Acta Crystallographica Section C
Crystal Structure Communications
ISSN 0108-2701

Ethyl 2-[*N*-(*tert*-butyloxycarbonyl)-L-alanylamino]-4-methyl-1,3-thiazole-5-carboxylate reveals a *trans* orientation of the preceding amide N—H with respect to the thiazole-ring sulfur

Umesh Prasad Singh,[a] Mini Thomas,[b] T. P. Seshadri[a] and Santanu Bhattacharya[b]*

[a]Department of Physics, Indian Institute of Science, Bangalore 560 012, India, and [b]Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, India
Correspondence e-mail: sb@orgchem.iisc.ernet.in

Received 4 July 2000
Accepted 8 September 2000

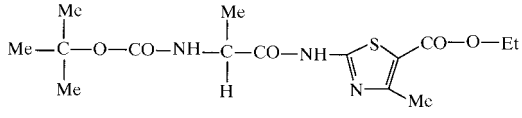
The title molecule, $C_{15}H_{23}N_3O_5S$, was prepared as a synthetic precursor to 4-methylthiazole-based DNA minor groove binders which would bear chiral amino acids in the sequence. The crystallographic evidence presented herein shows that the aromatic amide NH group preceding the thiazole ring points away from the direction of sulfur. The molecule is biplanar, with the dihedral angle between the N-terminus peptide moiety and the thiazole-containing plane being 49.7 (5)°, with a bend at the Cα carbon.

Comment

One approach to the general problem of developing DNA sequence-specific reagents is to modify the lead compounds based on the principles of DNA recognition. Distamycin (Dst) and Netropsin (Nt) are DNA binding compounds that bind in the minor groove of AT-rich sequences of DNA. The determinants in the DNA recognition of these molecules were deciphered from the crystal structures of free Nt (Berman *et al.*, 1979) and also those of the complexes of Nt (Kopka *et al.*, 1985) and Dst (Coli *et al.*, 1987) with the respective target DNA sequences. The sequence specificity in the binding of these molecules is a result of the hydrogen-bonding interactions between the amide NH groups and the adenine N3 or thymine O2 atoms, and the close van der Waals contact between C3(H) of the pyrrole residues and the thymine CH_3 groups. Replacement of pyrrole by other heterocyclic moieties capable of specific DNA recognition by hydrogen-bond acceptance, such as imidazole, pyridine or oxazole, led to the generation of molecules that recognize GC-rich sequences of DNA and are termed 'lexitropsins' or information-reading molecules [for recent reviews, see Wemmer & Dervan (1997) and Baily & Chaires (1998)].

Thiazole lexitropsins with amino and carboxy substituents at the 2- and 5-positions, respectively, along with alkyl substitution at position 4 (type II thiazole lexitropsins), are expected to be more discriminating than the parent antibiotics Dst and Nt against GC tolerance (Rao *et al.*, 1990), as this substitution pattern favours orientation of sulfur towards the minor groove of DNA.

Yet another important component of molecular recognition is ligand chirality and one way of modulating the sequence specificity and affinity of prototype lexitropsins is by introducing chiral amino acids (Herman *et al.*, 1998). As part of our continuing effort in the development of DNA binding molecules (Bhattacharya & Thomas, 1998, 2000*a*,*b*), we synthesized the title compound, (I), a dipeptide precursor in the synthesis of type II thiazole-containing lexitropsins. We attempted to solve the crystal structure of this molecule to determine (*a*) the relative orientation of the amide unit preceding the thiazole ring with respect to the S atom and (*b*) the overall shape of the molecule.



(I)

The title molecule (see Fig. 1) was found to be biplanar, with the dihedral angle between the mean plane *A* containing the N-terminus peptide moiety (C3—C4—O1—C0′—N1—C1*A*) and the mean plane *B* containing the thiazole ring (C1′—N2—C5—N3—C6—C7—S1—C8′—O9—C10) being 49.7 (5)°, with a bend at the Cα carbon. The maximum deviation of the atoms defining the two planes is 0.100 (4) and 0.040 (4) Å for planes *A* and *B*, respectively. The thiazole ring is planar, with a maximum deviation of the ring atoms from the plane of 0.005 (4) Å. Within the ring, the bonding angle at S1 (C5—S1—C7) is 88.36 (15)°, which is unusually small. The C—S bond lengths are also on the lower side, with an average of 1.715 (4) Å. The carbonyl groups that are on either side of the thiazole ring lie almost in the plane of the ring [deviations of 0.063 (4) and 0.039 (4) Å, respectively, for O1′ and O8′] and point in the same direction as sulfur. The non-bonded distances of O1′ and O8′ from S1 are 2.737 (4) and 2.939 (4) Å, respectively, which are significantly shorter than the sum of the van der Waals radii of sulfur and oxygen (3.50 Å; Bondi, 1964).

