Molecular and Crystal Structure of an Arabino Nucleoside. X-Ray Analysis of 1-β-D-Arabinofuranosyl-4-thiouracil Monohydrate[1]

Wolfram Saenger

Contribution from the Max-Planck-Institut für experimentelle Medizin, Abteilung Chemie, Göttingen, Germany. *Received April 9, 1971*

Abstract: Arabino-4-thiouridine monohydrate crystallized in the monoclinic space group $P2_1$ with a = 4.787, b = 15.050, and c = 8.661 Å, and β = 73.0°. The structure was solved from 1444 three-dimensional diffractometer data using the heavy atom technique and refined to R = 4.6%. The conformation of the arabino nucleoside is similar to the conformation observed for ribonucleosides: it is in the preferred anti conformation, the pucker of the arabinose moiety is C(3′) endo and the conformation about the C(4′)–C(5′) bond is such that O(5′) is above the sugar. The contact between O(2′) and the pyrimidine heterocycle is closer than the sum of the van der Waals radii. Within the crystal structure there are hydrogen bonds involving all noncarbon atoms except O(1′) and the molecules are arranged in a fishbone pattern.

The arabino nucleosides ara-U[2] and ara-T were first discovered in sponges of the species *Cryptotethya crypta*[3,4] while ara-A was extracted from culture filtrates from *S. antibioticus*.[5] Ara-A and the synthetic ara-C[6] show cytocidal effects against tumor and bacterial cells and ara-A is active *in vitro* against a broad spectrum of DNA viruses.[7,8] Both ara-A and ara-C are inhibitors for DNA polymerase and are incorporated into DNA *in vitro* as terminal residues.[9] Furthermore, ara-A could replace the 3′-terminal A in a still chargeable tRNA.[10]

Polyarabinouridylic acid does not function as messenger under conditions where the riboisomer does and no hyperchromicity in the uv absorption was observed when polyadenylic acid and polyarabinouridylic acid were mixed.[11] That no complex formation occurred was interpreted as being due to the steric hindrance caused by the O(2′)–H group.

From spectroscopic investigations by means of nuclear magnetic resonance,[12] optical rotatory dispersion (ORD), and circular dichroism (CD)[13] it was concluded that arabino nucleosides generally should display anti[14] conformation in solution and that the rotation of the aglycone with respect to the sugar moiety should be inhibited by the O(2′)–H group.

Because of its characteristic absorption band at 320 mμ 1-β-D-arabinofuranosyl-4-thiouracil (Ara-s^{4}U (Figure 1)) was used in our institute instead of ara-U in order to facilitate biochemical investigations. Since, so far, only the crystal structure of arabino-5-bromouracil has been reported[15] and apart from the position of the O(2′)–H group ara-s^{4}U is structurally similar to the recently investigated 4-thiouridine,[16] a detailed structural analysis of ara-s^{4}U was deemed necessary.

Materials and Methods

A sample of ara-s^{4}U was provided by Dr. V. Jacobi. The nucleoside crystallized from aqueous solution as stout yellow plates in the monoclinic space group $P2_1$. The crystallographic data presented in Table I were derived from photographic and diffractometer measurements. The calculated crystal density is in agreement with the density determined by the flotation method if the asymmetric unit is assumed to consist of the monohydrate of the nucleoside.

Table I. Crystallographic Data

Space group $P2_1$, extinction $0k0$ for k odd
a = 4.787 ± 0.002 Å
b = 15.060 ± 0.004 Å
c = 8.661 ± 0.002 Å
β = 73.0 ± 0.04°

Chemical formula, $C_9H_{12}N_2O_5S \cdot H_2O$
Molecular weight, 278.3
Density obsd (C_6H_{12}–CH_3I) = 1.54 g/cm^3
Density calcd (Z = 2) = 1.57 g/cm^3
Linear absorption coefficient (Mo Kα) = 3.02 cm^{-1}
Crystal dimensions, 0.2 × 0.4 × 0.35 mm

1444 intensity data were collected by means of an automatic four-circle diffractometer using Zr-filtered Mo Kα radiation and measuring up to a glancing angle θ = 27.5° in the $2\theta,\omega$ scan mode. The data were corrected for Lorentz factor and polarization but not for

---

(1) Preliminary publication: W. Saenger and V. Jacobi, *Angew. Chem.*, **83**, 174 (1971).

(2) Abbreviations used are: ara-U and ara-T for spongouridine and -thymidine (1-β-D-arabinofuranosyluracil and -thymine), ara-A, ara-C, and ara-s^{4}U for 1-β-D-arabinofuranosyladenine, -cytosine, and -4-thiouracil, respectively.

(3) W. Bergmann and R. J. Feeney, *J. Amer. Chem. Soc.*, **72**, 2809 (1950).

(4) For a review on arabinosyl nucleosides see R. J. Suhadolnik, "Nucleoside Antibiotics," Wiley-Interscience, New York, N. Y., 1970, and S. S. Cohen, "Progress in Nucleic Acid Research and Molecular Biology," Vol. 5, J. N. Davidson and W. E. Cohn, Ed., New York, N. Y., 1966, p 1.

(5) Parke, Davis and Co., Patent No. 671,557 (1967).

(6) E. R. Walwick, W. K. Roberts, and C. A. Dekker, *Proc. Chem Soc.*, 84 (1959).

(7) F. M. Schabel, Jr., *Chemotherapy*, **13**, 321 (1968).

(8) See Cohen, ref 4.

