) Frank-Kamenetskii, M. D.; Mirkin, S. M. *Annu. Rev. Biochem.* **1995**, *64*, 65–95. (c) Doronina, S. O.; Behr; J.-P. *Chem. Soc. Rev.* **1997**, 63–71. (d) Neidle, S. *Anti-Cancer Drug Design* **1997**, *12*, 433–442. (e) Plum, G. E. *Biopoly.* **1997**, *44*, 241–255. (f) Gowers, D. M.; Fox, K. R. *Nucleic Acids Res.* **1999**, *27*, 1569–1577.
2. (a) Huang, C.-Y.; Cushman, C. D.; Miller, P. S. *J. Org. Chem.* **1993**, *58*, 5048–5049. (b) Sasaki, S.; Nakashima, S.; Nagatsugi, F.; Tanaka, Y.; Hisatome, M.; Maeda, M. *Tetrahedron Lett.* **1995**, *36*, 9521–9524. (c) Huang, C.-Y.; Bi, G.; Miller, P. S. *Nucleic Acids Res.* **1996**, *24*, 2606–2613. (d) Lehmann, T. E.; Greenberg, W. A.; Liberles, D. A.; Wada, C. K.; Dervan, P. B. *Helv. Chim. Acta.* **1997**, *80*, 2002–2022 and references cited therein.
3. (a) Benhida, R.; Devys, M.; Fourrey, J.-L.; Lecubin, F.; Sun, J.-S. *Tetrahedron Lett.* **1998**, *39*, 6167–6170. (b) Bédu, E.; Benhida, R.; Devys, M.; Fourrey, J.-L. *Tetrahedron Lett.* **1999**, *40*, 835–838.
4. Zimmerman, S. C.; Schmitt, P. *J. Am. Chem. Soc.* **1995**, *117*, 10769–10770.
5. Throughout this work the C·G base pairs consisted of equimolecular mixtures of tri-*O*-acetylcytidine and tri-*O*-acetylguanosine. A lowfield shift of the imino and amino protons of guanosine ($\Delta\delta$=1.3 to 1.6 ppm) was observed by increasing the concentration (0.25 to 25 mM) of the C·G base pair whereas the signal of the cytidine amino protons was only observable at low concentrations.
6. Since the curve corresponding to H4 of ZB showed the same profile as the one of H7 of HB1, it was omitted in Fig. 2b for clarity.
7. Analytical and spectral data for final products: **HB1**: mp (AcOEt) 132–134°C. $^1$H NMR ($CDCl_3$, 250 MHz) 7.84 (s, 1H, H-2); 7.78 (s, 1H, NHPh); 7.58 (d, 1H, J=1.7 Hz, H4); 7.44 (dd, 1H, J=1.7 and 8.6 Hz, H6); 7.24 (d, 1H, J=8.6 Hz, H7); 5.60 (t, 1H, J=6.2 Hz, NHBu); 4.08 (t, 2H, J=7.5 Hz, $NCH_2$); 3.19 (m, 2H, $CH_2NH$); 1.85 (m, 2H, $CH_2$-); 1.35 (m, 14H, $7CH_2$); 0.85 (2t, 6H, J=7.1 Hz, $2\times CH_3$). $^{13}$C NMR ($CDCl_3$) 157.2, 143.7, 142.9, 134.7, 130.0, 117.6, 111.1, 109.5, 45.0, 39.7, 32.2, 31.6, 29.6, 28.9, 26.6, 22.4, 19.9, 13.9, 13.6. MS (FAB+) *m/z* 345 $(MH)^+$. Anal. calcd for $C_{20}H_{32}N_4O$: C 69.73; H 9.36; N 16.26.; found: C 70.12; H 9.54; N 15.87. **HB2**: mp (AcOEt) 160–162°C. $^1$H NMR ($CDCl_3$, 250 MHz) 7.9 (s, 1H, NHPh), 7.43 (d, 1H, J=2 Hz, H4), 7.11 (d, 1H, J=8.7 Hz, H7), 7.01 (dd, 1H, J=2.0 and 8.7 Hz, H6), 5.82 (t, 1H, J=6.3 Hz, NH), 3.04 (m, 2H, $CH_2N$), 2.48 (s, 3H, $CH_3$), 1.27 (m, 2H, $CH_2$), 1.15 (m, 2H, $CH_2$), 0.74 (t, 3H, J=7.1 Hz, $CH_3$). $^{13}$C NMR ($CDCl_3$) 164.3, 157.2, 147.1, 141.7, 135.8, 118.2, 111.5, 109.7, 39.9, 32.2, 20.0, 14.4, 13.7. MS (IC, *i*BuH.) *m/z* 248 $(MH)^+$. Anal. calcd for $C_{13}H_{17}N_3O_2$: C 63.14; H 6.93; N 16.99; found C 62.76; H 7.11; N 16.49. **HB3**: mp (AcOEt) 133–135°C. $^1$H NMR ($CDCl_3$, 250 MHz) 8.36 (s, 1H, NHPh), 7.65 (d, 1H, J=1.8 Hz, H7), 7.27 (d, 1H, J=8.5 Hz, H4), 6.84 (dd, 1H, J=1.8 and 8.5 Hz, H5), 6.10 (t, 1H, 6.3 Hz, NH), 3.07 (m, 2H, $CH_2N$), 2.45 (s, 3H, $CH_3$), 1.30 (m, 2H, $CH_2$), 1.18 (m, 2H, $CH_2$), 0.75 (t, 3H, J=7.1 Hz, $CH_3$). $^{13}$C NMR ($CDCl_3$) 163.2, 156.9, 151.4, 136.8, 136.4, 118.7, 116.5, 102.0, 39.8, 32.2, 20.0, 14.4, 13.7. MS (IC, *i*BuH.) *m/z* 248 $(MH)^+$. HRMS (IC) calcd for $C_{13}H_{17}N_3O_2$ 248.1398; found 248.1401.