It is interesting to note that the aromatic amide N—H group preceding the thiazole ring points in the direction opposite to that of the S atom. As mentioned earlier, minor groove binding of type II thiazole-containing lexitropsins would require the preceding amide N—H group to point in the same direction as the S atom. It is worthwhile noting at this point that in the NMR structure of a thiazole/pyrrole lexitropsin complexed with a dodecanucleotide (Kumar *et al.*, 1990), the N-terminus thiazole unit was found to be binding in an intercalative fashion rather than minor groove binding. This unusual observation can be explained in light of the present structure. In the absence of the preceding amide NH group

pointing towards the minor groove, the planar aromatic ring of thiazole-containing lexitropsins could act as an intercalating chromophore, as observed in the case mentioned above. Thus, the results of the observations emerging from the present structure should help the future design of thiazole-based DNA binding compounds.

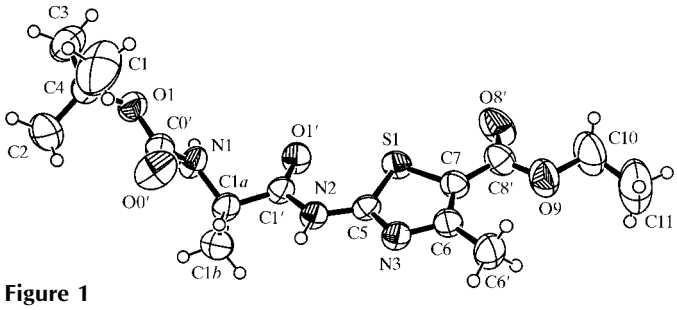


Figure 1
View of the title compound with the atom-numbering scheme. Displacement ellipsoids for non-H atoms are drawn at the 50% probability level.

Experimental

The synthesis of the title compound was carried out by coupling of *t*-Boc-alanine with ethyl 4-methyl-2-aminothiazole-5-carboxylate in the presence of dicyclohexyl carbodiimide and hydroxybenzotriazole. The pure compound was isolated in 96% yield upon column chromatography over silica gel using 20–30% petroleum ether/ethyl acetate as eluent. The diffraction-quality crystals were obtained by layering an acetone solution of (I) (300 µl, 50 m*M*) over 5 ml of water in a test tube. ^{1}H NMR, 200 MHz ($CDCl_3$, δ p.p.m.): 1.36 (*t*, 3H), 1.45 (*s*, 9H), 1.5 (3H merging with the Boc peak), 2.57 (*s*, 3H), 4.29 (*q*, 2H), 4.55 (*bs*, 1H), 5.47 (*bs*, 1H), 10.64 (*bs*, 1H). IR (Nujol, ν, cm^{-1}): 3269, 1707, 1667, 1527.

Crystal data

$C_{15}H_{23}N_3O_5S$
M_r = 357.42
Monoclinic, *C*2
a = 19.2711 (10) Å
b = 9.769 (3) Å
c = 10.2592 (10) Å
β = 103.666 (13)°
V = 1876.8 (6) Å^3
Z = 4
D_x = 1.265 Mg m^{-3}
Cu *K*α radiation
Cell parameters from 25 reflections
θ = 10–35°
μ = 1.784 mm^{-1}
T = 293 (2) K
Plate, colourless
0.13 × 0.06 × 0.04 mm

Data collection

Enraf–Nonius CAD-4 diffractometer
$\omega/2\theta$ scans
Absorption correction: ψ scan (North *et al.*, 1968)
T_{min} = 0.738, T_{max} = 0.931
1944 measured reflections
1885 independent reflections
1756 reflections with $I > 2\sigma(I)$
R_{int} = 0.021
θ_{max} = 69.94°
h = 0 → 23
k = 0 → 11
l = −12 → 12
2 standard reflections
frequency: 60 min
intensity decay: <2.0%

Refinement

Refinement on F^2
$R(F)$ = 0.038
$wR(F^2)$ = 0.109
S = 1.035
1886 reflections
207 parameters
H-atom parameters constrained
$w = 1/[\sigma^2(F_o^2) + (0.0740P)^2 + 0.4897P]$ where $P = (F_o^2 + 2F_c^2)/3$
$(\Delta/\sigma)_{max}$ = 0.018
$\Delta\rho_{max}$ = 0.34 e Å^{-3}
$\Delta\rho_{min}$ = −0.18 e Å^{-3}
Extinction correction: *SHELXL*97 (Sheldrick, 1997)
Extinction coefficient: 0.0027 (3)
Absolute structure: Flack (1983)
Flack parameter = −0.03 (2)

Table 1
Selected geometric parameters (Å, °).

