Overcoming the Inadequacy of X-ray Powder Diffraction in Reliable Hydrogen Location with the Aid of First Principles Calculations: Crystal Structure Determination of Orotaldehyde Monohydrate

Leonardo Guidoni,[†] Lorenzo Gontrani,[‡] Luigi Bencivenni,[§] Claudia Sadun,*[,§] and Paolo Ballirano[||]

Dipartimento di Fisica, Sapienza Università di Roma, P. le A. Moro 5, I-00185 Roma, Italy, CASPUR, Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca, Via dei Tizii 6, I-00185 Roma, Italy, Dipartimento di Chimica, Sapienza Università di Roma, P. le A. Moro 5, I-00185 Roma, Italy, and Dipartimento di Scienze della Terra, Sapienza Università di Roma, P. le A. Moro, 5 I-00185 Roma, Italy

Received: October 14, 2008; Revised Manuscript Received: November 12, 2008

First principle calculations of periodic crystal structure were successfully combined with powder X-ray diffraction measures to determine the structure of orotaldehyde monohydrate. This approach was particularly helpful to overcome the inadequacy of powder X-ray diffraction to reliably locate the hydrogen atoms of the intermolecular bond network of the crystal molecules. Density functional calculations were accomplished for the free molecule and its cyclic dimers showing that the most stable centrosymmetric dimer is the building block of the molecular crystal.

1. Introduction

The understanding of the relationships between molecular structure and properties is an essential part of knowledge in chemistry. This is particularly true for nucleic acids, because until their structures were unknown and X-ray diffraction studies were not able to provide any information about them, it was rather difficult to explain their biological functions. Genetic information carried by DNA and RNA is transmitted through complementary base pairing of the three pyrimidine derivatives, cytosine, thymine and uracil, forming hydrogen bonds with the respective complementary purines, namely guanine and adenine. The intermolecular interactions, due to hydrogen bonding network between the purinic and pyrimidinic bases, attach the two DNA strands. The specific interactions between the purinic and pyrimidinic bases, cornerstones of molecular biology, are highly ruled by NH···O or NH···N hydrogen bonds between the appropriate bases. These interactions are expected to depend both on proton acceptor ability of oxygen or nitrogen atoms and on intrinsic acidity of the NH bonds involved in the interaction.

Constant interest has been turned to structural properties of modified nucleic acid bases since most of them are widely implicated in a large number of biological roles. The structural modifications of the bases, acting on the ionization constant value of the protons involved in hydrogen bond formation, are in fact strictly correlated to the base mispair formation.

Much care was devoted to a series of uracil derivatives because they might cause genetic alterations. Uracil, the RNA nucleobase, is present in DNA as a breakdown product of cytosine[1] as its methylated derivative[2] and as formyl derivative,[3] an oxidation product of thymine methyl group.

The knowledge of solid state organization of purines and pyrimidines could be worthy because the two DNA strands are held together through the forces generated by the hydrophobic effect and π stacking which might also occur in these crystals.

The crystal structures of a wide number of uracil derivatives have been determined by X-ray diffraction leading to the chance to compare molecular dimensions and hydrogen-bond systems. The molecules replacing the C5 position of uracil, such as thymine,[4] 5-F-uracil[5,6] and 5-formyluracil,[7] have been deeply investigated both as anhydrous and monohydrate[8] crystals, reaching the conclusion that hydrate molecules packing might be different from that of the corresponding anhydrous forms.

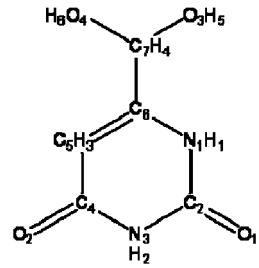
Pyrimidine nucleosides containing a carbon substituent at C6 are difficult to synthesize and therefore they have not been studied[9] as they would have deserved. Among these molecules, orotic acid is one of the most important species, being strongly related to uracil and thymine in its monohydrate crystal structure. Orotic acid, namely vitamin B13, is a very important biological substance. Its crystal and molecular structure was first described by Takusagawa and Shimada[10] and more recently it was redetermined by Portalone.[11] In order to achieve further information about pyrimidine derivatives, we have directed our investigation to orotaldehyde, a molecule which is derived from uracil by replacement of the C6 hydrogen atom by the aldehydic group (systematic name: 2,6-dioxo-1,2,3,6-tetrahydropyrimidine-4-carbaldehyde).

In spite of analogy between 5-formyluracil and orotaldehyde, the aldehydic group may react in different ways, depending on its proximity to a carbonylic or an imino group. The hydrate form, which is the object of the present study, does not present any crystallization water since water molecule reacting with the aldehydic group in the course of a nucleophilic addition reaction may produce a 1,1 diole, or a geminal diole molecule (systematic name: 6-dihydroxymethyl-pyrimidine-(1*H*,3*H*)-2,4-dione). The molecule with the numbering of the atoms is sketched below.

To our knowledge, no structural information seems to be available for this uracil derivative although a closely related

* Corresponding author.
[†] Dipartimento di Fisica, Sapienza Università di Roma.
[‡] CASPUR, Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca.
[§] Dipartimento di Chimica, Sapienza Università di Roma.
[||] Dipartimento di Scienze della Terra, Sapienza Università di Roma.

10.1021/jp809076t CCC: $40.75
Published on Web 12/16/2008



6-(dihydroxymethyl)pyrimidine-2,4(1H-3H)-dione

dihydroxymethyl form was identified studying the free radicals formed from orotic acid.[12] The structure of the substituent in the free radical is quite different from that of the dihydroxymethyl group, being planar as that of the carboxylic group. It is therefore interesting to study under different points of view the solid phase of this compound where the steric hindrance of the polar substituent actually takes part in an extended intermolecular network between hydrogen bonded dimers of monosubstituted uracil.

It is well-known that uracil and its derivatives may aggregate one another in crystal phase through four possible hydrogen-bond donor and acceptor couple sites. The smallest aggregates are hydrogen-bonded molecular dimers, actually found in the crystal structure of uracil and its monosubstituted derivatives.[3−6] These dimers are in turn held all together through a hydrogen bond network and it would be therefore valuable to determine the crystal structure of this molecule to ascertain the effect of the hindering substituent group on the hydrogen bond network of the crystal.