(9) R. L. Momparler, *Biochem. Biophys. Res. Commun.*, **34**, 465 (1969).

(10) See Suhadolnik, ref 4, p 151.

(11) J. Nagyvary, R. Provenzale, and J. M. Clark, Jr., *Biochem. Biophys. Res. Commun.*, **31**, 508 (1968).

(12) R. J. Cushley, K. A. Watanabe, and J. J. Fox, *J. Amer. Chem. Soc.*, **89**, 394 (1967).

(13) W. Guschlbauer and M. Privat de Garilhe, *Bull. Soc. Chim. Biol.*, **51**, 1511 (1969).

(14) J. Donohue and K. N. Trueblood, *J. Mol. Biol.*, **2**, 363 (1960).

(15) P. Tougard, *Biochem. Biophys. Res. Commun.*, **37**, 961 (1969).

(16) W. Saenger and K. H. Scheit, *J. Mol. Biol.*, **50**, 153 (1970).

Table II. Atomic Positional and Thermal Parameters Expressed in Fractions of the Unit Cell Dimensions and in the Form $T = \exp -\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}l^2 + 2\beta_{12}hk + 2\beta_{13}hl + 2\beta_{23}kl)^a$

Atom	X	Y	Z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
N(1)	0.6178 (6)	−0.0650 (2)	0.2616 (3)	0.0233 (12)	0.0026 (1)	0.0075 (4)	−0.0005 (3)	−0.0007 (5)	0.0000 (2)
C(2)	0.7009 (7)	−0.0890 (2)	0.1033 (4)	0.0308 (16)	0.0030 (2)	0.0081 (4)	−0.0000 (4)	−0.0022 (7)	−0.0000 (2)
N(3)	0.5502 (6)	−0.1587 (2)	0.0630 (3)	0.0350 (14)	0.0032 (1)	0.0072 (3)	−0.0000 (4)	−0.0023 (6)	−0.0005 (2)
C(4)	0.3245 (7)	−0.2032 (2)	0.1672 (4)	0.0291 (15)	0.0021 (1)	0.0101 (5)	0.0001 (4)	−0.0062 (7)	0.0003 (2)
C(5)	0.2469 (7)	−0.1736 (2)	0.3281 (4)	0.0303 (15)	0.0030 (2)	0.0083 (4)	−0.0016 (4)	0.0003 (7)	0.0001 (2)
C(6)	0.3922 (7)	−0.1073 (2)	0.3703 (4)	0.0291 (15)	0.0029 (1)	0.0068 (4)	−0.0002 (4)	−0.0012 (7)	0.0004 (2)
O(2)	0.8951 (6)	−0.0507 (2)	0.0026 (3)	0.0478 (15)	0.0055 (2)	0.0089 (3)	−0.0065 (4)	0.0034 (6)	−0.0008 (2)
S	0.1573 (2)	−0.2870 (1)	0.1034 (1)	0.0432 (5)	0.0032 (0)	0.0113 (1)	−0.0029 (1)	−0.0048 (2)	−0.0012 (1)
C(1′)	0.7676 (7)	0.0121 (2)	0.3080 (4)	0.0225 (13)	0.0029 (1)	0.0082 (4)	−0.0014 (4)	−0.0024 (6)	−0.0002 (2)
C(1′)	0.7855 (5)	−0.0011 (2)	0.4658 (3)	0.0238 (10)	0.0032 (1)	0.0090 (3)	0.0010 (3)	−0.0046 (4)	−0.0003 (2)
C(2′)	0.5958 (7)	0.0987 (2)	0.3121 (4)	0.0285 (14)	0.0028 (1)	0.0090 (4)	−0.0017 (4)	−0.0032 (7)	0.0008 (2)
C(2′)	0.4237 (5)	0.1001 (2)	0.2063 (3)	0.0385 (12)	0.0043 (1)	0.0119 (4)	−0.0014 (4)	−0.0089 (6)	0.0021 (2)
C(3′)	0.4122 (7)	0.1022 (2)	0.4863 (4)	0.0247 (14)	0.0024 (1)	0.0104 (5)	0.0001 (4)	−0.0024 (7)	0.0001 (2)
C(3′)	0.3210 (5)	0.1906 (2)	0.5302 (3)	0.0438 (13)	0.0028 (1)	0.0138 (4)	0.0031 (3)	−0.0036 (6)	−0.0003 (2)
C(4′)	0.6235 (7)	0.0668 (2)	0.5722 (4)	0.0226 (13)	0.0030 (2)	0.0095 (5)	0.0004 (4)	−0.0036 (6)	−0.0005 (2)
C(5′)	0.4915 (8)	0.0272 (3)	0.7362 (4)	0.0411 (18)	0.0038 (2)	0.0095 (5)	−0.0002 (5)	−0.0061 (8)	−0.0003 (2)
C(5′)	0.3206 (6)	−0.0479 (2)	0.7232 (3)	0.0490 (16)	0.0047 (1)	0.0085 (4)	−0.0045 (4)	0.0025 (6)	0.0002 (2)
O	0.2798 (8)	0.3127 (2)	0.2702 (3)	0.0914 (23)	0.0045 (2)	0.0113 (4)	0.0019 (5)	−0.0062 (8)	0.0014 (2)
HN(3)	0.4079	0.3200	0.0414	0.0292	0.0027	0.0089	0.0000	0.0047	0.0000
HC(5)	−0.1014	0.2835	−0.4121	0.0324	0.0030	0.0099	0.0000	−0.0052	0.0000
HC(6)	−0.3906	0.4075	−0.4712	0.0251	0.0023	0.0077	0.0000	−0.0040	0.0000
HC(1′)	−0.0204	0.0145	0.2295	0.0249	0.0023	0.0076	0.0000	−0.0040	0.0000
HC(2′)	−0.2444	0.1462	0.2581	0.0284	0.0026	0.0087	0.0000	−0.0046	0.0000
HC(3′)	0.2400	0.0667	−0.4869	0.0290	0.0027	0.0089	0.0000	−0.0047	0.0000
HC(3′)	0.1133	0.1822	−0.3845	0.0414	0.0038	0.0126	0.0000	−0.0067	0.0000
HC(4′)	−0.2085	0.1170	−0.4337	0.0292	0.0027	0.0089	0.0000	−0.0047	0.0000
HC(5′)	0.3744	0.0780	−0.1930	0.0378	0.0035	0.0115	0.0000	−0.0061	0.0000
HC(5′)	−0.3848	0.0198	−0.2033	0.0378	0.0035	0.0115	0.0000	−0.0061	0.0000
HC(5′)	−0.1447	0.4541	0.1847	0.0403	0.0037	0.0123	0.0000	−0.0065	0.0000
HC(2′)	−0.4589	0.1311	0.1075	0.0414	0.0038	0.0126	0.0000	−0.0067	0.0000

