Available online at www.sciencedirect.com
ScienceDirect
Mendeleev Commun., 2020, **30**, 660–662

Mendeleev Communications

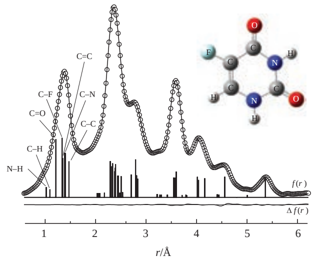
Molecular structure of 5-fluorouracil from gas-phase electron diffraction data and quantum-chemical calculations

Natalja Vogt,*[a,b] Denis N. Ksenafontov,[a] Denis Savelev[b] and Anatoliy N. Rykov[a]

[a] *Department of Chemistry, M. V. Lomonosov Moscow State University, 119991 Moscow, Russian Federation*
[b] *Section of Chemical Informational Systems, University of Ulm, D_89081 Ulm, Germany.*
E-mail: natalja.vogt@uni-ulm.de

DOI: 10.1016/j.mencom.2020.09.036

The accurate equilibrium molecular structure of a canonical tautomer of 5-fluorouracil was determined by electron diffraction taking into account anharmonic vibrational corrections calculated with *ab initio* force field. This structure was applied to benchmark the results of CCSD(T) computations used for the analysis of substitution effects in uracil derivatives.



Keywords: equilibrium molecular structure, 5-fluorouracil, gas-phase electron diffraction, coupled-cluster computation, substitution effects in uracil derivatives.

Due to its antibacterial activity, 5-fluorouracil (5FU) is used as a drug against infections.[1,2] It was suggested as an anticancer agent,[3] and it is commonly used in the treatment of head, neck, breast, colorectal, aerodigestive tract and other cancers.[4,5] 5FU belongs to a few compounds that can replace a base in nucleic acids, that is, to form ‘fraudulent’ nucleic acids.[6] In spite of its importance for medicine and pharmacy, the fundamental properties of 5FU, such as its accurate molecular structure, have not been studied well. Naumov *et al.*[7] determined the thermal-average structure of this molecule in a gas-phase electron diffraction (GED) study. However, this structure cannot be directly compared with other experimental structural data and used for benchmarking high-level quantum-chemical computations. The purpose of this work was the GED determination of the accurate equilibrium structure of 5FU taking into account vibrational corrections up to cubic terms. This structure will be used to benchmark *ab initio* calculations at the coupled-cluster level of theory.

The GED experiment was carried out on the EG-100M apparatus with an accelerated voltage of about 60 kV for long (LD) and short (SD) nozzle-to-film distances at a nozzle temperature of 484 K.[†] A commercial sample (declared purity, 99%) was used without further purification. The total electron scattering intensities $I(s)$[†] were obtained by the initial processing of recorded electron diffraction patterns using a method described elsewhere.[8]

All quantum-chemical calculations were carried out using the Gaussian program package,[9] except for the CCSD(T)[10,11] structure optimization performed with the Molpro program.[12,13] According to predictions of MP2[14] and B3LYP[15,16] calculations (with the cc-pVTZ basis set[17]), the diketo tautomer of 5FU is noticeably lower in energy than all other possible tautomers (a dienol and four keto-enol forms[†]). The next low-energy tautomer (dienol form) is higher in energy relative to the diketo form by 10.9 kcal mol^{-1} (MP2). Therefore, only the canonical tautomer [5-fluoro-2,4(1*H*,3*H*)-pyrimidinedione] can be expected in a gas phase at the temperature of the experiment. According to the results of calculations, this tautomer is planar (C_s symmetry) and relatively rigid with the lowest vibrational frequency higher than 130 cm^{-1}. The structure optimization was also carried out at the CCSD(T) level of theory in conjunction with the cc-pVTZ basis set. The final structure of the CCSD(T)_AE/aug-cc-pCVQZ quality (Table 1) was estimated using an additivity approximation as follows:

$$r_e^{BO} = r_e(\mathrm{CCSD(T)/cc\text{-}pVTZ}) + \Delta r_{corr} + \Delta r_{T\rightarrow aQ}, \quad (1)$$