The structure of the dihydroxymethyl product has been determined from parallel-beam angular dispersive X-ray powder diffraction (ADXD). This methodology does not require large crystals of several mm dimensions but just a sizable amount of monophasic microcrystalline sample.

As two couples of hydrogen bonds may occur in the four characteristic regions of neighboring proton donor and acceptor sites of the molecule, as well as in uracil and substituted uracils, *in vacuo* computational studies on the association of two dihydroxymethyl molecules were accomplished to provide some information about the structural changes occurring in the molecular precursor due to self-association and the stability of some cyclic dimers of this molecule. Further, due to X-ray diffraction hampering for the determination of the hydrogen sites by weak scattering power of hydrogen atom, full crystal density functional theory calculations were carried out. The joint use of both the experimental and theoretical methods would therefore enable a valuable determination of the molecular structure of orotaldehyde monohydrate.

2. Experimental and Computational Details

2.1. X-ray powder diffraction measurements. Microcrystalline powder of orotaldehyde monohydrate, purchased from Chess GmBH, was used for the present investigation without any further purification. X-ray powder diffraction data were collected using Cu Kα radiation on a parallel-beam Bruker-AXS D8 Advance automated diffractometer. The instrument was fitted with Soller slits on both incident and diffracted beam and a PSD detector. The sample was charged into a 0.7 mm o.d. borosilicate glass capillary. The investigated angular range was 10.000−140.722° 2θ with a step size of 0.02143° 2θ and a counting time of 10 s (Table 1). Sixty-two Bragg reflections up to 60° 2θ were located by profile-fitting technique using Topas V3 software.[13] The refinement was carried out using pseudo-Voigt functions with constrained widths and M coefficients. Autoindexing was performed using TREOR90[14] imposing as constraints maximum volume of 1000 Å^3 and the density value 1.7(1) g/cm^3, determined by flotation method. Integrated intensities were extracted and subsequently used for crystallographic *ab initio* direct methods structure solution using the EXPO2004 software.[15] Running the program in default mode it was possible to obtain the position of all the eleven non-hydrogen atoms. The derived fractional coordinates were used as starting value for the refinement of the structure with the GSAS suite of programs.[16,17] An absorption measurement has been carried out collecting the transmitted beam through the sample I_t(E) and the incident primary beam I_0(E), both in direct transmission. The derived μ_{eff} was found to be very small and for this reason no absorption correction was applied during the refinement. Peak shape was fitted with a pseudo-Voigt function[18] modified for

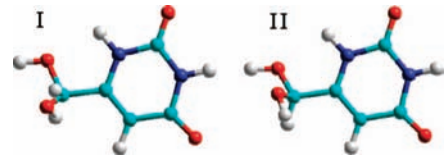


Figure 1. The two conformers of the molecule **I** and **II**. Color coding: C, cyan; O, red; N, blue; H, white.

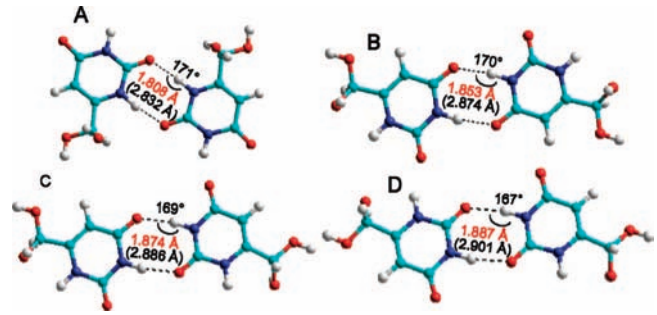


Figure 2. N−H---O hydrogen bonding within a plane for the dimers **A**, **B**, **C** and **D**, respectively. Color coding: C, cyan; O, red; N, blue; H, white. NH---O angle, H---O (in red) and N---O distances (in brackets) are indicated.

TABLE 1: Experimental Details of the X-ray Powder Diffraction Data Collection, Cell Parameters, And Parameters Describing the Peak Shape of the Final Refinement

instrument	Bruker D8Advance
X-ray tube	Cu at 40 kV and 40 mA
incident beam optic	multilayer (Göbel) X-ray mirror
sample mount	rotating capillary (60 rpm)
divergence slit	0.6 mm
Soller slits	incident beam: conventional; diffracted beam: radial
detector	VÅntec 1D
detector opening angle	6° 2θ
2θ range (deg)	10.000−140.722 (6101 data points)
step size (deg)	0.02143
counting time (s)	10
refined parameters	107
a (Å)	15.7742(3)
b (Å)	8.07596(8)
c (Å)	5.00886(8)
β (°)	111.009(1)
V (Å^3)	595.671(15)
GU, GV, GW	1172(4), −488(1), 8.1(2)
LX	12.66(3)
S/L = H/L	0.0117(1)

TABLE 2: Bond Lengths (Å) and Angles (deg) within the Pyrimidine Ring of a Series of Uracil Substituted Molecules[a]

