Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 3469–3474

DNA sequence recognition by an imidazole-containing isopropyl-substituted thiazole polyamide (thiazotropsin B)

Andrew J. Hampshire,[a] Hannah Khairallah,[a] Abedawn I. Khalaf,[b] Abdolrasoul H. Ebrahimabadi,[b] Roger D. Waigh,[c] Colin J. Suckling,[b] Tom Brown[d] and Keith R. Fox[a,*]

[a]*School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK*
[b]*Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK*
[c]*Department of Pharmaceutical Sciences, University of Strathclyde, 27 Taylor Street, Glasgow G4 0NR, UK*
[d]*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

Received 21 March 2006; accepted 2 April 2006
Available online 27 April 2006

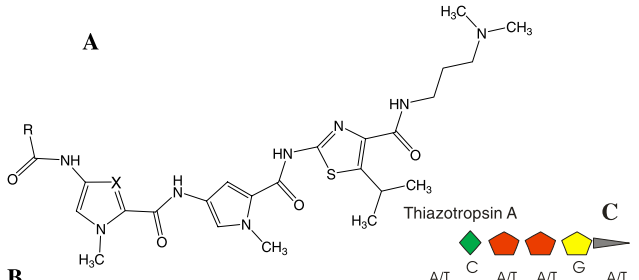
Abstract—We have used DNA footprinting and fluorescence melting experiments to study the sequence specific binding of an imidazole-containing isopropyl-substituted thiazole polyamide (thiazotropsin B) to DNA. While the parent compound (thiazotropsin A) binds to the hexanucleotide sequence ACTAGT, changing one of the *N*-methylpyrrole groups to *N*-methylimidazole changes the preferred binding sequence to (A/T)CGCG(T/A). Experiments with DNA fragments that contain variants of this sequence suggest that the ligand can also bind, with lower affinity, to sequences which differ from this by 1 bp in any position.

A large number of small molecules are known to bind to the minor groove of duplex DNA.[1,2] These compounds, of which the best known are the natural products distamycin and netropsin or the synthetic bisbenzimidazole Hoechst 33258, are generally selective for AT-rich sequences, to which they bind in a 1:1 mode.[3,4] The molecules make close contacts with the narrow minor groove that is typically found in AT-rich sequences; binding to GC-containing sequences is hindered by their wider groove and by steric clash with the 2-amino group of guanine. A large number of modifications to these simple structures have been attempted in order to change the sequence specificity, but in general these resulted in a relaxed specificity, with the inclusion of other sites as well as A/T tracts.[5] However, the observation that distamycin can bind to some sequences in an antiparallel side-by-side 2:1 mode[6] led to the development of the hairpin pyrrole–imidazole polyamides as specific sequence-reading agents. These compounds recognize sequences by the side-by-side pairing of *N*-methylpyrrole and *N*-methylimidazole groups in the DNA minor groove. The simple code recognizes CG with a pyrrole/imidazole (Py/Im) pair, Im/Py targets GC and the Py/Py pair binds to both AT and TA.[7–9]

Although these compounds are able to recognize specific DNA sequences, they are large molecules with limited biological activities. We have therefore developed novel distamycin analogues with heterocyclic rings of different shapes and sizes so as to produce molecules with improved potential for following the contours of the minor groove.[10–13] In our earlier studies we showed that a distamycin analogue (thiazotropsin A, Fig. 1A), in which one of the *N*-methylpyrroles is substituted by an isopropyl-substituted thiazole, is selective for the sequence ACTAGT.[10,11] Two molecules of this compound bind side-by-side in the minor groove but, because of the bulk of the substituted thiazole ring, the molecules are staggered relative to each other and the complex reads a total of 6 bp. In the present study, we have determined the sequence specificity of a closely related compound in which one of the *N*-methylpyrrole groups is replaced with *N*-methylimidazole: thiazotropsin B (Fig. 1A). By comparison with the Dervan rules for sequence recognition, and allowing for the staggered side-by-side binding of these compounds, we predict that this compound should bind to the sequence (A/T)CGCG(A/T). We