[a] Estimated standard deviations are presented in parentheses.

absorption due to the small linear absorption coefficient of the material.

From a Wilson[17] plot an overall isotropic temperature factor ($B = 2.81$ Å^2) was derived and the data were put on an absolute scale. A sharpened Patterson synthesis with reduced origin served to locate the C(4)–S vector. From an electron density calculation based on this information all the nonhydrogen atoms of the aglycone could be located. After the nonhydrogen atoms of the sugar residue and water molecule were gathered from a Fourier synthesis phased with the scattering of the aglycone skeleton, the whole structure was refined in three cycles of isotropic least-squares refinement[18] to a reliability factor $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$ of 9%. The function minimized was $\Sigma w(|F_o| - |F_c|)^2$ where w is a weight based on counter statistics[19] and data with $F_o < 3/\sqrt{w}$ were considered unobserved. The hydrogen atoms bound to C and N but not to O could be traced from a difference Fourier synthesis computed at this stage and the remaining hydrogen atoms except those attached to the water oxygen atom were found after four cycles of anisotropic refinement had reduced the R factor to 6%. In a further refinement cycle the average parameter shifts were less than one-third the standard deviations estimated from the correlation matrix. The final R factor is 4.0% for the 1330 observed data and 4.6% for all 1444 data.

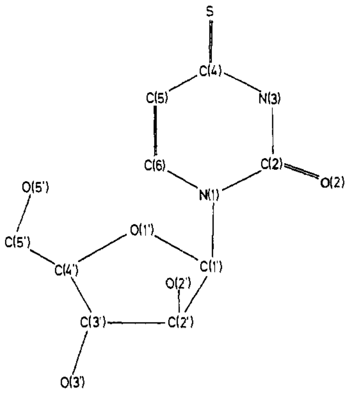


Figure 1. Chemical structure of ara-s^{4}U.

Results and Discussion

In Tables II and III[20] are listed the atomic parameters of ara-s^{4}U and the observed and calculated structure factors. The geometry of the molecule is described in Figure 2 and the thermal ellipsoids plot,[21] Figure 3, while in Tables IV and V some least-squares planes and dihedral angles necessary to describe the conformation of ara-s^{4}U are given. In Figure 5 a stereoscopic view of the packing of the molecules within the crystal lattice

---

(17) A. J. C. Wilson, *Nature* (*London*), 150, 152 (1942).

(18) W. R. Busing, K. O. Martin, and H. A. Levy, "A Fortran Crystallographic Least Squares Program," Oak Ridge National Laboratory Report No. TM-305, Oak Ridge, Tenn., 1962.

(19) G. H. Stout and L. H. Jensen, "X-Ray Structure Determination," Macmillan, New York, N. Y., 1968.

(20) Table III will appear following these pages in the microfilm edition of this volume of the journal. Single copies may be obtained from the Reprint Department, ACS Publications, 1155 Sixteenth St., N. W., Washington, D. C. 20036, by referring to author, title of article, volume, and page number. Remit $3.00 for photocopy or $2.00 for microfiche.

(21) C. K. Johnson, Oak Ridge National Laboratory Report No. BRNL-3794, Oak Ridge, Tenn., 1965.

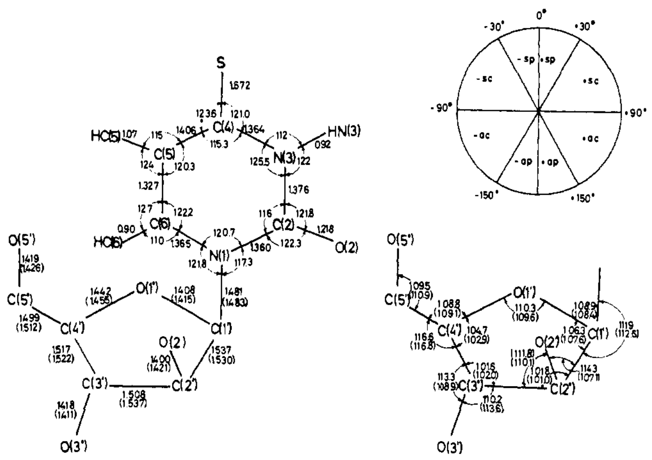


Figure 2. Intramolecular distances (ångströms) and angles (degrees) for ara-s^{4}U. The numbers in parentheses refer to the data obtained for the average C(3′) endo ribose unit.[32] The standard deviations σ for data not involving hydrogen atoms were estimated from the diagonal elements of the correlation matrix to be 0.003 Å for C(4)–S and 0.005 Å for the other bond lengths while bond angles are uncertain to 0.3°. Bond angles involving hydrogen atoms are 107 ± 6°, bond distances are 0.98 ± 0.07 Å. The insert describes the notation of conformational ranges[26] used in the text. The ranges ± synperiplanar, ± synclinal, ± antiperiplanar, and ± anticlinal are denoted as ±sp, ±sc, ±ap, and ±ac, respectively.

