Pergamon

Tetrahedron Letters 41 (2000) 5571–5575

TETRAHEDRON LETTERS

Facile synthesis of oligopeptide distamycin analogs devoid of hydrogen bond donors or acceptors at the N-terminus: sequence-specific duplex DNA binding as a function of peptide chain length

Santanu Bhattacharya* and Mini Thomas

Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, India

Received 13 March 2000; revised 2 May 2000; accepted 16 May 2000

Abstract

The first examples of distamycin analogs, which lack hydrogen bond interactor groups at the N-terminus, have been synthesized. The bispyrrole peptide did not exhibit any detectable binding with double-stranded (ds) DNA. However, all other homologues did bind to ds-DNA strongly, with the binding affinities increasing as a function of the number of repeating pyrrole carboxamide units, implying that a hydrogen bond donor or acceptor atom per se at the N-terminus is not essential for their DNA binding. Studies with poly d(GC) showed that the N-terminal formamide is not a prerequisite for GC binding, contrary to earlier postulations. © 2000 Elsevier Science Ltd. All rights reserved.

The design of synthetic organic molecules of low molecular mass that can regulate gene expression in a predictable manner is an area of current research interest.[1] Despite the fact that there are several minor groove binders such as Hoechst, DAPI, etc. the natural product distamycin (Dst) remains an ideal candidate for further elaboration of the central theme of designing synthetic gene-regulators. This is due to the simplicity of its chemical nature, consisting of a hetero-aromatic polyamide backbone, with a single positive charge emanating from the side chain at the C-terminus. A number of fundamental questions pertaining to the DNA recognition properties of Dst derivatives remain unanswered. For instance, is there a minimum size requirement for the onset of DNA binding? Can Dst analogs bind ds-DNA effectively in the absence of an N-terminus hydrogen bond donor or acceptor? Can one improve the intracellular stability of these compounds when the N-terminal formamide is absent? This is significant on account of the inherent susceptibility of the terminal amide group of distamycin towards cellular degradation. It is noteworthy that thioformyl distamycin, an amide isostere of distamycin, was synthesized by Lown et

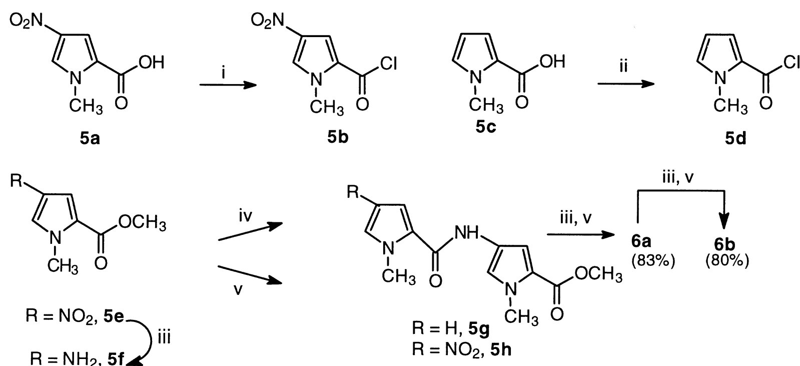
* Corresponding author. E-mail: sb@orgchem.iisc.ernet.in

0040-4039/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0040-4039(00)00802-9