Keywords: DNA recognition; Footprinting; Minor groove ligand.
* Corresponding author. Tel.: +44 23 8059 4374; fax +44 23 8059 4459; e-mail: K.R.Fox@soton.ac.uk

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.04.007



Name	Sequence
ACGCGT	5'-F-CCGACGCGTGC-3' 3'-Q-GGCTGCGCACG-5'
ACTAGT	5'-F-CCGACTAGTGC-3' 3'-Q-GCCTGATCACG-5'

Figure 1. (A) Structures of thiazotropsin A (X = C, R = H) and thiazotropsin B (X = N, R = CH_3). (B) Fluorescently labelled sequences used in the thermal melting experiments. F = fluorescein and Q = methyl red. (C). Schematic representation of the binding of thiazotropsins A and B in the DNA minor groove. Green, formyl (thiazotropsin A) or acetyl (thiazotropsin B) 'head'; red, *N*-methylpyrrole; yellow, thiazole; blue, *N*-methylimidazole; grey, DMAP 'tail'.

have used DNase I and hydroxyl radical footprinting and fluorescence melting to explore the sequence specificity of this compound.

Thiazotropsin A was prepared as previously described,[12,13] while thiazotropsin B was synthesized as described below.[14] Figure 2 shows the results of DNase I and hydroxyl radical footprinting experiments with these compounds on DNA fragments MS1 and MS2.[16,17] These footprinting substrates contain all the 136 possible tetranucleotide sequences;[16] MS1 and MS2 contain the same sequence, which is cloned in opposite orientations, allowing good resolution of target sites at both ends of the fragment. Although we anticipated that thiazotropsin B would possess a binding site that is longer than four base pairs (probably (A/T) CGCG(A/T), which is not represented in this sequence), this fragment was used for our initial studies with thiazotropsin A as it provides a better indication of possible target sequences than most natural fragments that possess a random distribution of potential sites. The first two panels of Figure 2 show the DNase I cleavage patterns, in which three footprints are evident on each strand, which are indicated alongside the sequence in Figure 4. The concentrations required to generate these footprints (10 μM and above) are higher than those previously reported for thiazotropsin A (1 μM). The footprint at site 1 persists to lower concentrations (10 μM) than the other two sites (25 μM). This site contains the sequence GCGCGA, which differs from the predicted target, (A/T)CGCG(A/T), at only the first base pair. We have previously shown that thiazotropsin A binds to the sequence ACTAGT and it is clear that thiazotropsin B binds at a different location. This is emphasized in the third panel of Figure 2 which directly compares the DNase I footprinting patterns of these two ligands. The