where $\Delta r_{corr} = r(\mathrm{cc\text{-}pwCVQZ\_AE}) - r(\mathrm{cc\text{-}pwCVQZ})$ are structural changes due to a core-valence correlation,[18] $\Delta r_{T\rightarrow aQ} = r(\mathrm{aug\text{-}cc\text{-}pVQZ}) - r(\mathrm{cc\text{-}pVTZ})$ are structural changes due to the enlargement of the basis set from triple- to quadruple-ζ size and the inclusion of diffuse functions.[19] The values of Δr_{corr} and $\Delta r_{T\rightarrow aQ}$ were estimated at the MP2 level.[†]

The harmonic and anharmonic (cubic) force fields required for the calculations of vibrational corrections to thermal-average internuclear distances were computed at the MP2/aug-cc-pVTZ level. The total corrections $\Delta r_{ij} = r_{a,ij} - r_{e,ij}^{se}$ including harmonic shrinkages, kinematic and dynamic anharmonicity effects and centrifugal distortions[†] were calculated according to Sipachev.[20–22] The semiexperimental equilibrium structure, r_e^{se}, was derived from the GED data augmented by the results of *ab initio* calculations (Table 1). Figure 1 shows the radial distribution functions derived by the Fourier transformation of final molecular electron scattering intensities $sM(s)$.[†]

As can be seen in Table 1, some parameters of the r_a structure previously determined by Naumov *et al.*[7] are inaccurate, for instance, the N(1)–C(6) bond length is too short (by 0.047 Å) and the N(1)–C(6)–C(5) angle is larger by 1.1° in comparison with the corresponding parameters refined in this work. The computed structure of CCSD(T)_AE/aug-cc-pCVQZ quality, r_e^{BO}, agrees well with the r_e^{se} one, namely within a few thousandths of angström and a few tenths of degree for the bond lengths and angles, respectively. The high accuracy of such quality r_e^{BO} structure was confirmed many times.[23–25] Discrepancies between the structures

---

[†] See details in Online Supplementary Materials.

Table 1 Experimental and *ab initio* structures of 5FU (see Figure 1 for atom numbering) in comparison with the *ab initio* structures of thymine and uracil (bond lengths in Å, and bond angles in °).

Parameter	5FU					uracil	thymine
	$r_e^{\mathrm{BO}\,a,b}$	$r_e^{\mathrm{se}\,a,c,d}$	r_a^{a}	$r_a^{d,e}$	$r(\mathrm{XRD})^{f}$	$r_e^{\mathrm{BO}\,g,h}$	$r_e^{\mathrm{BO}\,h,i}$
N(1)–C(2)	1.378_4	$1.378(2)^1$	1.390	$1.394(2)^1$	1.366(2)	1.379_6	1.377_6
C(2)–N(3)	1.382_8	$1.382(2)^1$	1.392	$1.394(2)^1$	1.383(2)	1.377_5	1.378_3
N(3)–C(4)	1.393_2	$1.393(2)^1$	1.404	$1.410(2)^1$	1.375(2)	1.400_5	1.393_4
N(1)–C(6)	1.379_8	$1.379(2)^1$	1.391	1.343^j	1.367(2)	1.372_9	1.378_9
C(4)–C(5)	1.460	$1.458(7)^j$	1.468	1.464(2)	1.444(2)	1.457_8	1.462_1
C(5)=C(6)	1.340_5	$1.342(5)^2$	1.348	$1.355(2)^2$	1.350(3)	1.344_6	1.345_5
C(2)=O	1.210_1	$1.212(1)^3$	1.215	$1.217(1)^3$	1.223(2)	1.211_1	1.210_3
C(4)=O	1.209_7	$1.212(1)^3$	1.215	$1.217(1)^3$	1.232(2)	1.213_1	1.214_5
C(5)–X^k	1.329_5	$1.329(5)^2$	1.337	$1.342(2)^2$	1.335(2)	1.076_4	1.493_3
N(1)–H	1.004_5	$1.004_5{}^l$	1.020	1.014^m	1.029(7)	1.004_8	1.004_1
N(3)–H	1.009^4	$1.009_4{}^l$	1.025	1.014^m	1.030(8)	1.009_0	1.008_5
C(6)–H	1.078_6	$1.078_6{}^l$	1.094	1.083^m	1.083(8)	1.079_4	1.080_2
N(1)–C(2)–N(3)	113.1_0	$113.1(1)^4$		$112.4(1)^4$	115.3(1)	113.5_1	112.7_3
C(2)–N(3)–C(4)	128.7_2	$128.7(1)^4$		$128.6(1)^4$	126.5(1)	127.7_3	128.0_8
N(3)–C(4)–C(5)	112.5_1	$112.6(2)^j$		$111.8(1)^4$	114.0(1)	113.8_2	114.9_7
C(4)–C(5)=C(6)	121.2_1	$121.2(2)^j$		$121.3(1)^4$	121.0(1)	119.6_6	117.7_8
C(4)–C(5)–X^k	117.3_3	$117.3_3{}^l$		115.9^n	116.9(1)	118.3_3	118.0_7
N(3)–C(4)=O	121.9_4	$121.9_4{}^l$		119.5^n	121.8(1)	120.4_1	120.4_4
N(3)–C(2)=O	123.4_9	$123.4_9{}^l$		–	121.9(1)	123.6_8	123.8_1
C(2)–N(1)–C(6)	123.8_2	$123.8(1)^4$		124.0(2)	122.8(1)	123.6_0	123.7_2
N(1)–C(6)=C(5)	120.6_3	$120.7(1)^4$		121.8(3)	120.5(1)	121.6_8	122.7_2