C3—C4	1.499 (2)	C5—N3	1.299 (4)
C4—O1	1.466 (2)	C5—S1	1.710 (3)
O1—C0′	1.346 (2)	N3—C6	1.374 (4)
C0′—O0′	1.206 (2)	C6—C7	1.381 (5)
C0′—N1	1.339 (2)	C7—C8′	1.463 (5)
N1—C1*A*	1.448 (2)	C7—S1	1.720 (3)
C1′—O1′	1.233 (4)	C8′—O8′	1.199 (6)
C1′—N2	1.362 (4)	C8′—O9	1.320 (5)
N2—C5	1.380 (4)		
C0′—O1—C4	120.2 (1)	C5—S1—C7	88.36 (15)
C7—C6—C6′	128.5 (3)	C8′—O9—C10	113.2 (4)
C3—C4—O1—C0′	−178.1 (1)	C1′—N2—C5—S1	1.1 (4)
C4—O1—C0′—N1	169.05 (15)	C6—C7—C8′—O8′	−177.4 (4)
O1—C0′—N1—C1*A*	−177.8 (1)	S1—C7—C8′—O8′	1.5 (5)
C0′—N1—C1*A*—C1′	−117.15 (14)	C7—C8′—O9—C10	179.3 (4)
C1*A*—C1′—N2—C5	176.6 (2)	C8′—O9—C10—C11	−169.0 (5)
C1′—N2—C5—N3	−177.6 (3)		

Molecule (I) crystallized in the monoclinic system, space group *C*2, *Cm* or *C*2/*m* from systematic absences; *C*2 was chosen and confirmed by the analysis. H atoms were treated as riding atoms (C—H_{CH_3} 0.96, C—H_{CH_2} 0.97, C—H_{CH} 0.98 and N—H 0.86 Å). The distance between C10 and C11 was fixed at 1.500 (3) Å to avoid shortening due to thermal vibration.

Data collection and cell refinement: *CAD*-4 *Software* (Enraf–Nonius, 1989); data reduction: *MolEN* (Fair, 1990); program(s) used to solve structure: *SHELXS*97 (Sheldrick, 1990); program(s) used to refine structure: *SHELXL*97 (Sheldrick, 1997); molecular graphics: *PLATON* (Spek, 1990) and *ORTEP*-3 (Farrugia, 1997).

This work was supported by DST, Government of India. UPS thanks UGC, India, for a Senior Research Fellowship.

Supplementary data for this paper are available from the IUCr electronic archives (Reference: VJ1112). Services for accessing these data are described at the back of the journal.

References

Baily, C. & Chaires, J. B. (1998). *Bioconjugate Chem.* **9**, 513–538.
Berman, H. M., Neidle, S., Zimmer, C. & Thrum, H. (1979). *Biochem. Biophys. Acta*, **561**, 124–131.
Bhattacharya, S. & Thomas, M. (1998). *J. Indian Chem. Soc.* **75**, 716–724.
Bhattacharya, S. & Thomas, M. (2000*a*). *Biochem. Biophys. Res. Commun.* **267**, 139–144.
Bhattacharya, S. & Thomas, M. (2000*b*). *Tetrahedron Lett.* **41**, 5571–5576.
Bondi, A. (1964). *J. Phys. Chem.* **68**, 441–451.
Coli, M., Frederick, C. A., Wang, A. H. J. & Rich, A. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 8385–8389.
Enraf–Nonius (1989). *CAD*-4 *Software*. Version 5.0. Enraf–Nonius, Delft, The Netherlands.
Fair, C. K. (1990). *MolEN*. Enraf–Nonius, Delft, The Netherlands.
Farrugia, L. J. (1997). *J. Appl. Cryst.* **30**, 565.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Herman, D. M., Baird, E. E. & Dervan, P. B. (1998). *J. Am. Chem. Soc.* **120**, 1382–1391.
Kopka, M. L., Yoon, C., Goodsell, D., Pzura, P. & Dickerson, R. E. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 1376–1380.
Kumar, S., Jaseja, M., Zimmerman, J., Yadagiri, B., Pon, R. T., Sapse, A. M. & Lown, J. W. (1990). *J. Biomol. Struct. Dyn.* **8**, 99–120.
North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968). *Acta Cryst.* A**24**, 351–359.
Rao, K. E., Bathni, Y. & Lown, J. W. (1990). *J. Org. Chem.* **55**, 728–737.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (1997). *SHELXL*97. University of Göttingen, Germany.
Spek, A. L. (1990). *Acta Cryst.* A**46**, C-34.
Wemmer, D. E & Dervan, P. B. (1997). *Curr. Opin. Struct. Biol.* **7**, 335–361.