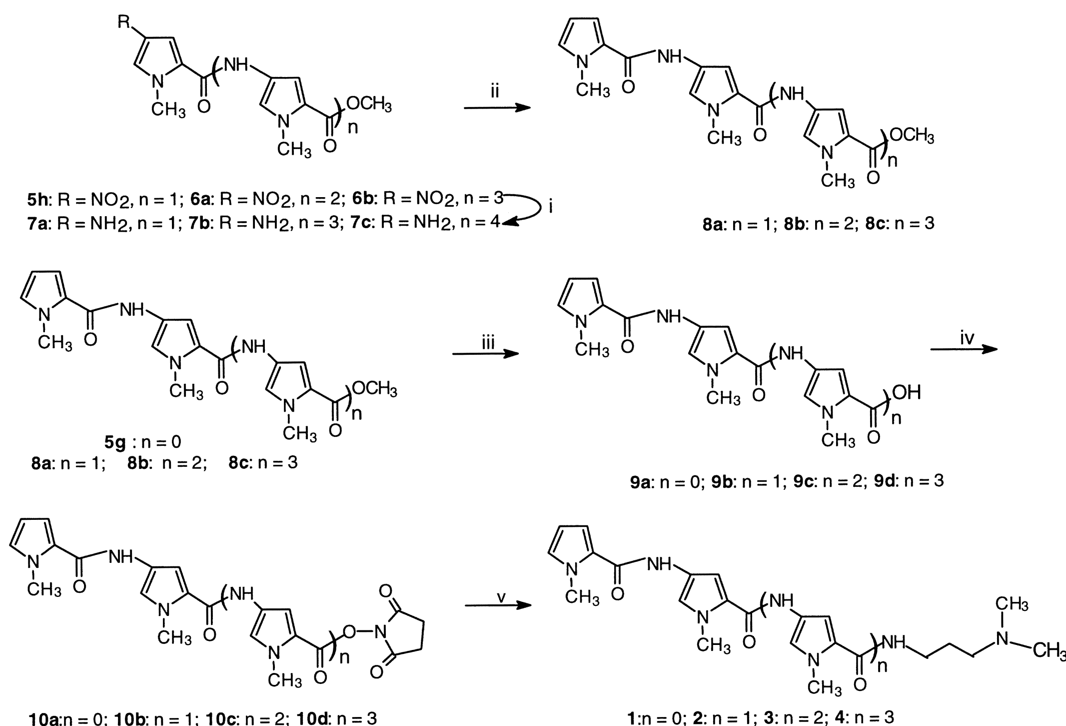
al. on account of the above considerations.[2] Design of appropriate Dst analogs would also be useful to test the validity of the hypothesis that the N-terminus formamide group is important for the GC binding mode.[3]

However, the seemingly simple synthesis of Dst and its derivatives has been complicated by the low reactivity of the aromatic carboxylic acids and the extremely unstable nature of the amines.[4] To circumvent some of these difficulties, a solid-phase-based method was recently developed by Dervan et al. toward the synthesis of higher homologues of distamycin.[5] We felt that a viable solution-phase method would still be quite desirable, as it would be complimentary and would offer considerable general applicability, especially when other bio-active moieties such as chemical nucleases,[1c,6] need to be appended. Herein, we present the convenient synthesis and results of the DNA binding experiments of the first examples of Dst class of minor groove binders that are devoid of a hydrogen bond donor or an acceptor group at the N-terminus.

The key starting material for the synthesis, **5e**, was obtained by the nitration and subsequent esterification of commercially available *N*-methyl pyrrole 2-carboxylic acid, **5c** according to a published procedure.[4b] This was then reduced using H_2-Pd/C in DMF to obtain the free amine **5f**, which was coupled with the acid chloride, **5b**, in the presence of Et_3N in DMF to obtain the nitro-dipeptide analog **5h** in 87% yield (Scheme 1). Further, elongation of the peptide backbone was achieved by repetition of the reduction and the amide coupling steps. Thus, the nitro derivatives **6a** and **6b** were obtained in 83 and 80% yields, respectively, from their precursors as shown in Scheme 1. The chain growth was terminated upon coupling with *N*-methyl pyrrole-2-carbonyl chloride **5d**, once the required length of the polyamide chain was achieved. All the nitro compounds **5e**, **5h**, **6a** and **6b** were reduced separately and coupled with the acid chloride, **5d**, to obtain **5g**, **8a**, **8b** and **8c** respectively, in ca. 80% yields as illustrated in Schemes 1 and 2. It is important to note that none of the free amines were isolated for characterization, on account of their unstable nature. They were separated from Pd/C in the reaction mixture by filtration and were coupled with the appropriate acid chlorides directly. We term the latter step as 'end-capping' as it prevents further extension of the amide chain from the N-terminus. The 'end-capped' methyl esters were hydrolyzed to the corresponding acids, which were subsequently converted to the activated esters using *N*-hydroxy succinimide (HOSU) and dicyclohexyl carbodiimide (DCC) (Scheme 2). DCU was removed from the HOSU esters by filtration and were subjected to aminolysis with *N*,*N*-dimethyl-1,3-diaminopropane in $CHCl_3$ to obtain the final compounds, **1**–**4** in 80, 88, 90 and