is represented, the hydrogen bonds being indicated by broken lines.

Table IV. Least-Squares Planes through Parts of Ara-S^{4}U and Distances of Some Atoms from These Planes[a]

Coefficients of plane equation	Distances, Å, of atoms from the plane	
	(a) Aglycone	
$A = 0.7360$	+N(1), 0.001	O(2), −0.054
$B = -0.6500$	+C(2), −0.014	S, 0.016
$C = 0.1891$	+N(3), 0.005	C(1′), −0.061
$D = -3.6408$	+C(4), 0.007	O(1′), 0.661
	+C(5), −0.011	C(2′), −1.504
	+C(6), 0.003	O(2′), −2.476
	(b) Four-Atom Sugar Plane	
$A = 0.8196$	+O(1′), 0.020	C(5′), −0.771
$B = 0.5394$	+C(1′), −0.018	O(2′), −0.708
$C = -0.1930$	+C(2′), 0.011	O(3′), −0.221
$D = -3.3382$	+C(4′), −0.012	O(5′), −2.084
	C(3′), −0.624	N(1), −1.274
		O, 0.493
	(c) Three-Atom Sugar Plane	
$A = 0.7942$	+C(4′), 0.0	O(2′), −0.587
$B = 0.5708$	+O(1′), 0.0	O(3′), −0.107
$C = -0.2083$	+C(1′), 0.0	C(5′), −0.793
$D = -3.1765$	C(2′), 0.091	N(1), −1.265
	C(3′), −0.554	

[a] The plane equations are of the form $AX + BY + CZ + D = 0$, where X, Y, and Z are measured in ångström units along the a axis, b axis and c^+, respectively. Atoms which define the planes are marked +. Angles between normals to the planes are: a,b = 77.5°, a,c = 76.7°, b,c = 2.5°.

Table V. Selected Dihedral Angles A–B–C–D within the Ara-s^4-U Molecule[a]

O(5′)–C(5′)–C(4′)–O(1′) = −55.5
O(5′)–C(5′)–C(4′)–C(3′) = 62.2
O(1′)–C(1′)–C(2′)–C(3′) = −27.5
C(1′)–C(2′)–C(3′)–C(4′) = 39.0
C(2′)–C(3′)–C(4′)–O(1′) = −38.3
C(3′)–C(4′)–O(1′)–C(1′) = 22.1
C(4′)–O(1′)–C(1′)–C(2′) = 3.5
N(1)–C(1′)–C(2′)–C(3′) = 90.9
N(1)–C(1′)–C(2′)–O(2′) = −28.5
N(1)–C(1′)–O(1′)–C(4′) = −116.9
C(1′)–C(2′)–C(3′)–O(3′) = 157.1
O(2′)–C(2′)–C(3′)–O(3′) = −81.3
O(2′)–C(2′)–C(3′)–C(4′) = 160.6
C(5′)–C(4′)–C(3′)–C(2′) = −158.3
C(5′)–C(4′)–C(3′)–O(3′) = 85.6
C(5)–C(4′)–O(1′)–C(1′) = 146.6
O(1′)–C(1′)–N(1)–C(6) = 36.0
C(2′)–C(1′)–N(1)–C(6) = −80.7

[a] The angles are defined as zero if the bond C–D is colinear with bond A–B when looking down the B–C bond. The angle is measured positive (degrees) if the far bond is rotated clockwise. The first two dihedral angles are ρ_{OO} and ρ_{OC},[25] the next five are the endocyclic sugar dihedral angles, and the last angle, τ_{CN}, defines the conformation about the glycosidic bond.

(a) Conformation of Ara-s^{4}U. Besides a few exceptions which were found to exhibit a half-chair conformation, ribose units in nucleosides in the crystalline state most frequently occur in an envelope form corresponding to four nearly coplanar and one out-of-plane atom.[22,23] The arabinofuranosyl residue in ara-s^{4}U is in a similar conformation and puckered such that atom C(3′) is protruding by 0.624 Å from the best plane through atoms C(4′), O(1′), C(1′), C(2′) (Table IV), and on the same side as atom C(5′), thus rendering the conformation C(3′) endo.[24] This pucker can also be described by the five endocyclic dihedral angles given in Table V, and by the deviations of atoms C(2′), C(3′), and O(3′) from the plane C(4′)–O(1′)–C(1′), Table IV; atoms C(2′) and C(3′) are 0.094 and 0.522 Å, respectively, away from this plane and on opposite sides. Due to the C(3′) endo conformation of the sugar residue, atom O(3′) is almost within the plane.

