Bioorganic & Medicinal Chemistry Letters 20 (2010) 3530–3533

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Crystal structure of an intermolecular 2:1 complex between adenine and thymine. Evidence for both Hoogsteen and 'quasi-Watson–Crick' interactions

Sosale Chandrasekhar [a,*], Tangali R. Ravikumar Naik [a], Susanta K. Nayak [a,b], Tayur N. Guru Row [b,*]

[a] Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, India
[b] Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India

ARTICLE INFO

Article history:
Received 7 December 2009
Revised 16 April 2010
Accepted 28 April 2010
Available online 20 May 2010

Keywords:
Adenine
Base pair
Hoogsteen
Thymine
Watson–Crick
X-ray diffraction

ABSTRACT

The titled complex, obtained by co-crystallization (EtOH/25 °C), is apparently the only known complex of the free bases. Its crystal structure, as determined by X-ray diffraction at both 90 K and 313 K, showed that one A–T pair involves a Hoogsteen interaction, and the other a Watson–Crick interaction but only with respect to the adenine unit. The absence of a clear-cut Watson–Crick base pair raises intriguing questions about the basis of the DNA double helix.

Ever since the discovery of the double helical structure of deoxyribonucleic acid (DNA) by Watson and Crick in the early 1950's,[1] the complementary base pairing scheme therein has been a corner-stone of modern biology. In particular, the mutual recognition of specific base pairs not only forms the bed-rock of molecular biology,[2] but also resonates in key areas of chemical pathology and medicinal chemistry, for example, virus action[2b] and gene therapy.[3]

The Watson–Crick base pairing scheme has been extensively validated, both in molecular-biological and physico-chemical studies. The molecular-biological studies have essentially involved the DNA polymerase catalyzed incorporation of complementary nucleotides into a growing oligonucleotide strand, upon a template strand of defined base sequence.[4] The physico-chemical studies have focused on stability and structure, largely involving DNA melting[5] and X-ray crystal structure determination of oligonucleotides.[6]

Interestingly, however, direct experimental evidence for specific complexation between the DNA bases themselves is scarce. A major problem is the poor solubility of the free bases in most solvents, so obtaining detectable concentrations of the base pairs (say, for spectroscopy) is often difficult. All the bases have considerable positive free energies of solution: ~1.5–5.0 kcals/mol in water, but less in DMSO.[7a] A few early studies, based on solubility[7b] and IR spectroscopy,[7c] provided valuable insights into purine–pyrimidine interactions without, however, significant structural detail.

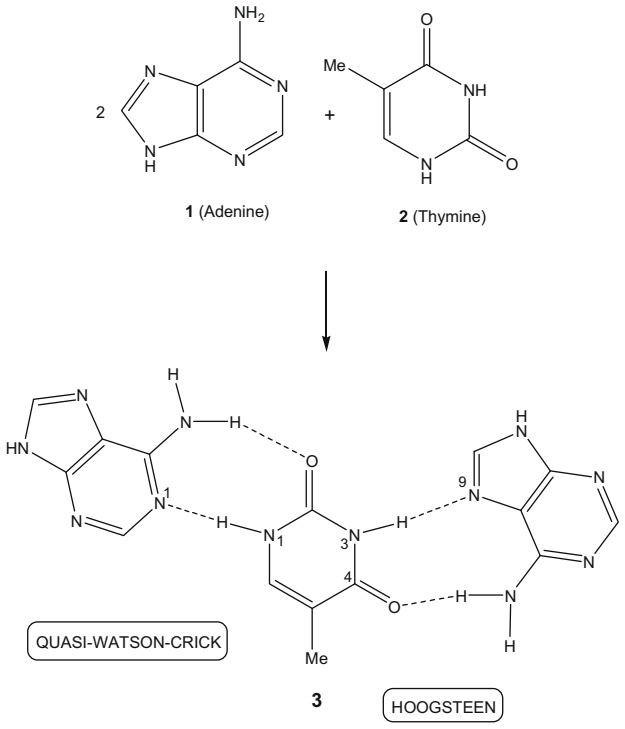
We have been interested in preparing co-crystals of the DNA bases, particularly with a view to determining their crystal structures via X-ray diffraction. Also, because of the above difficulty in observing the base pair complexes in solution, a crystallization-driven approach seemed reasonable and worth exploring. Two pioneering studies, by Etter et al. and by Hoogsteen, report crystal structures of equimolar complexes of 9-methyladenine and 1-methylthymine (but not the free bases).[8]

We have now succeeded in forming a crystalline 2:1 complex (**3**) of the complementary base pairs adenine (**1**) and thymine (**2**) (Scheme 1),[9] and have determined its crystal structure by X-ray diffraction (Fig. 1).