bond lengths	I	II	III	IV	V	VI	VII	av	VIII
O1−C2	1.224	1.215	1.234	1.214	1.219	1.221	1.207	1.219	1.221
O2−C4	1.242	1.245	1.231	1.237	1.223	1.236	1.220	1.233	1.245
N1−C6	1.366	1.358	1.382	1.383	1.373	1.363	1.343	1.367	1.369
N1−C2	1.370	1.371	1.355	1.379	1.363	1.360	1.386	1.369	1.398
C2−N3	1.377	1.376	1.361	1.379	1.376	1.378	1.358	1.372	1.377
N3−C4	1.372	1.371	1.391	1.375	1.380	1.374	1.384	1.378	1.376
C4−C5	1.439	1.430	1.447	1.432	1.441	1.437	1.458	1.441	1.433
C5−C6	1.352	1.340	1.349	1.346	1.334	1.337	1.360	1.345	1.347
bond angles	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**av**	**VIII**
C6−N1−C2	122.4	122.7	122.8	120.6	122.9	123.2	121.7	122.4	121.2
O1−C2−N1	123.8	123.7	122.7	123.3	123.9	123.3	122.3	123.4	123.6
O1−C2−N3	121.3	122.3	122.1	121.3	121.4	122.1	122.5	121.7	121.6
N1−C2−N3	114.8	114.0	115.2	115.4	114.7	114.6	115.2	114.8	114.7
C4−N3−C2	126.3	126.7	126.3	126.3	126.9	126.6	127.5	126.5	126.2
O2−C4−N3	119.6	119.2	118.3	120.0	121.0	120.9	119.2	119.8	119.6
O2−C4−C5	124.6	125.3	126.1	123.9	125.2	124.9	126.8	125.0	124.1
N3−C4−C5	115.7	115.5	115.6	116.1	113.8	114.2	114.0	115.1	116.2
C6−C5−C4	118.6	118.9	118.2	118.3	120.9	120.5	118.6	119.1	118.1
C5−C6−N1	122.1	122.3	121.8	123.3	120.7	120.8	123.1	121.8	123.2

[a] Results from the Rietveld refinement carried out with restraints from the average of columns I−VII are reported for comparison. Key: I, orotic acid, ref 11; II, uracil, ref 35; III, thymine, ref 4; IV, methylthymine, ref 36; V 5-fluorouracil, ref 5; VI, thymine, ref 5; VII, formyluracil, ref 37; Average, average of columns I−VII; VIII, this work.

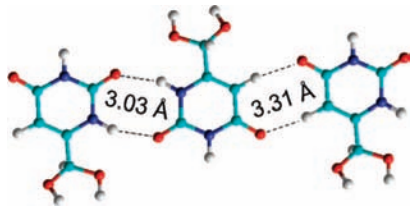


Figure 3. Molecules sequence in the ribbon-like molecular arrangement.

asymmetry.[19] Refined variables were GU, GV, and GW Gaussian, LX Lorentzian, and S/L, and H/L asymmetry parameters along with the sample-displacement from the focusing circle.

Background was fitted by a 30-terms Chebyshev polynomial of the first kind. A total of 38 restraints on bond distances and contacts were imposed to either avoid divergence or convergence toward false minima.[20]

2.2. Computational Details. The geometries of the two 6-dihydroxymethyl-uracil isomers and of the most stable cyclic hydrogen bonded dimers were optimized at the BLYP[21] and B3LYP[22] density functional level with the 6-311++G** basis set. Gaussian-03 package[23] was employed for all the calculations accomplished using a Linux Opteron based cluster, being part of the Biogrid Project at CASPUR. In particular, eight HP DL 585 servers, each one equipped with four AMD Opteron CPUs model 850 at 2.4 GHz and 8GB Ram and four HP DL 145 servers equipped with two AMD Opteron CPUs model 252 at 2.6 GHz and 4GB RAM were made available for the calculations.

Periodic calculations of the crystal structure were performed using the density functional theory starting from the atomic positions first refined by the diffraction data. We expanded the Kohn−Sham orbitals in plane waves up to a cutoff of 70 Rydberg using only Γ point. For the exchange and the correlation part of the universal functional we used the BLYP generalized gradient corrections.[24,25] Core electrons were taken into account using norm-conserving Troullier−Martins type pseudopotentials.[24] Since dispersion interactions might be relevant for accounting $\pi-\pi$ crystal packing we also performed calculations with optimized dispersion-corrected atom-centered potentials (DCACP)[25] as described in ref 26. After geometry optimization, 2 ps Car−Parrinello molecular dynamics run was performed at 500 K to investigate the presence of other local minima. After the dynamics, the geometry was reoptimized and no difference with the starting one was detected. Car−Parrinello molecular dynamics were performed in the *NVT* ensemble using a Nose−Hoover[27] thermostat, a fictitious electron mass of 400 au, and a time step of 3 au.

3. Results and Discussion

The equilibrium geometries of the molecule and of a selected number of its cyclic hydrogen-bonded dimers were obtained by analytical gradient based technique using the density functional theory at the BLYP/6-311++G** and B3LYP/6-311++G** levels..[21,22,28] Harmonic frequencies of vibration were computed for all these molecular species confirming that each structure is a true energy minimum, having all the calculated harmonic frequencies real values. Both the BLYP and B3LYP density functionals provide with the 6-311++G** basis set comparable equilibrium geometries for the two stable conformers of the monomer and for the cyclic dimers. The relative stability of the two conformers of the monomer was determined from single-energy-point MP2 calculations using the 6-311++G** and 6-311++G(3df,3pd) basis sets. Each stable conformer of the molecule is due to different orientations of the OH groups and both the structures have comparable stability. Between these structures, indicated as **I** and **II** and shown in Figure 1, structure **I** is the most stable one being just 5 kJ mol^{-1} lower in energy than structure **II,** as determined from MP2/6-311++G** and MP2/6-311++G(3df,3pd) single point energy calculations.

The different stability of these two structures of the molecule is basically due to the orientation of the hindering group and ultimately from the orientation of the two oxidrilic groups. The shortest intramolecular NH···OH bond distances calculated for structure **I** 2.170 Å is in fact quite shorter that the corresponding one 2.350 Å determined for structure **II** at the B3LYP/6-311++G** level and the two intramolecular OH···OH bond distances are 2.337−3.103 and 2.598−2.483 Å for structures **I** and **II**, respectively. The BLYP/6-311++G** determination of