[a] This work. [b] See equation (1). [c] Parenthesized uncertainties are 3σ. [d] Parameters with equal superscripts were refined in one group, structural differences in each group were assumed at the values from theoretical computations. [e] Ref. 7. [f] Refs. 26,27. [g] Ref. 24. [h] Structure of CCSD(T)_AE/cc-pwCVQZ quality. [i] Ref. 23. [j] Dependent parameter. [k] X is F, H or C (methyl) atom. [l] Assumed at the r_e^{BO} value. [m] Assumed at the value from B3LYP/6-31G* calculation. [n] Assumed.

in a gas phase and a solid state[26,27] are relatively large (up to 0.02 Å and 2° in the bond lengths and angles, respectively, see Table 1) reflecting the strong effect of intermolecular interactions in the crystal.

The high accuracy of the computed structure allows for the observation of fine structural effects in uracil derivatives. The C(5)–C(4) and C(5)=C(6) bond lengths in 5-methyluracil (thymine)[25] with an electron-donating methyl group are longer than those in 5FU with fluorine as an electron acceptor by 0.002 and 0.004 Å, respectively (see Table 1[†]). The adjacent C(4)–C(5)=C(6) angle in 5FU is larger than that in thymine (by 3.4°). These structural effects correlate with charge distributions in the molecules.[†] Due to a negative partial charge on the C(5) atom in thymine, the C(5)–C(4) and C(5)=C(6) bonds are weaker and longer than those in 5FU with a positive partial charge on the C(5) atom. Note that the N(3)–C(4) bond in 5FU is of the same length as that in thymine.

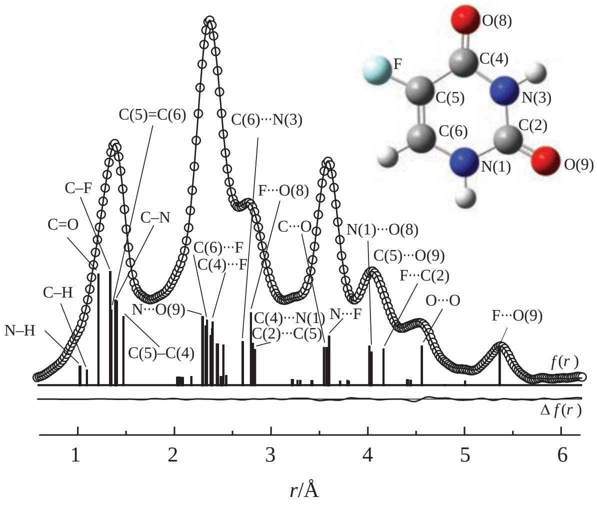


Figure 1 Experimental radial distribution curve $f(r)_{\mathrm{exp}}$ (open circles) derived with a damping factor of 0.0016 and its theoretical counterpart $f(r)_{\mathrm{theor}}$ (solid line) with vertical bars of the molecular terms. Difference curve, $\Delta f(r) = f(r)_{\mathrm{exp}} - f(r)_{\mathrm{theor}}$ (disagreement factor R_{f} = 3.07%).