The crystallographic data is summarized in Table 1, and includes a variable temperature study.[10–12] The rather high R_{int} values are due to the poor quality of the crystals obtained. The variable temperature study was undertaken to evaluate the stability of the hydrogen-bonded network and the importance of solvent water molecules. In this complex (**3**), one of the adenine units bonds with the thymine via a Hoogsteen hydrogen bonding interaction,[8,13] involving N_3–H and C_4=O of thymine, and N_7 and NH_2 of adenine. Interestingly, another adenine molecule interacts via its N_1 and NH_2 units with the N_1–H and C_2=O units of the same thymine molecule; this may be termed a 'quasi-Watson–Crick' interaction, as only the adenine engages in Watson–Crick fashion. Four molecules of water stabilize the hydrogen-bonded network in the unit cell.

* Corresponding author. Tel.: +91 80 2293 2689; fax: +91 80 2360 0529 (S.C.).
E-mail addresses: sosale@orgchem.iisc.ernet.in, sosalechandra@hotmail.com (S. Chandrasekhar), ssctng@sscu.iisc.ernet.in (Tayur N. Guru Row).

0960-894X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2010.04.131



Scheme 1. The formation of a 2:1 hydrogen-bonded complex **3** between adenine (**1**) and thymine (**2**) and its structure as determined by X-ray diffraction (cf. Fig. 1). The thymine forms a Hoogsteen pair with one adenine, and a 'quasi-Watson–Crick' pair with the other (hydration is not shown; atom numbering differs from Fig. 1).

The lattice is composed of π-stacked layers formed by the hydrogen-bonded bases (Fig. 2). These further generate a hexagonal supramolecular motif with a central void (Fig. 3); this is filled by four water molecules (designated O_1, O_2, O_3 and O_4, Figs. 1–3), which form a channel.

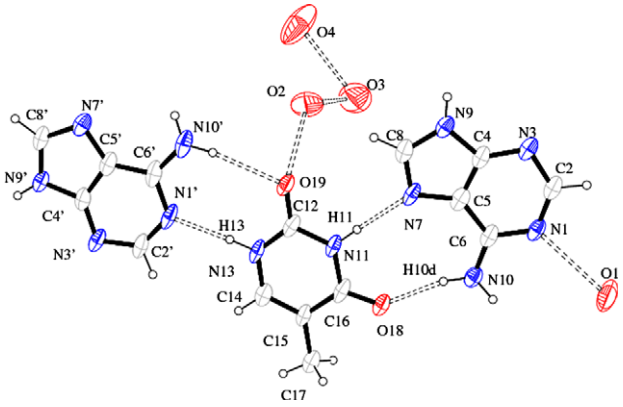


Figure 1. ORTEP diagram of complex **3** (cf. Scheme 1) with 30% ellipsoidal probability at 90 K. Hydrogen bonding interactions are shown as dotted lines (atom numbering follows crystallographic convention).

At 90 K, these are hydrogen-bonded among themselves, and (but for O_4) also with the bases, with well-defined O–H···O and N–H···O hydrogen bonds (Tables 2 and 3). On heating to 313 K, O_4 escapes from the lattice; this modifies the unit cell parameters, but the space group and the void are unaltered. Further heating destroys crystallinity.

The apparently preferred formation of the Hoogsteen base pair in **3** is remarkable, in view of the established fact that the Watson–Crick interaction is the structural basis of DNA. However, there has been a resurgence of interest in the Hoogsteen interaction,[13a,b] and the relative stabilities of the two modes of base pairing remain contentious.[13c] This work suggests that, between the two free bases studied, the Hoogsteen is preferred.