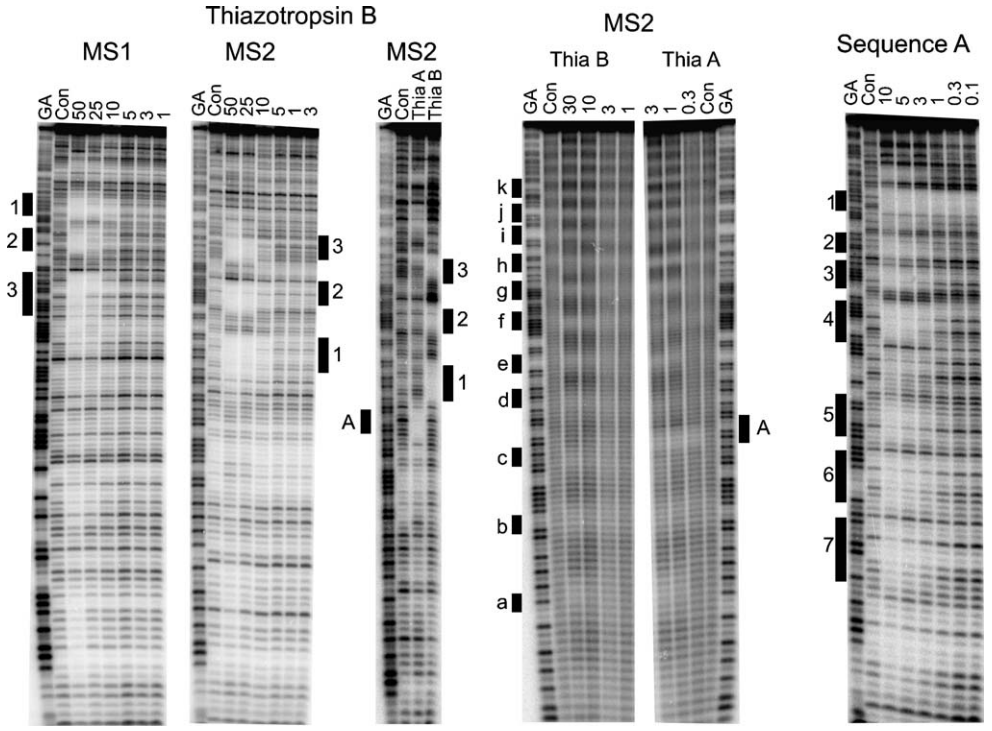


Figure 2. The first two panels show DNase I footprints for thiazotropsin B on MS1 (top strand, Fig. 4A) and MS2 (bottom strand, Fig. 4B). The third panel compares the DNase I footprints for thiazotropsin A (10 μM) and thiazotropsin B (50 μM) on MS2. 'A' indicates the location of the binding site for thiazotropsin A. The fourth and fifth panels show hydroxyl radical cleavage patterns for these ligands on the MS2 fragment. The final panel shows DNase I cleavage patterns of Sequence A in the presence of thiazotropsin B. Ligand concentrations (μM) are shown at the top of each gel lane. GA corresponds to marker lanes specific for purines, while 'con' is cleavage in the absence of added ligand. The filled boxes show the location of the best binding sites. The letters and numbers alongside each footprint correspond to those in Figure 4.

footprint for thiazotropsin A (labelled 'A') is not affected by thiazotropsin B and the three thiazotropsin B sites are not affected by thiazotropsin A. It is clear that these two ligands have very different sequence binding requirements, even though they only differ by a single atom. Site 2 contains the sequence TCCCGT, which also differs from the predicted target at 1 bp in the third position. Site 3, which contains the sequence TAGCAA, is less closely related and differs from the predicted sequence in the second and fourth positions.

The third and fourth panels of Figure 2 compare the hydroxyl radical footprints for thiazotropsins A and B on fragment MS2. This reveals many sites of attenuated cleavage for both ligands, which are indicated in Figure 4. Some of these footprints correspond to the DNase I footprints (sites d, f and h correspond to sites 1–3, respectively) but many others are not seen with DNase I. We have previously noted the presence of additional footprints in the hydroxyl radical cleavage patterns with higher concentrations of thiazotropsin A. These secondary sites are in the same positions for both thiazotropsins A and B.

The final panel of Figure 2 shows DNase I footprinting patterns for thiazotropsin B on Sequence A. This is an unrelated synthetic footprinting substrate that contains several $(G/C)_4$ tracts, including the predicted target site ACGCGT. Several footprints are evident on this fragment (numbered 1–7 in Fig. 2) and their location is indicated in Figure 4B. The footprint at site 1 (ACGCGT) persists to lower concentrations (0.1 μM) than all the other sites, which typically reveal protections at around 5–10 μM. Several of these weaker sites are located around the sequences (A/T)GGCC(A/T) (sites 2, 4 and 5) or (A/T)CCGG(A/T) (site 7). These results suggest that thiazotropsin B possesses considerable sequence selectivity which is clearly different from that of thiazotropsin A.

Based on these results we prepared a synthetic DNA footprinting template which contains several variations around the sequence (A/T)$(G/C)_4$(A/T) (sequence 1, Fig. 4C). The results of DNase I and hydroxyl radical footprinting experiments with this fragment are shown in Figure 3. The first two panels show the patterns of DNase I cleavage on this fragment labelled on the top or bottom strands, in which a number of regions of protection are evident. Three clear regions of protection are produced, which are clearest on the labelled top strand (first panel). Sites 1–3 contain the sequences ACGCGA, TCGCGA and ACGCGT, respectively. Although sites 1 and 2 cover about 8 bases, site 3 is longer and presumably contains a second overlapping binding site, which