Scheme 1. *Reagents, conditions and yields*: (i) $SOCl_2$, THF, reflux, 1 h; (ii) $SOCl_2$, THF, 0°C 1 h, then rt, 30 min; (iii) H_2-Pd/C (5%), 1 atm. rt, 12 h; (iv) **5d**, Et_3N, –5°C, 1 h, then 1 h at rt (85%); (v) **5b**, Et_3N, –5°C, 1 h, then 1 h at rt (87%)



Scheme 2. *Reagents, conditions and yields*: (i) H_2-Pd/C(5%), 1 atm. rt, 18 h; (ii) **1d**, Et_3N, –5°C, 1 h, then 1 h at rt (80, 78 and 85% for **8a**–**8c**, respectively); (iii) 0.25 (N) NaOH, EtOH/H_2O, reflux, 1 h, then 0°C, 0.5 (N) HCl (96, 94, 92 and 90% for **9a**–**9d**, respectively); (iv) DCC, HOSU, DMF, 0°C, 30 min, then rt, 4 h; (v) *N*,*N*-dimethyl-1,3-diaminopropane, rt, 2 h (80, 88, 90 and 82% for **1**–**4**, respectively)

82% yields, respectively.[7] It is noteworthy that the *N*,*N*-dimethyl-1,3-diaminopropionamido group was introduced at the C-terminus only in the last step of the synthesis since the amines produced at the N-terminus upon reduction from their corresponding nitro derivatives were relatively more unstable when the *N*,*N*-dimethyl-1,3-diaminopionamido group was already present at the C-terminus. Column chromatography was employed only for the purification of the 'end-capped' methyl esters and the final compounds.

The DNA binding abilities of these oligopeptides were assessed by thermal helix-to-coil 'melting' temperature (T_m) measurements (pH = 7.4, 10 mM tris buffer, 20 mM NaCl, D/P = 0.5). Compound **1** did not alter the T_m of ds-DNA. However, the ΔT_m increased progressively as a function of the peptide length for the analogs **2**–**4**. The ΔT_m values of the complexes of **2**–**4** and poly d (AT) were ca. 6.1, 17.2 and 26.6°C, respectively. With CT (calf thymus) DNA, the ΔT_m were 1.5, 4.5 and 8°C, respectively, for **2**–**4** at the same [D]/[P] ratio suggesting that these compounds have retained their AT specificity despite the removal of the N-terminal amide unit. This was further corroborated by the ethidium bromide displacement assay.[8] The apparent binding constants (K_{app}) obtained for **2**–**4** with CT DNA were $\sim 5.8\times10^4$, 1×10^6 and 4.5×10^6 M^{-1}, respectively, whereas the corresponding values for poly d (AT) were 4.2×10^5, 3.3×10^7 and 1.2×10^8 M^{-1}, respectively. The K_{app} values obtained for Dst under the same conditions were 7.7×10^5 and 3.5×10^7 M^{-1} for CT DNA and poly d (AT) respectively,[8] suggesting that the compounds **3** and **4** bind to ds-DNA with greater affinity than their natural counterpart, Dst, apart from preserving the sequence-specificity.