Furthermore, the wide significance of these results is indicated by the analogy between thymine (**2**) and pseudouridine (ψ, via replacement of its sugar residue by the C_5 methyl group in **2**).[14] Indeed, pseudouridine also forms an elaborate three dimensional

Table 1
Crystallographic data for complex **3** (cf. Scheme 1) at two different temperatures

Property	90 K	313 K
Crystal size (mm)	0.35 × 0.11 × 0.05	0.35 × 0.11 × 0.05
Molecular formula	$2(C_5H_5N_5)$ $(C_5H_6N_2O_2)(4H_2O)$	$2(C_5H_5N_5)$ $(C_5H_6N_2O_2)(3H_2O)$
CCDC No.	741243	739016
Formula weight (g mol^{-1})	460.4	444.4
Wave length (Å)	0.71073	0.71073
Crystal system	Monoclinic	Monoclinic
Space group	$P\,2_1/c$	$P\,2_1/c$
a (Å)	14.134(1)	14.114(1)
b (Å)	22.365(4)	22.591(1)
c (Å)	6.692(1)	6.835(6)
β (°)	100.96(1)	103.618(8)
Volume (Å^3)	2076.8(5)	2118.4(4)
Z	4	4
Density (g cm^{-3})	1.47	1.39
F (0 0 0)	951.7	919.7
$\theta_{(min,max)}$	3.1, 26.0	3.1, 26.0
$h_{min,max}$	(−17, 17)	(−17, 11)
$k_{min,max}$	(−27, 26)	(−27, 27)
$l_{min,max}$	(−8, 8)	(−8, 8)
No. of measured reflections	16472	11621
No. of unique reflections	4067	4089
No. of parameters	299	290
R_{int}	0.061	0.104
R_{-obs},R_{-all}	0.095, 0.188	0.146, 0.289
wR_{2-obs},wR_{2-all}	0.288, 0.32	0.322, 0.365
$\Delta\rho_{max,min}$ (e Å^{-3})	0.794, −0.490	0.453, −0.437
G. o. F	1.027	1.053

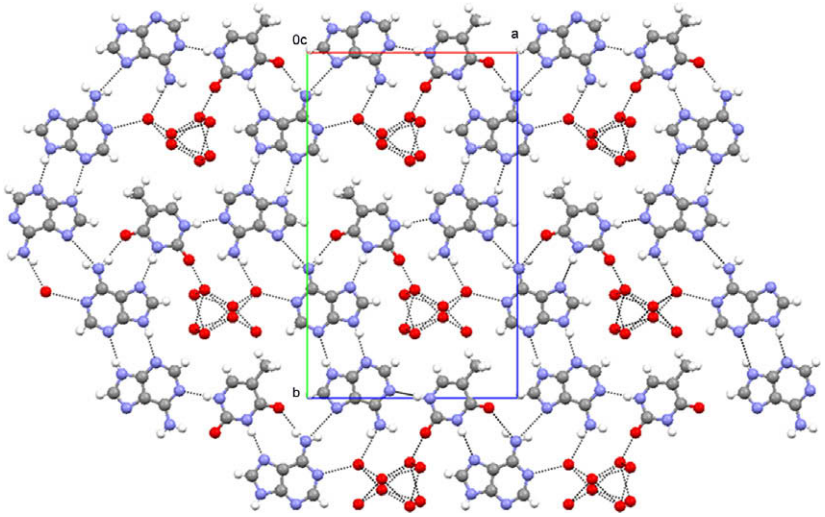


Figure 2. Layered structure of complex **3** (cf. Scheme 1) at 90 K, showing four molecules of water with their hydrogen bonds, as viewed along the *c*-axis of the unit cell.