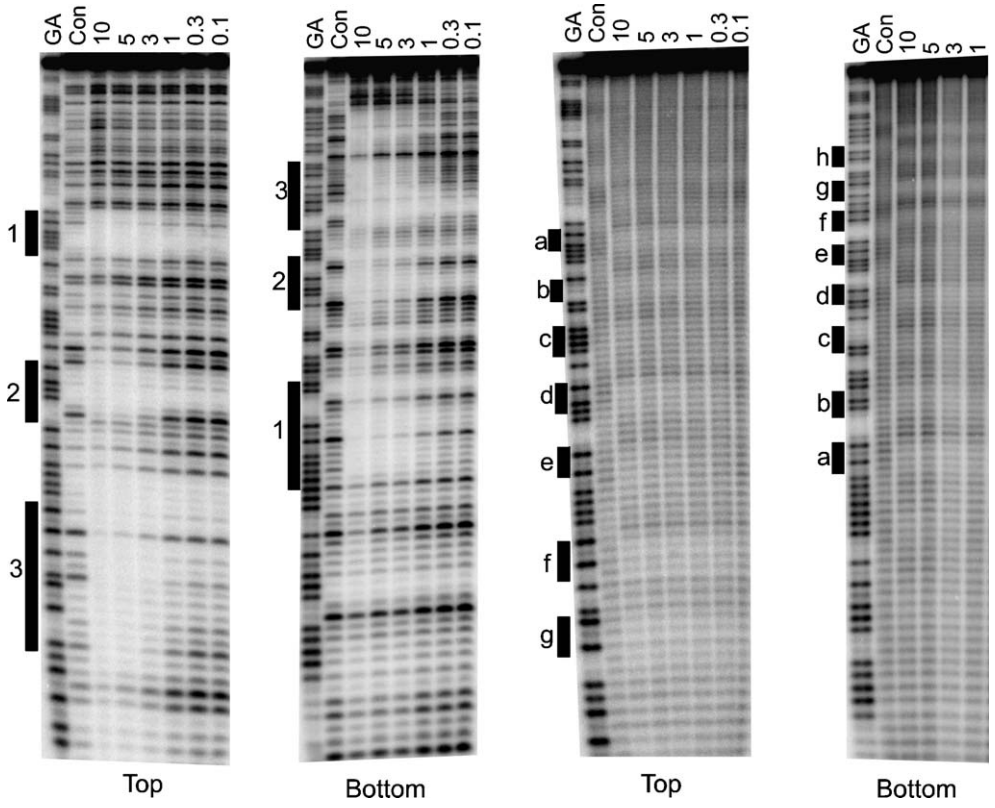


Figure 3. Footprinting patterns showing the interaction of thiazotropsin B with Sequence 1. The first two panels correspond to DNase I cleavage patterns in the presence of different concentrations of thiazotropsin B, while the third and fourth panels are hydroxyl radical cleavage patterns. In each panel the ligand concentration (μM) is shown at the top of each gel lane. GA corresponds to marker lanes specific for purines, while 'con' is cleavage in the absence of added ligand. The filled boxes show the location of the best binding sites, as described in the text. The numbers alongside each footprint correspond to those in Figure 4. 'Top' and 'bottom' indicate which strand of the sequence shown in Figure 1B bears the radioactive label.

may be TAGCGT. Footprints are also evident in similar locations on the labelled bottom strand. It is also informative to note the regions that are not protected by this ligand such as in the upper portion of the gel for the labelled top strand, which contain the sequences TCCGCA and ACCCCT and between sites 1 and 2 (containing TCGCCT and TCGGGA). The third and fourth panels of Figure 3 show hydroxyl radical footprinting patterns for thiazotropsin B on this sequence. Several regions of attenuated cleavage are evident, which are underlined in Figure 4C. Sites a (ACGCGA), d (TCGCGA) and f (ACGCGT) produce the clearest footprints, which persist to the lowest ligand concentrations, and which correspond to the DNase I footprints at sites 1–3. Other footprints are evident at sites b (TCGCCT), c (TCGGGA), e (TCGCGC), g (TAGCGT) and h (AGCGCA) most of which differ from the predicted site by 1 bp.