Since the (2′)- and (3′)-hydroxyl groups are trans and not cis as in ribose units the dihedral angle O(2′)–C(2′)–C(3′)–O(3′) is −81.3° (for the definition of the dihedral angle, see Table V) compared to +40 to +50° in ribose units, *i.e.* in ara-s^{4}U the bond C(3′)–O(3′) is rotated anticlockwise with respect to the bond C(2′)–O(2′). A similar value for the corresponding angle was observed in the C(3′) endo furanose ring of sucrose sodium bromide dihydrate[25] (cited in ref 23).

Furthermore, in ara-s^{4}U the dihedral angle N(1)–C(1′)–C(2′)–O(2′) is only −28.5° (− synperiplanar, −sp)[26] and atoms N(1) and O(2′) are almost staggered, while in ribonucleosides this angle is about −140° (−anticlinal, −ac)[26] and close to "trans." The proximity of atom O(2′) and the heterocycle should restrict the rotation of the heterocycle with respect to the sugar moiety, while in ribonucleosides, hydrogen atom HC(2′) acts as the main rotational barrier.[14,27] It was also concluded from the high molar rotation and rotational strength of arabino nucleosides toward optical rotatory dispersion (ORD) and circular dichroism (CD) that the rotation about the glycosidic bond is considerably limited.[13]

(22) M. Spencer, *Acta Crystallogr.*, **12**, 59 (1959).
(23) M. Sundaralingam, *J. Amer. Chem. Soc.*, **87**, 599 (1965).
(24) C. D. Jardetzky, *ibid.*, **82**, 229 (1960).
(25) C. A. Beevers and W. Cochran, *Proc. Roy. Soc. (London), Ser. A*, **190**, 257 (1947).
(26) W. Klyne and W. Prelog, *Experientia*, **16**, 521 (1960).
(27) A. E. V. Haschemeyer and A. Rich, *J. Mol. Biol.*, **27**, 369 (1969).

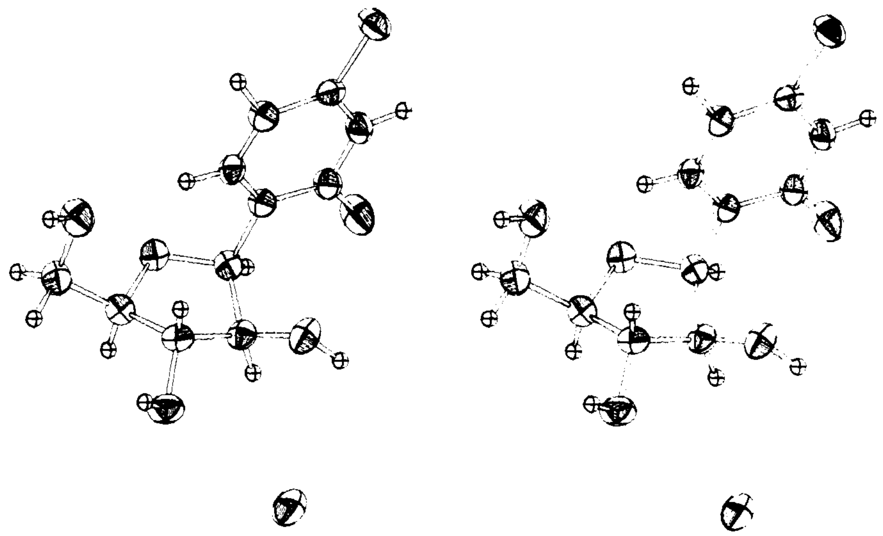


Figure 3. A stereoscopic view along a^* of the asymmetric unit showing the 50% probability thermal ellipsoids for the nonhydrogen atoms.

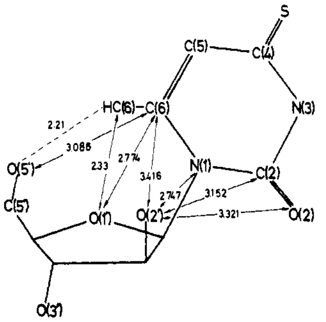


Figure 4. Intramolecular short nonbonded contacts. It should be noted that the distances O(2′)···N(1), O(1′)···C(6) and O(5′)···C(6) are shorter than the corresponding van der Waals contacts (O···N, 2.9 Å; O···aromatic system, 3.2 Å).

The position of atom O(5′) relative to the sugar ring is defined by φ_{OO} and φ_{OC}[28] (Table V), which are both gauche. Thus, atom O(5′) is "above" the sugar ring, a conformation occurring in most of the ribonucleosides investigated in the crystalline state.

Table VI. Geometrical Data for Hydrogen Bonds, A···H–D, and Close Intermolecular Contacts

Distances, A				Angles, deg		
A···D		A···H	H–D	A···D–H	A···H–D	D···A–X
S(x, y, z), O(2′) ($\bar{x} - 1, y - {}^1/_2, \bar{z}$)	3.311	2.32	0.99	5	173	114.8 (X = C(4))
S(x, y, z), O(3′) ($\bar{x}, y - {}^1/_2, \bar{z}$)	3.343	2.45	1.06	27	142	91.4 (X = C(4))
O(2) (x, y, z), O(5′) ($x + 1, y, z - 1$)	2.669	1.71	0.98	9	166	151.3 (X = C(2))
O($\bar{x}, y - {}^1/_2, \bar{z}$), N(3) ($x, y, z$)	2.792	1.90	0.92	12	162	
O(5′) ($\bar{x} + 1, y + {}^1/_2, \bar{z}$) O($x, y, z$)	2.852	close intermolecular contacts				100.3 (X = C(5′))
O(3′) (x, y, z), O(x, y, z)	2.959					117.6 (X = C(3′))
C(2′) (x, y, z), O(x, y, z)	3.288					167.3 (X = C(2′))