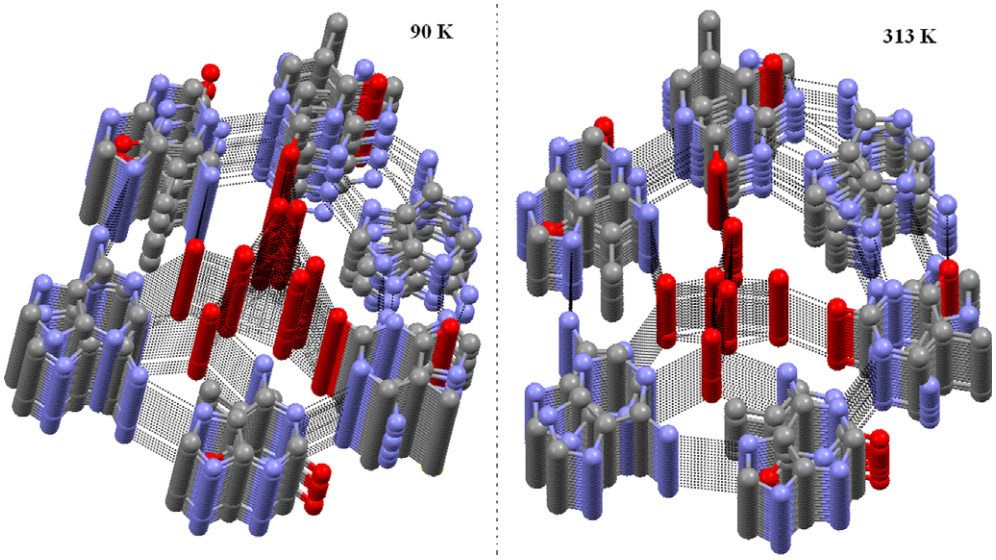


Figure 3. Molecular stacking in the complex **3** (cf. Scheme 1), with the water channels (in red) in the hexagonal voids.

Table 2
Intra- and intermolecular interactions in complex **3** (cf. Scheme 1) at 90 K and 313 K

Crystal **3**	D–H···A	D···H/Å	D···A /Å	H···A /Å	∠D–H···A /°	Symmetry
90 K	N13–H13···N1′	0.88	2.782(6)	1.90	176	x, y, z
	N10′–H10A···O19	0.88	3.178(6)	2.31	168	x, y, z
	N11–H11···N7	0.88	2.845(5)	1.96	177	x, y, z
	N10–H10D···O18	0.88	2.845(5)	1.99	166	x, y, z
	C2′–H2′···O3	0.95	3.199(1)	2.49	131	−x+1, −y+1, −z+1
	N10′–H10B···O1	0.88	2.943(6)	2.11	157	x+1, +y, +z
	N9–H9···N3′	0.88	2.822(7)	1.96	165	−x+1, +y+½, −z+½
	N9′–H9′···N3	0.88	2.879(7)	2.02	164	−x+1, +y−½, −z+½
	N10–H10C···N7′	0.88	2.973(5)	2.21	144	x−1, +y, +z
	C8′–H8′···O18	0.95	3.199(7)	2.27	165	x+1, +y, +z
	C2′–H2′···O2	0.95	3.589(1)	2.72	151	−x+1, −y+1, −z+1
313 K	N13–H13···N1′	0.86	2.821(6)	1.96	174	x, y, z
	N10′–H10A···O19	0.86	3.140(6)	2.29	168	x, y, z
	N11–H11···N7	0.86	2.877(5)	2.02	173	x, y, z
	N10–H10D···O18	0.86	2.845(5)	2.00	166	x, y, z
	N10′–H10B···O1	0.86	3.064(6)	2.25	157	x, y, z
	N9–H9···N3′	0.86	2.821(8)	1.97	167	−x+1, +y+½, −z+½
	N9′–H9′···N3	0.86	2.934(7)	2.09	164	−x+1, +y−½, −z+½
	N10–H10C···N7′	0.86	3.067(5)	2.34	142	x−1, +y, +z
	C8′–H8′···O18	0.95	3.234(1)	2.30	167	x+1, +y, +z
	C2′–H2′···O2	0.95	3.81(2)	2.99	147	−x+1, −y+1, −z+1
	*C2′–H2′···O3					

* Exceeds the sum of the van der Waals radii.