We further examined the sequence selectivity of this ligand by measuring its effect on the melting of short DNA duplexes, using the fluorescence melting technique that we have previously described.[19] These duplexes (Fig. 1B) are labelled with fluorescein at the 5′-end of one strand and with methyl red at the 3′-end of the complement. These groups are in close proximity in the duplex and the fluorescence is quenched. When this melts the fluorophore and quencher are separated and the fluorescence increases. One sequence contains the binding site for thiazotropsin A (ACTAGT), while the other contains one of the potential sites for thiazotropsin B (ACGCGT). Figure 5 presents typical fluorescence melting curves for the ligands on these sequences, and the calculated T_m values are presented in Table 1. It can be seen that thiazotropsin A increases the T_m of ACTAGT, as previously reported,[11] but has almost no effect on ACGCGT. In contrast thiazotropsin B stabilizes ACGCGT, but has only a small effect on ACTAGT.

These results show that the preferred binding sites for thiazotropsins A and B are very different. On fragments MS1 and MS2 thiazotropsin A binds to the sequence ACTAGT, while thiazotropsin B shows no interaction with this site, but binds around the sequence GCGCGA. It is clear that changing the *N*-methylpyrrole group to *N*-methylimidazole has caused a marked change in the sequence binding characteristics. We previously suggested that thiazotropsin A binds as a staggered side-by-side 2:1 complex[10,11] (formation of the unstaggered complex is prevented by the bulk of the isopropyl-thiazole moiety) in which the central AT base pairs are close to Py/Py pairs (see Fig. 1C). Changing the terminal pyrrole to an imidazole produces Im/Py and Py/Im pairs which, by comparison with the Dervan rules for hairpin polyamides, would be expected to recognize GC and CG, respectively. We would therefore predict that this ligand should bind best to (A/T)CGCG(A/T) (see Fig. 1C). The

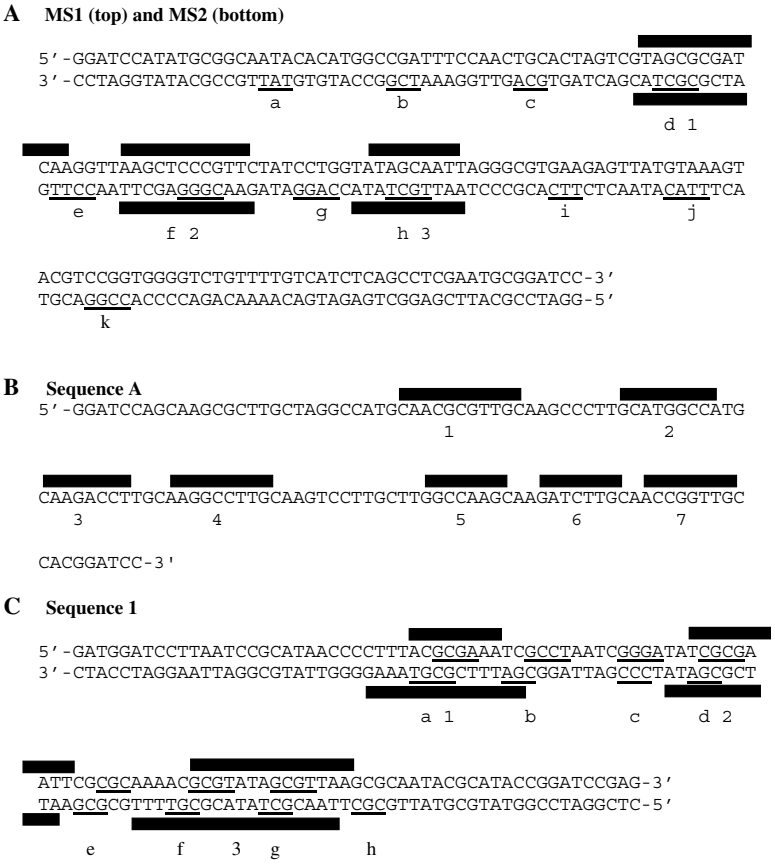


Figure 4. Sequences of the footprinting substrates indicating the regions protected from DNase I cleavage (filled boxes) and hydroxyl radicals (underlined). All the sequences were cloned into the *Bam*HI site of pUC19 (only the insert is shown).