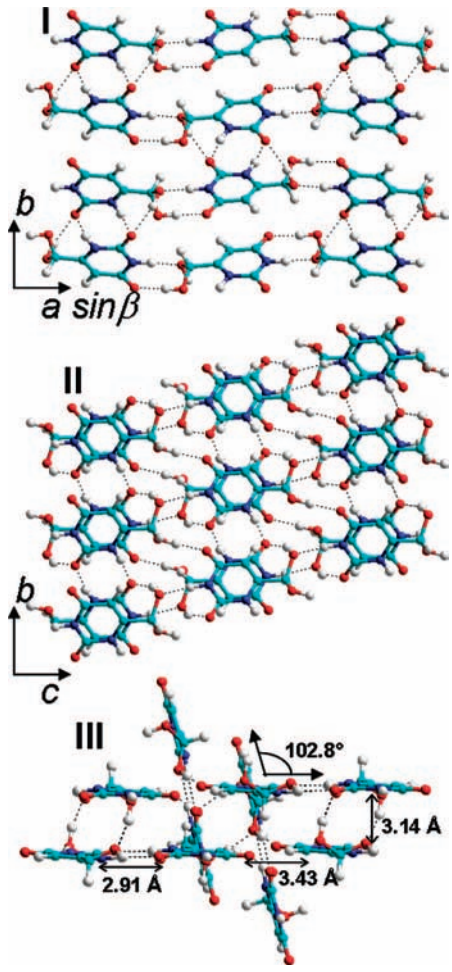


Figure 4. Packing of 6-dihydroxymethyl-pyrimidine-(1H,3H)-2,4dione molecules in the solid state, as determined by X-ray powder diffraction with DCACP restraint. (I) Structure as seen along [001], and (II) along [100]. (III) View of a part of the crystal arrangement showing the distance and the angle between the strands, the NH---O hydrogen bond and of the CH----O lengths respectively as obtained after geometry optimization.

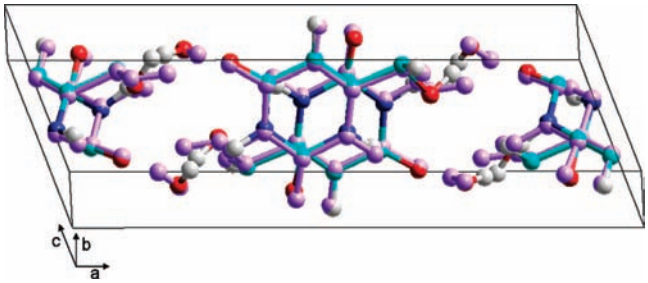


Figure 5. Superimposition of the content of the asymmetric unit as determined by both ADXD and DCAPC calculation. Color coding: colors ADXD, monochromatic DCAPC.

these parameters is thoroughly in line with the B3LYP/6-311++G** conclusions, being the shortest NH···OH bond distances 2.195 Å for conformer **I** and 2.378 Å for **II** while the two OH···OH bond distances are 2.505−2.629 Å for **I** and 2.362−3.131 Å for the less stable conformer **II**.

The four structures considered for the hydrogen-bonded cyclic dimers are shown in Figure 2. BLYP/6-311++G** and B3LYP/6-311++G** calculations suggest that the centrosymmetric dimer **A** is the lowest energy one, whereas the dimers **B**, **C** and **D** are, at the B3LYP/6-311++G** level, 10, 12 and 13.5 kJ mol^{-1} less stable than dimer **A**, respectively. The dimerization

TABLE 3: Calculated Equilibrium Geometries[a] of Orotaldehyde Hydrate Molecule[b] and Its Lowest Energy Centrosymmetric Dimer[c]

	monomer[b]		dimer[c]		crystal	
bond lengths	BLYP	B3LYP	BLYP	B3LYP	BLYP	DCACP
O1−C2	1.225	1.212	1.244	1.230	1.248	1.247
O2−C4	1.229	1.216	1.229	1.215	1.257	1.254
C4−N3	1.429	1.411	1.434	1.415	1.401	1.402
N1−C6	1.385	1.375	1.388	1.378	1.375	1.376
N1−C2	1.409	1.392	1.390	1.374	1.391	1.394
C2−N3	1.397	1.384	1.386	1.374	1.375	1.377
C4−C5	1.463	1.457	1.460	1.454	1.440	1.446
C5−C6	1.363	1.350	1.364	1.351	1.362	1.365
C6−C7	1.531	1.522	1.529	1.519	1.524	1.529
C7−O3	1.428	1.408	1.422	1.403	1.414	1.417
C7−O4	1.433	1.412	1.433	1.412	1.453	1.450
N1−H1	1.020	1.012	1.040	1.031	1.038	1.036
N3−H2	1.020	1.013	1.020	1.013	1.046	1.047
C5−H3	1.086	1.080	1.085	1.078	1.083	1.086
C7−H4	1.099	1.092	1.101	1.095	1.101	1.103
O3−H5	0.977	0.966	0.974	0.964	1.009	1.002
O4−H6	0.974	0.964	0.977	0.967	1.001	0.999
N1−H1···O1′−C2′			1.819	1.808	1.856	1.883
N1···O1′			2.853	2.831	2.879	2.906
	monomer[b]		dimer[c]		crystal	
bond angles	BLYP	B3LYP	BLYP	B3LYP	BLYP	DCACP
C6−N1−C2	124.1	123.8	123.1	123.0	122.0	122.0
O1−C2−N1	122.8	122.9	123.0	123.0	122.8	122.8
O1−C2−N3	124.5	124.1	122.7	122.4	122.1	122.1
N1−C2−N3	112.6	113.0	114.3	114.6	115.0	115.1
C4−N3−C2	128.0	127.9	127.3	127.2	126.3	126.3
O2−C4−N3	120.3	120.4	119.9	120.0	120.7	120.7
O2−C4−C5	126.3	126.0	126.9	126.6	124.2	124.2
N3−C4−C5	113.4	113.6	113.2	113.4	115.1	115.1
C6−C5−C4	120.5	120.2	120.5	120.1	119.4	119.5
C5−C6−N1	121.3	121.4	121.5	121.6	121.9	121.9
H2−N3−C2	115.7	115.7	116.2	116.2	113.8	113.8
H2−N3−C4	116.2	116.3	116.4	116.6	119.9	119.9
C2−N1−H1	116.5	116.6	115.7	115.8	115.6	115.6
H1−N1−C6	118.8	118.9	121.2	121.2	121.1	121.2
N1−C6−C7	116.2	116.1	114.4	114.3	116.7	116.6
C5−C6−C7	122.5	122.4	124.0	124.0	121.5	121.5
H3−C5−C4	117.8	118.0	118.4	118.6	118.7	118.6
H3−C5−C6	121.6	121.9	121.1	121.3	121.8	121.8
O3−C7−C6	112.7	112.5	107.7	107.9	108.9	108.9
O4−C7−C6	106.9	106.9	111.0	110.9	108.3	108.3
O3−C7−O4	111.9	111.9	112.3	112.3	112.3	112.3
H5−O3−C7	108.8	109.5	107.2	107.9	107.8	107.8
H6−O4−C7	107.8	108.5	107.7	108.3	103.9	103.9
H4−C7−O3	111.2	111.3	111.6	111.6	111.9	111.9
H4−C7−O4	104.2	104.7	103.9	104.3	104.9	104.9
N1−H1···O1′			172.6	171.1	167.8	167.9
C2−O1···H1′			126.4	127.4	133.3	133.3