Table 3
Hydrogen bonding interactions of the water molecules in the hexagonal void in **3** (cf. Scheme 1; the involved atoms are designated as X and Y)

Temperature	X···Y	X···Y (Å)
90 K	O1···O4	2.583(1)
	O2···O4	2.770(2)
	O3···O4	2.872(1)
	O2···O3	2.731(2)
	O2···O19(T)	2.815(6)
	O1···N1(A)	2.788(2)
313 K	O1···O3	2.630(2)
	O2···O3	2.789(2)
	O2···O19(T)	2.799(1)
	O1···N1(A)	2.787(1)

hydrogen-bonded network of stacked layers (involving inter-base N–H···O hydrogen bonds).[15d] Although ψ is a modified nucleoside occurring in minor amounts in RNA, it is apparently critical to the structure and function of transfer RNA in particular. Hence, ψ is being intensively studied by a variety of techniques, to unravel the modes of hydrogen bonding believed to mediate its role.[15] The present results are particularly relevant as thymine and *ψ* each possess two free *NH* groups, so their hydrogen bonding modes may well be similar. Interestingly, the crystal structure of *ψ* also suffered from a high R value.[15d]

Finally, in the context of **3** again, it is noteworthy that the 9-methyladenine and 1-methylthymine crystalline complex also indicated preferred Hoogsteen pairing.[8] These, of course, are analogs of the corresponding nucleosides; thus, Hoogsteen pairing seems to be preferred regardless of the number of free *NH* groups in the bases. Therefore, there appears to be gathering evidence that indicates that the basis of the stability of DNA may need to be re-examined.

Acknowledgments

We thank UGC and CSIR for generous fellowships (to T.R.R.N. and S.K.N., respectively). We are grateful to DST (FIST program) for a major grant toward the X-ray diffraction facility.