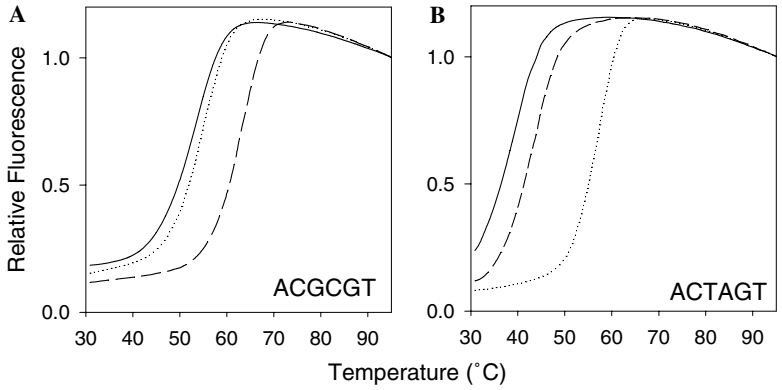


Figure 5. Fluorescence melting profiles for sequences ACGCGT and ACTAGT in the presence of thiazotropsins A and B. Filled line, no added ligand; dotted line, thiazotropsin A (50 μM); dashed line, thiazotropsin B (25 μM for ACGCGT and 50 μM for ACTAGT).

Table 1. Effect of thiazotropsins A and B on the melting temperature of sequences ACGCGT and ACTAGT

	T_m (°C)	ΔT_m (°C)				
		50 μM	25 μM	10 μM	5 μM	1 μM
Thiazotropsin B						
ACGCGT	53.7	12.0	9.8	5.4	1.6	0.1
ACTAGT	39.3	4.0	1.1	0.5	0.2	0.2
Thiazotropsin A						
ACGCGT		3.1	1.2	0.3	0.1	0.0
ACTAGT		19.3	16.6	13.2	9.8	0.6

The first column shows the melting temperatures of the oligonucleotides alone, while the other columns show the change in T_m caused by the ligands. The values shown are the average of three determinations for which the T_m values differed by 0.5 °C or less.

footprinting data confirm that this is a good binding site, but reveal that it can bind to several other related sequences. Sites 1 and 2 in the MS1/MS2 fragments differ from the proposed site by one base pair in the first and third positions, respectively. Site 1 in Sequence A corresponds to the predicted site and produces the best footprint (persisting to the lowest ligand concentrations) on this fragment. Sites 1–3 on Sequence 1 also correspond to the suggested site. It therefore appears that, although we have not exhaustively examined all the possible hexanucleotide binding sites for this ligand, (A/T)CGCG(A/T) is amongst the best binding sites. These studies confirm that the staggered overlap of small molecules can increase the binding site size, thereby achieving the recognition of longer DNA sequences without increasing their molecular size.

It is also clear that both thiazotropsins A and B bind to a range of secondary sites at higher concentrations; these are especially evident in the hydroxyl radical cleavage patterns. These secondary sites are more evident with thiazotropsin B, as this ligand binds to its preferred site less well than thiazotropsin A. The weaker binding for imidazole-containing ligands is consistent with observations with the hairpin polyamides. In contrast to thiazoles and pyrroles, which promote strong binding to DNA, the imidazole ring leads to lower affinity. These secondary sites appear at a diverse range of sequences and many are common for both thiazotropsins A and B. It is possible that these arise from simple 1:1 binding.