[a] Bond distances (Å), bond angles (deg). [b] Lowest energy structure (see Figure 1). [c] See Figure 2.

energy, including the BSSE corrections based on the counterpoise method[29,30] was determined for each cyclic dimer and the counterpoise corrected dimerization energy values calculated at the B3LYP/6-311++G** level for each dimer is −69, −48, −45 and −43 kJ mol^{-1} for the dimers **A**, **B**, **C**, and **D**, respectively. The high stability of the centrosymmetric dimer **A** is due to its structure which closely mirrors that of the most stable centrosymmetric HB4 dimer of uracil[31] which was studied theoretically from other authors. As far as structural effects due to self-association are concerned, the NH, CO, and CN bonds more directly involved in the intermolecular interaction undergo some valuable modifications. Hydrogen binding causes NH and CO bond lengthening, whereas the CN bond increases its partial double-bond character. In particular, on the basis of the B3LYP/6-311++G** calculations, the NH and CO bonds belonging to the hydrogen bond network of the of the most stable dimer

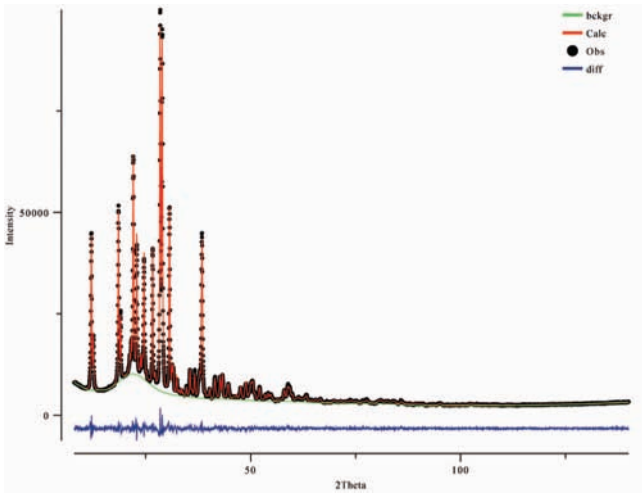


Figure 6. FPA Rietveld refinement of the sample. From top to bottom: background (green), calculated (red), experimental (black), and difference (blue) plots.

A (see Figure 2) are stretched by 0.02 and 0.01 Å respectively with regard to the lowest energy monomer while the average shortening of the CN bonds is 0.015 Å. Among the remaining CN bonds, they are calculated to increase single bond character, lengthening by 0.003 and 0.004 Å. At last, the calculated intermolecular distance NH---O is 1.808 Å, suggesting thus strong hydrogen bond interaction. Hydrogen bond strength reflects on the value of the NH---O (or alternatively N---O) intermolecular bond distance. Such an intermolecular interaction is in fact particularly strong for the most stable dimer **A**, being NH---O = 1.808 Å (N---O = 2.832 Å) while it gets gradually weaker for the other dimers, as inferred from the valuable increase of the bond distances proceeding from dimer **A** to the less stable dimers **B** (NH---O = 1.853 Å; N---O = 2.874 Å), **C** (NH---O = 1.874 Å; N---O = 2.886 Å) and **D** (NH---O = 1.887 Å; N---O = 2.901 Å). This trend holds over at the BLYP level, overestimating this structural parameter by a few hundredth of Å with respect to the B3LYP density functional (NH---O =1.8185 Å; N---O = 2.853 Å for dimer **A**; NH---O = 1.866 Å; N---O = 2.898 Å for dimer **B**; NH---O = 1.889 Å; N---O = 2.917 Å for dimer **C** and NH---O = 1.913 Å; N---O = 2.934 Å for dimer **D**). Bearing in mind the different physical meaning between equilibrium and crystallographic structure, it is of some consolation to state that the calculated geometries of the molecule in the cyclic dimers and the intermolecular NH---O (or N---O) distances compare with the X-ray results fairly well. Density functional calculations provide *in vacuo* equilibrium geometries, neglecting thermal and vibrational effects corrections, whereas crystallographic structure obtained from X-ray determination does not. Furthermore, the *in vacuo* determination of the equilibrium geometry does not reproduce effects of surrounding crystal environment on the molecular structure of the dimers. We will take into account these effects by performing periodic calculations.

As far as the X-ray diffraction measurements are concerned, the autoindexing procedure of the powder diffraction data led to a solution in the monoclinic system for the following cell parameters $a = 15.738(3)$ Å, $b = 8.058(2)$ Å, $c = 4.998(1)$ Å, $\beta = 111.01(3)°$ and the following figures-of-merit M_{20}=14;[32] $F_{30} = 28(0.0204,62)$.[33] The evaluation of reflection conditions provides an unambiguous space-group assignment. In fact, as indicated by the presence of reflection conditions $h00$ for $h = 2n$ and $0k0$ for $k = 2n$, the only possible space-group is $P12_1/a1$ (14) and therefore $P12_1/c1$ after a and c axis switching.