References and notes

1. Watson, J. D.; Crick, F. H. C. *Nature (London)* **1953**, *171*, 737.
2. Tropp, B. E. *Molecular Biology: Genes to Proteins*, 3rd ed.; Jones & Bartlett: Sudbury, MA, 2008 (a) pp 99–173; (b) 211–316.
3. Swayze, E. E.; Griffey, R. H.; Bennett, C. F. In *Comprehensive Medicinal Chemistry II*; Taylor, J. B., Triggle, D. J., Moos, W. H., Eds.; Elsevier: Oxford, 2007; Vol. 2, pp 1037–1052.
4. Recent work is reviewed in: (a) Kunkel, T. A.; Bebenek, K. *Annu. Rev. Biochem.* **2000**, *69*, 497; (b) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191.
5. (a) Mizrahi, V.; Benkovic, S. *J. Adv. Enzymol. Relat. Areas Mol. Biol.* **1988**, *61*, 437; (b) Aboul-ela, F.; Koh, D.; Tinoco, I., Jr *Nucleic Acids Res.* **1985**, *13*, 4811.
6. (a) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties and Functions*; University Science Books: Sausalito (CA), 2000. pp 101–102, 176–180; (b) Kennard, O.; Hunter, W. N. *Q. Rev. Biophys.* **1989**, *22*, 327.
7. (a) Benoit, R. L.; Frechette, M. *Thermochim. Acta* **1988**, *126*, 155; (b) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963**, *85*, 1289; (c) Chen, M. C.; Lord, R. C. *Biochim. Biophys. Acta* **1974**, *340*, 90.
8. (a) Etter, M. C.; Reutzel, S. M.; Choo, C. G. *J. Am. Chem. Soc.* **1993**, *115*, 4411; (b) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907.
9. Crystallizations were performed in 0.22 mM EtOH in a glass test tube initially sealed with 'parafilm'. This was punctured with a small hole to permit slow evaporation over several days at 27 °C. Colorless crystals were thus obtained, which were collected, dried and used directly for the diffraction study.
10. Full details of the crystallographic study have been deposited in the Cambridge Crystallographic Data Base, Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ (UK), (e-mail: deposit@ccdc.cam.ac.uk), and can be obtained by quoting the Nos. 741243 and 739016.
11. A single crystal of the size 0.35 × 0.11 × 0.05 mm^3 was selected and all X-ray datasets were collected on Oxford Xcalibur CCD Diffractometer with Mova MoKα micro-source radiation (λ = 0.7107 Å) Eos detector with X-ray generator operating at 49.30 kV and 0.98 mA. The cell refinement and the data reduction were done using CrysAlis RED.[16] Variable temperature studies were performed with Oxford diffraction Cryojet system. Since the crystals were not of good quality, the data were collected initially at 292(2) K (room temperature) with the cell dimensions being monitored at regular intervals during data collection. The same crystal was cooled to 90 K and a full data set was collected. The crystal was then slowly warmed to 313 K at a rate of 5 K/h, unit cell parameters were collected at every 5 K interval to check for any significant changes. The cell dimensions showed a shortening of the *a*- and *c*-axes, an elongation along the *b*-axis and a change of 2.6° in the β angle, with a volume change of +42 Å^3 (Table 1). Full datasets were collected at 313 K, which indicated the loss of one water molecule. On further heating the crystal deteriorated irreversibly. The structures with data sets at 90 K and 313 K were solved by direct method using SHELXl97[17] from the WinGX program suite (version 1.70.01).[18] The molecular diagrams were generated using ORTEP-3[19] and the packing diagrams were generated using Mercury-2.2.[20] Geometrical calculations were done using PARST 95[21] and PLATON.[22] The non-hydrogen atoms were refined anisotropically; the hydrogen atoms bonded to C and N atoms were positioned geometrically and refined using a riding model with distance restraints of N–H = 0.86–0.88 Å, aromatic C–H = 0.93–0.95 Å, methyl C–H = 0.96–0.98 Å, and with Uiso(H) = 1.2 Ueq (N, C) and 1.5 Ueq (N, C). Since the data quality at 90 K was not good, the R_{int} being 0.061, no attempts were made to locate hydrogen atoms via difference Fourier methods for the water molecules in both data sets.
12. At room temperature, the lattice has *Z* = 4 with stacking distances ranging from 3.35 to 3.80 Å. As the temperature increased to 313 K, the stacking becomes less compact (3.43–3.75 Å) with the loss of one water molecule from the lattice. The water oxygen atoms participate in strong N–H···O, O–H···N as well as weak C–H···O (O_2 and O_3) interactions with the adenine and thymine moieties (Tables 2 and 3). These weaken with increase of temperature. On further heating the structure collapses irreversibly. It must be pointed out that the role of water molecules with respect to their hydrogen bonding potentials would be understood more clearly when better quality crystals are obtained. However, sustained effects to get these crystals so far have failed.
13. (a) Ghosal, G.; Muniyappa, K. *Biochem. Biophys. Res. Commun.* **2006**, *343*, 1; (b) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. *Structure* **2005**, *13*, 1569; (c) Quinn, J. R.; Zimmerman, S. C.; Del Bene, J. E.; Shavitt, I. *J. Am. Chem. Soc.* **2007**, *129*, 934.
14. We are grateful to a referee for kindly indicating this analogy.
15. (a) Desaulniers, J.-P.; Chang, Y.-C.; Aduri, R.; Abeysirigunawardena, S. C.; SantaLucia, J., Jr.; Chow, C. S. *Org. Biomol. Chem.* **2008**, *6*, 3892; (b) McCrate, N. E.; Varner, M. E.; Kim, K. I.; Nagan, M. C. *Nucleic Acids Res.* **2006**, *34*, 5361; (c) Newwby, M. I.; Greenbaum, N. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12697; (d) Hempel, A.; Lane, B. G.; Camerman, N. *Acta Crystallogr., Sect. C* **1997**, *53*, 1707.
16. Oxford Diffraction (2009). CrysAlis CCD and CrysAlis RED, Version 1.171.33.31. Oxford Diffraction Ltd. Abingdon, Oxfordshire, England.
17. Sheldrick, G. M. *Acta Crystallogr., Sect. A* **2008**, *64*, 112.
18. Farrugia, L. J. *J. Appl. Crystallogr.* **1999**, *32*, 837.
19. (a) Farrugia, L. J. *J. Appl. Crystallogr.* **1997**, *30*, 565; (b) Burnett, M. N.; Johnson, C. K. *Report ORNL-6895*; Oak Ridge National Laboratory: Oak Ridge (TN), 1996.
20. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41*, 466 [Mercury 2.2, CCDC, http://www.ccdc.cam.ac.uk/products/mercury/].
21. Nardelli, M. *J. Appl. Crystallogr.* **1995**, *28*, 569.
22. Spek, A. L. *Acta Crystallogr., Sect. A* **1990**, *46*, c34.