References and notes

1. Neidle, S. *Nat. Prod. Rev.* **2001**, *18*, 291.
2. Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.
3. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
4. Abu-Daya, A.; Brown, P. M.; Fox, K. R. *Nucleic Acids Res.* **1995**, *23*, 3385.
5. Lown, J. W.. In *Molecular Aspects of Anticancer Drug–DNA Interactions*; Macmillan: London, 1993; Vol. 1, p 322.
6. Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723.
7. Dervan, P. B.; Poulin-Kerstien, A. T.; Fechter, E. J.; Edelson, B. S. *Top. Curr. Chem.* **2005**, *253*, 1.
8. Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439.
9. Wemmer, D. E.; Dervan, P. B. *Curr. Opin. Struct. Biol.* **1997**, *3*, 688.
10. Anthony, N. G.; Fox, K. R.; Johnston, B. F.; Khalaf, A. I.; Mackay, S. P.; McGroarty, I. S.; Parkinson, J. A.; Skellern, G. G.; Suckling, C. J.; Waigh, R. D. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 1353.
11. James, P. L.; Merkina, E. E.; Khalaf, A. I.; Suckling, C. J.; Waigh, R. D.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2004**, *32*, 3410.
12. Khalaf, A. I.; Ebrahimabadi, A. H.; Drummond, A. J.; Anthony, N. G.; Mackay, S. P.; Suckling, C. J.; Waigh, R. D. *Org. Biomol. Chem.* **2004**, *2*, 3119.
13. Anthony, N. G.; Khalaf, A. I.; Mackay, S. P.; Parkinson, J. A.; Suckling, C. J.; Waigh, R. D. *J. Am. Chem. Soc.* **2004**, *126*, 1138.
14. Synthesis of thiazotropsin B: *N*-[3-(dimethylamino)propyl]-5-isopropyl-2-{[(1-methyl-4-nitro-1*H*-pyrrol-2-yl)-carbonyl]amino}-1,3-thiazole-4-carboxamide[12] (105 mg, 0.25 mmol) and Pd/C–10% (100 mg) were suspended in methanol (20 mL) and hydrogenated for 2.5 h at room temperature. Filtration of catalyst over Kieselguhr under N_2, followed by removal of the solvent under reduced pressure, gave the amine as an off-white solid, which was used without further purification. 4-Acetamido-1-methyl-2-imidazolecarboxylic acid sodium salt[15] (120 mg, 0.7 mmol) was dissolved in DMF (0.5 mL, dry), to which was added a solution of HBTU (137 mg, 0.36 mmol) and

NMM (0.2 mL, dry) in DMF (1 mL, dry) with stirring for 30 min at room temperature. This mixture was added to the amine and stirring was continued at room temperature overnight, followed by purification by HPLC to give the desired material as TFA salt (30 mg, 22% yield) with no distinct melting point. ^{1}H NMR, δ_H [DMSO-d_6]: 1.27–1.28 (6H, d, J = 6.8 Hz, isopropyl); 1.87 (2H, m); 2.02 (3H, s); 2.78 (6H, s, NMe_2); 3.07 (2H, m); 3.32 (2H, m); 3.89 (3H, s, NMe); 3.95 (3H, s, COMe); 4.19 (1H, m, isopropyl); 7.44 (2H, m); 7.48 (1H, m); 7.97 (1H, m); 9.3 (1H, br s); 10.00 (1H, s); 10.23 (1H, s); 12.03 (1H, s). IR [KBr]: 1667, 1550, 1470, 1288, 1198, 1133 cm^{-1} HR-FABMS: found: 558.26332; calculated for $C_{25}H_{36}N_9O_4S$: 558.26110.

15. Tao, Z.; Fujiwara, T.; Saito, I.; Sugiyama, H. *Angew. Chem., Int. Ed.* **1999**, *38*, 650.
16. Lavesa, M.; Fox, K. R. *Anal. Biochem.* **2001**, *293*, 246.
17. Footprinting substrates were cloned into the *Bam*HI site of pUC19. Radiolabelled fragments were isolated from the plasmids by cutting with *Hin*dIII and *Sac*I, and were labelled at the 3′-end of the *Hin*dIII site with [α-^{32}P]dATP. Footprinting reactions were performed as previously described.[18]
18. Fox, K. R.; Waring, M. J. *Methods Enzymol.* **2001**, *340*, 412.
19. Darby, R. A. J.; Sollogoub, M.; McKeen, C.; Brown, L.; Risitano, A.; Brown, N.; Barton, C.; Brown, T.; Fox, K. R. *Nucleic Acids Res.* **2002**, *29*, e39.