TABLE 4: Agreement Parameters, *J* Texture Index, and Relevant Bond Distances and Angles of the Various Refinements Carried out with Decreasing F Statistical Weight Associated to Bond Restraints[a]

		F100	F50	F20	F10
χ^2		7.05	6.58	6.09	5.77
restr contribution to χ^2 (%)		10.11	8.80	7.33	6.93
wRp (%)		3.34	3.25	3.16	3.08
Rp (%)		2.49	2.43	2.38	2.34
RF^2 (%)		5.41	5.14	4.94	4.83
J		1.049	1.044	1.038	1.035
bond lengths	DCACP	F100	F50	F20	F10
O1−C2	1.247	1.237(2)	1.234(2)	1.228(3)	1.224(3)
O2−C4	1.254	1.254(3)	1.256(3)	1.261(3)	1.266(3)
N1−C6	1.376	1.376(2)	1.376(3)	1.375(3)	1.375(3)
N1−C2	1.396	1.395(2)	1.400(2)	1.411(3)	1.420(3)
N3−C4	1.402	1.397(2)	1.394(3)	1.390(3)	1.385(3)
C2−N3	1.377	1.376(3)	1.377(3)	1.379(4)	1.381(4)
C4−C5	1.446	1.442(2)	1.440(2)	1.436(3)	1.433(3)
C5−C6	1.365	1.362(3)	1.359(3)	1.353(4)	1.349(4)
C6−C7	1.529	1.525(2)	1.523(2)	1.519(3)	1.513(3)
C7−O3	1.417	1.409(2)	1.406(3)	1.404(3)	1.402(3)
C7−O4	1.450	1.443(2)	1.440(2)	1.435(3)	1.428(3)
N3−H2	1.047	1.055(3)	1.058(4)	1.064(6)	1.072(8)
N1−H1	1.036	1.035(4)	1.033(5)	1.028(7)	1.028(9)
C5−H3	1.086	1.088(3)	1.091(3)	1.095(4)	1.100(6)
C7−H4	1.103	1.099(3)	1.097(4)	1.096(6)	1.095(8)
O3−H5	1.002	1.007(3)	1.011(4)	1.016(5)	1.020(7)
O4−H6	0.999	0.994(4)	0.992(5)	0.986(6)	0.975(8)
H2---O4	1.727	1.781(3)	1.804(4)	1.840(6)	1.873(8)
H1---O1	1.883	2.040(5)	2.040(6)	2.010(7)	1.985(9)
H5---O2	1.744	1.721(3)	1.729(3)	1.739(5)	1.744(6)
H6---O1	2.003	1.949(4)	1.951(5)	1.962(7)	1.979(9)
bond angles	DCACP	F100	F50	F20	F10
C6−N1−C2	122.1	121.4(2)	121.3(2)	121.3(2)	121.5(2)
O1−C2−N1	122.8	121.9(2)	121.7(2)	121.4(2)	121.4(2)
O1−C2−N3	122.7	122.6(2)	122.8(2)	123.3(2)	123.6(2)
N1−C2−N3	115.0	115.0(1)	115.0(2)	114.8(2)	114.6(2)
C4−N3−C2	126.4	126.2(1)	125.9(2)	125.5(2)	125.0(2)
O2−C4−N3	120.6	119.2(1)	118.7(2)	117.9(2)	117.3(2)
O2−C4−C5	124.4	125.1(2)	125.2(2)	125.3(2)	125.1(2)
N3−C4−C5	115.0	115.7(2)	116.1(2)	116.8(2)	117.6(2)
C6−C5−C4	119.4	118.5(1)	118.4(2)	118.5(2)	118.5(2)
C5−C6−N1	121.8	122.8(1)	122.9(2)	122.8(2)	122.5(2)
N1−C6−C7	116.6	116.1(2)	116.1(2)	116.2(2)	116.3(2)
O3−C7−C6	108.9	107.5(1)	106.8(1)	106.0(2)	105.3(2)
O4−C7−C6	108.3	107.5(1)	107.4(2)	107.2(2)	107.0(2)
C5−C6−C7	121.5	121.0(1)	120.9(2)	120.8(2)	120.9(2)
O3−C7−O4	112.3	110.7(1)	110.1(2)	109.6(2)	109.7(2)

[a] Bond distances and angles from DCACP calculation are reported for comparison.

A preliminary structure refinement was first carried out by restraining the bond and angles of the molecule at the respective mean values obtained from averaging the molecular geometries of uracil and of some selected monosubstituted uracils (see Table 2) and a further refinement was accomplished using the results of *in vacuo* density functional BLYP and B3LYP calculations. During the latter refinement, all hydrogen atoms of the molecule were constrained at the calculated positions obtained from the density functional computations. Further Rietveld refinements of orotaldehyde monohydrate were carried out imposing structural constraints for the uracil ring, deriving from the four cyclic dimers **A**, **B**, **C**, and **D**, due to the different choice of the couple of donors and acceptors atoms involved in hydrogen bonds. Bond distances and angles of the pyrimidine ring were restrained to a total of eighty-five observations. This approach produced

TABLE 5: Fractional Coordinates and Isotropic Displacement Parameters from the Rietveld Refinement Carried Out with $F = 20$

site	x	y	z	U_{iso}
N1	0.53436(14)	0.3445(3)	0.3482(4)	0.0316(1)[a]
N3	0.39439(12)	0.2446(3)	0.3519(5)	0.0316(1)[a]
C2	0.43814(16)	0.3461(3)	0.2205(5)	0.0316(1)[a]
C4	0.43669(16)	0.1584(3)	0.6039(5)	0.0316(1)[a]
C5	0.53356(17)	0.1697(3)	0.7319(5)	0.0316(1)[a]
C6	0.57807(15)	0.2645(3)	0.6031(6)	0.0316(1)[a]
C7	0.68078(16)	0.2706(3)	0.7149(5)	0.0316(1)[a]
O1	0.39930(11)	0.4186(3)	−0.0051(4)	0.0316(1)[a]
O2	0.38657(10)	0.07427(23)	0.7017(4)	0.0316(1)[a]
O3	0.29724(11)	0.58832(24)	0.4100(3)	0.0316(1)[a]
O4	0.29000(12)	0.70775(21)	0.9826(4)	0.0316(1)[a]
H3	0.5626(5)	0.1505(13)	0.9645(11)	0.05[b]
H4	0.2111(5)	0.3429(8)	0.6692(18)	0.05[b]
H5	0.2303(4)	0.5956(14)	0.3848(24)	0.05[b]
H6	0.3097(7)	0.5985(9)	0.9455(18)	0.05[b]

[a] Constrained to be equal. [b] Unrefined.

a very satisfactory fit of the experimental data, the best agreement indices being consistently obtained for the lowest-energy dimer **A**.