The conformation about the glycosidic bond is anti[14] as in all the pyrimidine nucleosides analyzed so far except 4-thiouridine.[16] The dihedral angle τ_{CN}, C(2′)–C(1′)–N(1)–C(6), −80.7°, is in the anti range 0 to −180° and the angle τ_{CN-90}, −170°[16], is −antiperiplanar (−ap).[26]

(b) Aglycone. The geometrical data of the thiopyrimidine heterocycle of ara-s⁴U bear resemblance to the data obtained for the 4-thiouracil residues in 4-thiouridine,[16] the 1-methyl-4-thiouracil–9-methyladenine base pair,[29] and the 4-thiothymidyl residue.[30] The S–C(4) and C(2)–O(2) bond lengths are characteristic for double bonds,[31] *i.e.*, the heterocycle is in the usual keto–keto form. The distribution of the exocyclic angles around N(1) and C(1′) is indicative of the anti conformation of the nucleoside.[16]

The six atoms comprising the heterocycle and the sulfur atom are fairly coplanar, Table IV, while atoms O(2) and C(1′) are deviating by 0.054 and 0.061 Å from the best six-atom plane.

(c) Arabinose Moiety. In Figure 2 the geometrical data for the arabinofuranosyl residue are presented together with data (in parentheses) averaged for several ribose moieties in the same (C(3′) endo) conformation.[32,33] Generally, the angles and distances are comparable to within 1.6° and 0.013 Å, but some differences are worth mentioning. Thus the bond distance C(2′)–C(3′) is decreased by 6 σ in ara-s⁴U (1.508 Å) with respect to ribose units (1.537 Å), and the C(3′)–O(3′) bond is longer than the C(2′)–O(2′) bond while in ribose units it is shorter (Figure 2). Furthermore, the exocyclic angles at C(2′) and C(3′) are different; the angles at C(3′) are of the same magnitude

---

(28) E. Shefter and K. N. Trueblood, *Acta Crystallogr.*, 18, 1067 (1965), and ref 23.

(29) W. Saenger and D. Suck, *J. Mol. Biol.*, 60, 87 (1971).

(30) W. Saenger and D. Suck, *FEBS (Fed. Eur. Biochem. Soc.) Lett.*, 5, 262 (1969).

(31) W. Saenger and D. Suck, *Acta Crystallogr., Sect. B*, 27, 1178 (1971).

(32) W. Saenger and F. Eckstein, *J. Amer. Chem. Soc.*, 92, 4712 (1970).

(33) M. Sundaralingam and L. H. Jensen, *J. Mol. Biol.*, 13, 930 (1965).

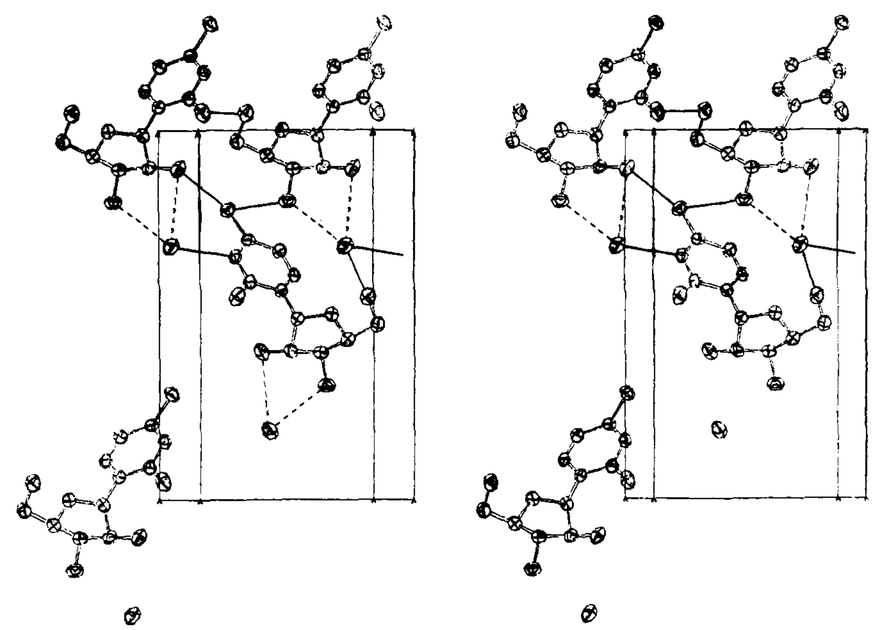


Figure 5. Packing diagram of the ara-s⁴U molecules within the crystal structure viewed along *a**. Hydrogen bonds and short contacts to the water molecule are indicated by broken and full lines, respectively; distances and angles are given in Table VI.

in ara-s⁴U and the averaged ribose unit but in reversed order, and the angle C(1′)–C(2′)–O(2′), 114.3°, is increased in ara-s⁴U by 7.2° compared to the 107.1° in ribose units while C(3′)–C(2′)–O(2′) is similar in both sugars.