Experiments with **3**/poly d (GC) employing circular dichroism showed intense ICD signals and thus suggested that the N-terminal formamide is not a prerequisite for GC binding, as was postulated earlier. This complex was highly electrostatic in nature, as the ICD signal disappeared completely in ca. 200 mM (NaCl).

In summary, a convenient, general and adaptable method has been developed for the synthesis of four Dst analogs, which form the first examples of Dst class of compounds that bear neither a hydrogen bond donor nor an acceptor at the N-terminus. Our methodology offers several advantages over the existing procedures. First, the total number of steps in the synthetic scheme is kept at a minimum, as each intermediate nitro compound is the precursor for the next higher analog. Secondly one of the least yielding steps in the synthesis of Dst, viz. the introduction of the terminal formamide group has been completely avoided. Thirdly, this modification, along with the introduction of the C-terminus side chain only in the last step, permit the synthetic procedure to have enough flexibility such that it would accommodate the introduction of natural amino acids, etc. in the sequence, if desired, for further modulation of the DNA binding properties. It is also demonstrated that the presence of a hydrogen bond donor or acceptor at the N-terminus per se, as in the case of Dst and imidazole/pyrrole oligopeptide analogs, respectively, is not a prerequisite for the maintenance of DNA binding. The minimum number of pyrrole carboxamide units for the onset of DNA binding, in the absence of the N-terminus amide is shown to be three. It is also shown that the N-terminus formamide is not a prerequisite for GC binding. Work is now underway in our laboratory toward the synthesis of novel minor groove binders with greater affinity and longer sequence recognition properties.

Acknowledgements

This work was supported by the Swarnajayanti Fellowship Grant of the Department of Science and Technology, Government of India.

References

1. (a) See, for reviews: Dervan, P. B. *Science* **1986**, *232*, 464. (b) Geierstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct*. **1995**, *24*, 463. (c) Bailly, C.; Chaires, J. B. *Bioconjugate Chem*. **1998**, *9*, 513. (d) Iida, H.; Jia, G.; Lown, J. W. *Curr. Opin. Biotechnology*. **1999**, *10*, 29.
2. Singh, M. P.; Kumar, S.; Joseph, T.; Pon, R. T.; Lown, J. W. *Biochemistry* **1992**, *31*, 6453, and references cited therein.
3. Luck, G.; Zimmer, C.; Rinert, K. E.; Arcamone, F. *Nucleic Acids Res*. **1977**, *4*, 2655.
4. (a) Penco, S.; Redaelli, S.; Arcamone, F. *Gazz. Chim. Ital*. **1967**, *97*, 1110. (b) Bialer, M.; Yagen, B.; Mecholum, R. *Tetrahedron* **1978**, *34*, 2389. (c) Lown, J. W.; Krowicki, K. *J. Org. Chem*. **1985**, *50*, 3774. (d) Grehn, L.; Ragnarsson, U. *J. Org. Chem*. **1981**, *46*, 3492. (e) Rao, K. E.; Bathni, Y.; Lown, J. W. *J. Org. Chem*. **1990**, *55*, 728. (f) Ding, L.; Grehn, L.; De Clercq, E.; Andrei, G.; Snoeck, R.; Balzarini, J.; Fransson, B.; Ragnarson, U. *Acta Chem. Scand*. **1994**, *48*, 495. (f) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc*. **1992**, *114*, 8783.
5. Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc*. **1996**, *118*, 6141.
6. (a) Bhattacharya, S.; Mandal, S. S. *J. Chem. Soc., Chem. Commun*. **1995**, 2489. (b) Bhattacharya, S.; Mandal, S. S. *J. Chem. Soc., Chem. Commun*. **1996**, 1515.
7. All the intermediates and final compounds were characterized by IR, NMR and electrospray mass spectrometry. Selected spectral data for compounds **1**–**4**. Compound **1**: ^{1}H NMR (300 MHz, $CDCl_3$) δ (ppm) 1.67–1.75 (m, 2H), 2.27 (s, 6H), 2.42 (t, 2H, $J=6$ Hz), 3.4–3.46 (m, 2H), 3.84 (s, 3H), 3.91 (s, 3H), 6.11 (dd, 1H, $J_1=4$ Hz, $J_2=3$Hz), 6.44 (d, 1H, $J=2$ Hz), 6.70 (dd, 1H, $J_1=4$ Hz, $J_2=3$ Hz), 6.75 (t, $J=2$ Hz, 1H), 7.16 (d, 1H, $J=2$ Hz),