The main features of the molecular packing obtained after imposing the structural constraints might be visualized in terms of a ribbon-like molecular arrangement. The molecules, lying on a tape, are 180° rotated each with respect to the successive in a coplanar sequence, so two adjacent molecules could be hydrogen-bonded. The lengths of the N−H···O bonds, between one side of the first molecule and the other side of the second molecule, and the weak C−H···O bonds, between the second and the third molecule, are equal to 3.03 Å and 3.31 Å, respectively. The large difference between the two determined hydrogen bond distances provides strong evidence on which conformer is present the molecule sequence (see Figure 3), Although these bond lengths are beyond the experimental error, the average value of a hydrogen bond being 2.850 Å.

The molecular arrangement in the crystal could be illustrated in terms of adjacent stacks of narrow ribbons forming between them 102.8° angles. The ribbons are linked together through hydrogen bonds between molecules belonging to adjacent strands. This bond occurs by means of the NH opposite to the dihydroxymethyl group of the first molecule and one OH function of the dihydroxymethyl group of the adjacent molecule lying on the near ribbon. The ribbons stacks distance is 3.14 Å, but the stacking interactions are weaker than hydrogen bonding, there is no hydrogen bond between consecutively strands, and there is no π-interaction between molecules belongings to two parallel tapes.

The low number of hydrogen bonds obtained for this crystal structure and the relatively high temperature (403 K) of water loss determined by thermogravimetry data prompted us to suppose that the hydrogen bond network were more considerable as one might expect at a first glance. Owing to this consideration, it was decided to proceed to perform fully periodic density functional theory calculations. In general, gas-phase and periodic quantum chemical calculations might provide a valuable performance whenever both are jointly flanked by experimental determination of crystal structures.[34]

Starting from the structural data resulting from the Rietveld refinement, we performed DFT geometry optimization of full crystal structure with periodic boundary conditions. The obtained optimized geometry represents a local minimum in the proximity of the starting Rietveld refinement. Although the position of the heavy atoms do not differ significantly by the one refined by the X-ray data, their orientation, the position of hydrogen atoms, and the changes of intermolecular distances showed the presence of an additional out-of-plane hydrogen bond which was unrevealed before.

In the searching for the minimum energy structure of complex systems, different minima might exist around a starting configuration, slightly differing by hydrogen orientation or other subtle local rearrangements. Since we would like to inquire into the stable hydrogen bonding network governing the crystal packing at final temperature, we investigated the robustness of the obtained optimized structure performing Car−Parrinello molecular dynamics in the *NPV* ensemble. During the simulated time the OH groups, although very mobile, maintain the hydrogen bond network of the optimized geometry indicating that this network represents a solid and stable interaction and not just a shallow local minima. Geometry reoptimization of the structure after the dynamics obtained the same configurational minimum supporting again the robustness of our structure. Differences between BLYP and BLYP–DCACP optimized geometries are not significant.

As mentioned before in this work, the geometry optimized structure was in overall agreement with the first refined structure, suggesting the presence of a larger hydrogen bond networks involving both the OH groups of dihydroxymethyl (Figure 4). As one might expect, the main differences are found for the positions of the hydrogen atoms, due to the ADXD technique unfeasibility to accurately locate them.

Moreover the most significant differences which were found are concerning with the position of the dihydroxymethyl group; Figure 5 shows the crystallographic cell content obtained from the first Rietveld refinement, superimposed with the cell content obtained by density functional theory calculations with periodic boundary conditions and in Table 3 are reported the bond distances.

The resulting coordinates of the density functional optimized *in vacuo* structure were employed for the further refinements of the ADXD data. Four different refinements were therefore carried out using decreasing values of the statistical weight F associated with the new restraints imposed using the values obtained for the crystal from calculations. This was done in order to evaluate the influence of restraints on the final atomic coordinates. Only the displacement parameters of non-hydrogen atoms were refined whereas those of hydrogen atoms were kept fixed to the U_{iso} value 0.05 Å^2. The experimental, calculated, and difference plots are shown in Figure 6 for the refinement carried out keeping $F = 20$. The refinements and selected bond distances and angles are shown in Table 4 and positional and displacement parameters in Table 5.

The Rietveld refinement, carried out with these structural constraints, as expected, improved the agreement indices reducing *F*. The effect of this reduction is counterbalanced by a small increase of the difference between the bond distances obtained theoretically from the *in vacuo* density functional calculations and the Rietveld refinement. However, such differences are not significant being smaller than 0.02 Å, therefore within the effect of using different exchange correlation functionals.

Moreover, comparison with the average bond distances of the pyrimidine ring (Table 2) obtained from orotic acid, uracil, thymine, methyl thymine, and thymidine, shows no significant differences and local deviations are related to different hydrogen bond networks except for the N1C2 bond reaching the highest value (1.397 Å) among all the CN bond lengths, indicating thus a valuable single bond character. The use of first principles

calculations performed on the full crystal structure with periodic boundry conditions was crucial to identify further hydrogen binding between the C2−O1 and C4−O2 bonds of the molecule with the oxidrilic groups of the dihydroxymethyl substituent of neighboring molecules (see Figure 4). The resulting hydrogen bond networks turned out to differ from the one of the first refined structure because of the formation of additional hydrogen bond connecting different stripes of head-to-end linked molecules. The hydrogen bond network is therefore meaningfully wider because both the OH bonds of the dihydroxymethyl group do actively participate to crystal building. The main features of the crystal structure are still preserved with regard to the first refinement, notwithstanding the hydrogen bond network is now larger since all the donor and acceptor groups of the uracil framework and the two OH bonds of the dihydroxymethyl group are effectively engaged in the hydrogen bond formation.