In two other structures containing "arabinofuranosyl" rings, raffinose[34] and ethyl 1-thio-α-D-glucofuranoside,[35] the angles corresponding to C(1′)–C(2′)–O(2′) in ara-s⁴U are 111.2 and 107.8°, respectively. Thus, the opening of this angle by 7.2° in ara-s⁴U could indicate a steric interaction between atoms O(2′) and N(1) of the heterocycle which is only 2.747 Å apart, Figure 4, while the van der Waals distance is 2.9 Å. However, it should be noted that the angle C(2′)–C(1′)–N(1) is not increased but close to the average 112.6° derived for C(2′) endo ribonucleosides and that similar exocyclic angles at C(2′) as in ara-s⁴U—but in reversed order—were observed in C(2′) endo ribose units.[32]

It is striking that in the arabinose moiety the exocyclic angles at C(2′) and C(3′) are distributed symmetrically and pairwise similarly. Thus, C(1′)–C(2′)–O(2′), 114.3°, corresponds to C(4′)–C(3′)–O(3′), 113.3°, and C(3′)–C(2′)–O(2′), 111.8°, corresponds to C(2′)–C(3′)–O(3′), 110.2°, while in ribose moieties the distribution of these angles is asymmetric probably due to a steric interaction of the cis hydroxyl groups. The same steric effects might be responsible for the above-mentioned differences in bond distances involving atoms C(2′) and C(3′).

(d) Crystal Structure. The packing arrangement of the ara-s⁴U molecules is depicted in the stereoscopic view, Figure 5. By the symmetry operation of the twofold screw axis a typical fishbone pattern of molecules is generated. Along the *a* axis the heterocycles are arranged parallel to each other, about 3.4 Å apart and tilted against this axis by 42°. Thus the heterocycles are "stacked" but the mutual overlap is only through atoms N(1), C(2), O(2), and N(3), and C(5), C(4), N(3), and S (Figure 6). Similar packing arrangements have been observed in other pyrimidine nucleoside crystal structures.[36]

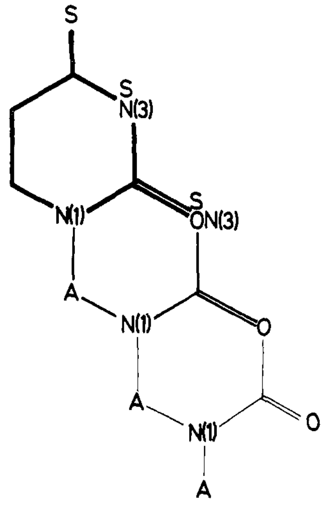


Figure 6. Schematic drawing of the overlap of successive 4-thiouracil residues in the "base-stack" when viewed approximately along the normals to the heterocyclic planes.

Virtually every noncarbon heavy atom except O(1′) and N(1) is involved in hydrogen bonding. The water molecule is in hydrogen bonding contact with atoms N(3), 2.794 Å, and O(5′), 2.852 Å (Table V), and in close contact with atoms O(3′), 2.959 Å, and O(2′), 3.288 Å, *i.e.*, it is surrounded by oxygen and nitrogen atoms which are arranged in a flat, distorted tetrahedron.

The sulfur atom is the acceptor for two hydrogen atoms, from O(2′), 3.311 Å, and O(3′), 3.343 Å, while O(2) is involved in an O(2)···H–O(5′) bond of 2.669-Å length.

---

(34) H. Berman, *Acta Crystallogr., Sect. B*, 26, 290 (1970).

(35) R. Parthasarathy and R. E. Davis, *ibid.*, 23, 1049 (1967).

(36) C. E. Bugg, J. M. Thomas, M. Sundaralingam, and S. T. Rao, *Biopolymers*, 10, 175 (1971).

When the crystal structures of the analogs ara-s^{4}U, 4-thiouridine,[16] the 1-methyl-4-thiouracil–9-methyladenine complex,[29] and 3′-*O*-acetyl-4-thiothymidine[30] are compared with each other it becomes evident that the hydrogen bonding of the heterocycles in these structures is largely determined by the packing arrangement: in 4-thiouridine and in the base–pair complex where the heterocycles are markedly stacked, O(2) and N(3) but not S are involved in hydrogen bonding, while in 3′-*O*-acetyl-4-thiothymidine, S but not O(2) and N(3) forms a hydrogen bond. In ara-s^{4}U, however, O(2), N(3), and S do have a hydrogen-bonded partner.

The computations were carried out at the Aerodynamische Versuchsanstalt Göttingen (IBM 7040) and at the Gesellschaft für wissenschaftliche Datenverarbeitung mbH., Göttingen (UNIVAC 1108). The thermal ellipsoids plot was performed at Deutsches Rechenzentrum Darmstadt (IBM 7094).

Acknowledgment. The author is indebted to Dr. V. Jacobi for supplying the crystals for this investigation and to Professor F. Cramer for his support of and interest in this work. The valuable advice of Dr. B. Krebs, Universität Göttingen, during the progress of this work is gratefully acknowledged.