Uracil, thymine and 5FU have nearly identical partial charge distributions in the H–N(3)–C(4)=O fragments. Because this fragment is docking to adenine in nucleic acids, 5FU might be able to surrogate a corresponding base.

This work was supported by the Dr. B. Mez-Starck Foundation (Germany).

Online Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi: 10.1016/j.mencom.2020.09.036.

References

1 A. Nyhlén, B. Ljungberg, I. Nilsson-Ehle and I. Odenholt, *Chemotherapy*, 2002, **48**, 71.

2 C. A. Bodet, 3rd, J. H. Jorgensen and D. J. Drutz, *Antimicrob. Agents Chemother.*, 1985, **28**, 437.

3 C. Heidelberger, N. K. Chaudhuri, P. Danneberg, D. Mooren, L. Griesbach, R. Duschinsky, R. J. Schnitzer, E. Pleven and J. Scheiner, *Nature*, 1957, **179**, 663.

4 D. B. Longley, D. P. Harkin and P. G. Johnston, *Nat. Rev. Cancer*, 2003, **3**, 330.

5 A. Argiris, D. J. Haraf, M. S. Kies and E. E. Vokes, *Onkologist*, 2003, **8**, 350.

6 H. G. Mandel, *Prog. Mol. Subcell. Biol.*, 1969, **1**, 82.

7 V. A. Naumov, S. A. Shlykov and A. V. Potanin, *Russ. J. Gen. Chem.*, 2009, **79**, 475 (*Zh. Obshch. Khim.*, 2009, **79**, 486).

8 Y. V. Vishnevskiy, *J. Mol. Struct.*, 2007, **833**, 30.

9 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao,

H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian, Wallingford, CT, 2009.

10 G. D. Purvis and R. J. Bartlett, *J. Chem. Phys.*, 1982, **76**, 1910.

11 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479.

12 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby and M. Schütz, *WIREs Comput. Mol. Sci.*, 2012, **2**, 242.

13 H.-J. Werner, P. J. Knowles, F. R. Manby, M. Schütz, P. Celani, G. Knizia, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A.W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni and T. Thorsteinsson, *MOLPRO*, 2009.

14 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618.

15 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.

16 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

17 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.

18 K. A. Peterson and T. H. Dunning, *J. Chem. Phys.*, 2002, **117**, 10548.

19 R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796.

20 V. A. Sipachev, *J. Mol. Struct.: THEOCHEM*, 1985, **121**, 143.

21 V. A. Sipachev, *Struct. Chem.*, 2000, **11**, 167.

22 V. A. Sipachev, *J. Mol. Struct.*, 2001, **567**–**568**, 67.

23 N. Vogt, J. Demaison, D. N. Ksenafontov and H. D. Rudolph, *J. Mol. Struct.*, 2014, **1076**, 483.

24 N. Vogt, L. S. Khaikin, O. E. Grikina and A. N. Rykov, *J. Mol. Struct.*, 2013, **1050**, 114.

25 L. S. Khaikin, N. Vogt, A. N. Rykov, O. E. Grikina, J. Vogt, I. V. Kochikov, E. S. Ageeva and I. F. Shishkov, *Mendeleev Commun.*, 2018, **28**, 236.

26 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr.*, 2016, **B72**, 171.

27 K. N. Jarzembska, M. Kubsik, R. Kamiński, K. Woźniak and P. M. Dominiak, *Cryst. Growth Des.*, 2012, **12**, 2508.

Received: 27th May 2020; Com. 20/6227