4. Conclusions

In the present work, the combined use of experimental powder X-ray diffraction and quantum-mechanical calculations successfully describe the complex structural pattern of intermolecular interactions occurring in the molecular crystal of the uracil derivative orotaldehyde.

To accomplish such a goal, two different kinds of computational approaches were employed to refine the X-ray experimental data.

During the first round, through models obtained from high level molecular orbital calculations using hybrid density functionals and extended basis sets, a detailed description of geometry and energy of orotaldehyde cyclic dimer was obtained. It was also found that the Rietveld refinement using the constraints built with the lowest energy dimer geometry provides the best agreement indices.

In the fitted structure at this stage, a partial degree of hydrogen bonding (between adjacent strands only) was found, in contrast with the high temperature water loss detected by thermogravimetry data. A second model, based on a density functional simulation of the whole periodic crystal, reveals indeed a more extended hydrogen bond network with further out-of-plane interactions. This new structure leads to a better Rietveld refinement and is compatible with the observed high temperature water loss of the crystal.

Summarizing, we have shown that the coupling between theoretical models of increasing complexity and powder diffraction refinement methods may overcome the limitations of X-ray scattering in describing hydrogen bonding patterns, a subject of paramount importance in the description of biologically interesting molecules.

Acknowledgment. We would like to thank Dr. Jolanda Francolini, "Sapienza" University of Rome, for TGA measures. We are also grateful to Prof. Ruggero Caminiti, "Sapienza" University of Rome, for valuable suggestions. This research was supported by the Ministry of Education, University, and Research, Italy.

Supporting Information Available: Crystal structure report for orotaldehyde monohydrate in cif format. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Takahashi, I.; Marmur, J. *Nature* **1963**, *197*, 794–795.
(2) Portalone, G.; Ballirano, P.; Maras, A. *J. Mol. Struct.* **2002**, *608*, 35–39.
(3) Masaoka, A.; Terato, H.; Kobayashi, M.; Ohyama, Y.; Ide, H. *J. Biol. Chem.* **2001**, *276*, 16501–16510.
(4) Gerdil, R. D. *Acta Crystallogr.* **1961**, *14*, 333–344.
(5) Barnett, S. A.; Hulme, A. T.; Tocher, D. A. *Acta Crystallogr.* **2006**, *C62*, 412–415.
(6) Hulme, A. T.; Price, S. L.; Tocher, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 1116–1117.
(7) Baldini, M.; Belicchi Ferrari, M.; Bisceglie, F.; Pelosi, G.; Pinelli, S.; Tarasconi, P. *Inorg. Chem.* **2003**, *42*, 2049–2055.
(8) Perrier, P. R.; Byrn, S. R. *J. Org. Chem.* **1982**, *47*, 4671–4676.
(9) Klein, R. S.; Fox, J. J. *J. Org. Chem.* **1972**, *37*, 4381–4386.
(10) Takusagawa, F.; Shimada, A. *Bull. Chem. Soc. Jpn.* **1973**, *46*, 2011–2019.
(11) Portalone, G. *Acta Crystallogr.* **2008**, *E64*, o656.
(12) Hiittermann, J.; Ward, J. F.; Myers, L. S. J. *J. Phys. Chem.* **1970**, *74*, 4022–4029.
(13) Bruker AXS. Topas V3: General profile and structure analysis for powder diffraction data. *User's manual*; Bruker AXS: Karlsruhe, Germany, 2005.
(14) Werner, P. E.; Eriksson, L.; Westdahl, M. J. *Appl. Cryst.* **1985**, *18*, 367–370.
(15) Altomare, A.; Caliandro, R.; Camalli, M.; Cuocci, C.; Giacovazzo, C.; Moliterni, A. G. G.; Rizzi, R. J. *Appl. Cryst.* **2004**, *37*, 1025–1028.
(16) Larson, A. C.; Von Dreele, R. B. GSAS. General Structure Analysis System. LAUR 86−748; Los Alamos National Laboratory: Los Alamos, NM, 1985.
(17) Toby, B. H. *J. Appl. Crystallogr.* **2001**, *34*, 210–213.
(18) Thompson, P.; Cox, D. E.; Hastings, J. B. *J. Appl. Crystallogr.* **1987**, *20*, 79–83.
(19) Finger, L. W.; Cox, D. E.; Jephcoat, A. P. *J. Appl. Crystallogr.* **1994**, *27*, 892–900.
(20) Baerlocher, Ch. Restraints and constraints in Rietveld refinement. In *The Rietveld Method*; Young, R. A., Ed.; Oxford Science: Oxford, U.K., 1993; pp 186−196.
(21) Becke, A. D, *Phys. Rev.* **1988**, *38*, 3098–3100.
(22) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004. Gaussian 03, Revision C.02; Gaussian, Inc., Wallingford CT, 2004.
(24) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.
(25) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004/1–153004/4.
(26) Lin, I. C.; Coutinho-Neto, M. D.; Felsenheimer, C.; von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2007**, *75* (20), 205131/1–205131/5.
(27) Hoover, W. G. *Phys. Rev.* **1985**, *A 31*, 1695–1697.
(28) Lee, W. Y.; Parr, R. G. *Phys. Rev.* **1988**, *B 37*, 785–789.
(29) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024–11031.
(30) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
(31) Frey, J. A.; Müller, A.; Losada, M.; Leutwyler, S. *J. Phys. Chem.* **2007**, *B 111*, 3534–3542.
(32) de Wolff, P. M. *J. Appl. Crystallogr.* **1968**, *1*, 108–113.
(33) Smith, G. S.; Snyder, R. L. *J. Appl. Crystallogr.* **1979**, *12*, 60–65.
(34) Panck, J.; Stare, J.; Hadzi, D. *J. Phys. Chem.* **2004**, *A108*, 7417–7423.
(35) Stewart, R. F.; Jensen, L. H. *Acta Crystallogr.* **1967**, *23*, 1202.
(36) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 28–38.
(37) Portalone, G.; Colapietro, M. *Acta Crystallogr.* **2007**, *C63*, o650–o654.

JP809076T