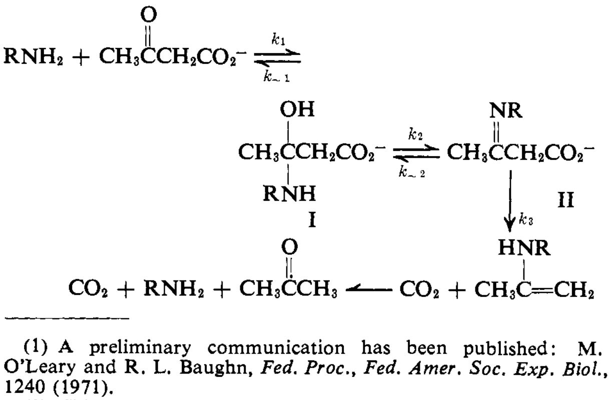
Acetoacetate Decarboxylase. Identification of the Rate-Determining Step in the Primary Amine Catalyzed Reaction and in the Enzymic Reaction[1,2]

Marion H. O'Leary* and Richard L. Baughn

Contribution from the Department of Chemistry, University of Wisconsin,
Madison, Wisconsin 53706. *Received April 12, 1971*

Abstract: The decarboxylation of acetoacetic acid in aqueous solution at 30° catalyzed by aminoacetonitrile shows carboxyl carbon isotope effects (k^{12}/k^{13}) of 1.031 at pH 3.6, 1.032 at pH 4.1, and 1.036 at pH 5.0. The magnitude of the isotope effect and the variation of the isotope effect with pH indicate that decarboxylation is partially rate determining. Quantitative evaluation of all of the rate constants in the mechanism is possible by assuming a value for the isotope effect on the decarboxylation step and calculating individual rate ratios from the pH dependence of the isotope effect. The decarboxylation catalyzed by acetoacetate decarboxylase shows isotope effects of 1.019 at pH 5.3, 1.017 at pH 6.0, and 1.019 at pH 7.2. As in the case of the amine-catalyzed reaction, decarboxylation is partially rate determining, but no single step is entirely rate determining.

The amine-catalyzed decarboxylation of acetoacetic acid proceeds according to the general mechanism[3–6] of eq 1. The first two steps in the reaction,



$$\mathrm{RNH_2 + CH_3\overset{\overset{\displaystyle O}{\|}}{C}CH_2CO_2^- \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \underset{\displaystyle\mathrm{I}}{CH_3\overset{\overset{\displaystyle OH}{|}}{\underset{\underset{\displaystyle RNH}{|}}{C}}CH_2CO_2^-} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \underset{\displaystyle\mathrm{II}}{CH_3\overset{\overset{\displaystyle NR}{\|}}{C}CH_2CO_2^-} \xrightarrow{k_3} CO_2 + CH_3\overset{\overset{\displaystyle HNR}{|}}{C}{=}CH_2 \longrightarrow CO_2 + RNH_2 + CH_3\overset{\overset{\displaystyle O}{\|}}{C}CH_3} \qquad (1)$$

formation of the carbinolamine (I) and of the Schiff base (II), are well known, and have been studied in detail in many cases not involving decarboxylation.[7] However, in spite of several recent studies,[4–6] the details of the mechanism of decarboxylation are unclear. In particular, little is known about the relative rates of the various steps in eq 1.

The decarboxylation of acetoacetic acid catalyzed by acetoacetate decarboxylase from *Clostridium acetobutylicum* also proceeds according to the general mechanism[8] of eq 1. Decarboxylation occurs *via* a Schiff base between acetoacetate and the ε-amino group of a lysine residue of the enzyme.[9] This particular amino group is abnormal in that it is the only amino group acetylated by acetic anhydride[10] at pH 6, and kinetic studies[11] indicate that it has an unusually

---

(1) A preliminary communication has been published: M. H. O'Leary and R. L. Baughn, *Fed. Proc., Fed. Amer. Soc. Exp. Biol.*, **30**, 1240 (1971).

(2) This research was supported by National Institutes of Health Grant No. NS 07657 and by a grant from the Research Committee of the University of Wisconsin Graduate School.

(3) K. J. Pedersen, *J. Amer. Chem. Soc.*, **51**, 2098 (1929); H. von Euler and A. Ōlander, *Z. Anorg. Chem.*, **147**, 295 (1925); E. Widmark and C. A. Jeppsson, *Skand. Arch. Physiol.*, **42**, 43 (1922); E. Widmark and C. A. Jeppsson, *Chem. Z.*, **93**, 526 (1922); K. J. Pedersen, *J. Phys. Chem.*, **38**, 559 (1934); K. J. Pedersen, *J. Amer. Chem. Soc.*, **60**, 595 (1938).

(4) R. W. Hay, *Austr. J. Chem.*, **18**, 337 (1965).

(5) J. P. Guthrie and F. H. Westheimer, *Fed. Proc., Fed. Amer. Soc. Exp. Biol.*, **26**, 562 (1967).

(6) J. P. Guthrie, Ph.D. Dissertation, Harvard University, Cambridge, Mass., 1968.

(7) W. P. Jencks, "Catalysis in Chemistry and Enzymology," McGraw-Hill Book Co., New York, N. Y., 1969, Chapter 10; T. C. Bruice and S. J. Benkovic, "Bioorganic Mechanisms," Benjamin, New York, N. Y., 1966, Vol. 2, Chapter 8.

(8) F. H. Westheimer, *Proc. Chem. Soc.*, 253 (1963); G. A. Hamilton and F. H. Westheimer, *J. Amer. Chem. Soc.*, **81**, 6332 (1959); I. Fridovich and F. H. Westheimer, *ibid.*, **84**, 3208 (1962).

(9) S. Warren, B. Zerner, and F. H. Westheimer, *Biochemistry*, **5**, 817 (1966); R. A. Laursen and F. H. Westheimer, *J. Amer. Chem. Soc.*, **88**, 3426 (1966).

(10) M. H. O'Leary and F. H. Westheimer, *Biochemistry*, **7**, 913 (1968).

(11) (a) S. M. Coutts, Ph.D. Dissertation, Harvard University, Cambridge, Mass., 1967; (b) D. E. Schmidt and F. H. Westheimer, *Biochemistry*, **10**, 1249 (1971).