7.55 (bs, 1H), 7.70 (s, 1H). LRMS (EI; *m*/*z*) calcd for $C_{17}H_{25}N_5O_2$ (M^+): 331; found: 331. Compound **2**: ^1H NMR (300 MHz, $CDCl_3$) δ (ppm) 1.71–1.77 (m, 3H), 2.31 (s, 6H), 2.44 (t, 2H, J = 6 Hz), 3.43–3.49 (m, 2H), 3.92 (s, 3H), 3.94 (s, 3H), 3.99 (s, 3H), 6.14 (dd, 1H, $J_1 = 4$ Hz, $J_2 = 3$ Hz), 6.41 (d, 1H, $J = 2$ Hz), 6.67 (dd, 1H, $J_1 = 4$ Hz, $J_2 = 3$ Hz), 6.73 (d, $J = 2$ Hz), 6.77 (t, 1H, $J = 2$ Hz), 7.09 (d, 1H, $J = 2$ Hz), 7.18 (d, 1H, $J = 2$ Hz), 7.52 (s, 1H), 7.62 (s, 1H), 7.74 (bs, 1H). ESI-MS; calcd for $C_{23}H_{32}N_7O_3$ (MH^+): 454.5; found: 454.5. Compound **3**: ^1H NMR (300 MHz, $CDCl_3$) δ (ppm) 1.68–1.74 (m, 2H), 2.26 (s, 6H), 2.42 (t, 2H, $J = 6$ Hz), 3.40–3.43 (m, 2H), 3.86 (s, 3H), 3.87 (s, 3H), 3.90 (s, 3H), 4.0 (s, 3H), 6.11 (dd, 1H, $J_1 = 4$ Hz, $J_2 = 3$ Hz), 6.54 (d, 1H, $J = 2$ Hz), 6.60 (d, 1H, $J = 2$ Hz), 6.73 (d, 1H, $J = 2$ Hz), 6.75–6.76 (m, 2H), 7.12 (t, 2H, $J = 2$ Hz), 7.2 (d, 1H, $J = 2$ Hz), 7.68 (s, 1H), 7.70 (bs, 1H), 8.05 (bs, 2H). ESI-MS calcd for $C_{29}H_{38}N_9O_4$ (MH^+): 576.3; found: 576.3. Compound **4**: ^1H NMR (500 MHz, $CDCl_3$) δ (ppm) 1.66–1.69 (m, 2H), 2.20 (s, 6H), 2.36 (t, 2H, $J = 6$ Hz), 3.37–3.39 (m, 2H), 3.77 (s, 6H), 3.81 (s, 3H), 3.83 (s, 3H), 3.93 (s, 3H), 6.06 (bs, 1H), 6.57 (bs, 1H), 6.61 (bs, 1H), 6.65 (bs, 1H), 6.73 (bs, 1H), 6.74 (bs, 1H), 6.78 (bs, 1H), 7.04 (bs, 1H), 7.13 (bs, 1H), 7.15 (bs, 1H), 7.18 (bs, 1H), 7.61 (bs, 1H), 8.09 (s, 1H), 8.18 (s, 1H), 8.49 (bs, 2H). ESI-MS calcd for $C_{35}H_{44}N_{11}O_5$ (MH^+): 698.4; found: 698.4.

8. Lee, M.; Rhodes, A. L.; Wyatt, M. D.; Forrow, S.; Hartley, J. A. *Biochemistry* **1993**, *32*, 4237.