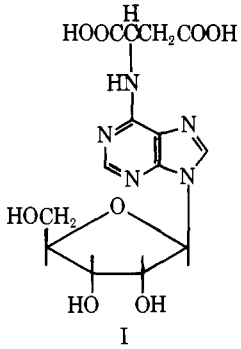
Model Compounds for Protein Nucleic Acid Interactions. 3. Synthesis and Structure of a Nucleoside Dipeptide, *N*-(9-β-D-Ribofuranosylpurin-6-yl)glycyl-L-alanine Sesquihydrate

Poojappan Narayanan, Helen M. Berman,*[1] and Robert Rousseau

Contribution from The Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, and ICN Nucleic Acid Research Institute, Irvine, California 92664. *Received May 21, 1976*

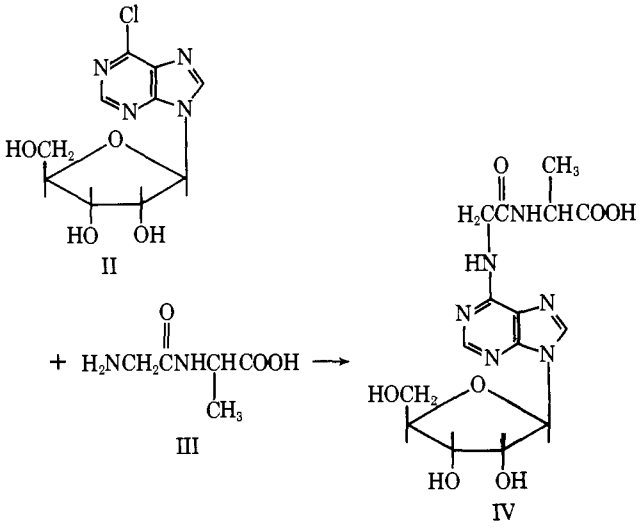
Abstract: The synthesis and crystal structure determination of a nucleoside dipeptide, *N*-(9-β-D-ribofuranosylpurin-6-yl)glycyl-L-alanine, are reported. The cell dimensions are $a = 17.535$, $b = 4.879$, and $c = 22.414$ Å; $\beta = 107.2°$. The space group is $A2$ with four molecules in the cell. The isotropic refinement converged to an R value of 0.135. Unlike other N(6)-substituted adenine structures, there are no intramolecular hydrogen bonds. The carboxyl group of one molecule hydrogen bonds to the N(6) and N(7) of the adenine ring. This provides a model for a way in which adenine, base paired to a uracil, might interact with a carboxyl functional group of glutamic or aspartic acid.

Nucleoside and nucleotide peptides occur naturally in tRNA[2] and as intermediates in biochemical pathways. An example of the latter, adenylosuccinic acid, is an intermediate in the biochemical interconversion of inosinic and adenylic acid.[3] The aglycon, *N*-(purin-6-yl)aspartic acid (I), is present



I

on the myceleum of penicillium chrysogenym.[4] Other *N*-(purin-6-yl)amino acids[5,6] have been synthesized and tested as possible antimetabolites and antithrombotic agents.[7] By coupling a nucleoside (II) with a dipeptide (III), a nucleoside dipeptide, *N*-(purin-6-yl) peptide, *N*-(9-β-D-ribofuranosylpurin-6-yl)glycyl-L-alanine (IV), is formed.



II

III

IV

As part of a program of study to elucidate new amino acid-base interactions,[8-10] we present both the synthesis and the crystal structure of this *N*-(purin-6-yl) peptide.

Experimental Section

Synthesis of *N*-(9-β-D-Ribofuranosylpurin-6-yl)glycyl-L-alanine (IV). 6-Chloro-9-(9-β-D-ribofuranosyl)purine (500 mg), anhydrous sodium carbonate (185 mg), and glycyl-L-alanine[11] (510 mg) were heated at 80 °C for 5 h in a mixture of 5 ml of H_2O and 5 ml of dimethylformamide. The solution was allowed to cool to room temperature, added to a Dowex 1-X8 (acetate form 50 ml), and subsequently washed with 4 l. of water and then 2 l. of 2 N acetic acid. Fractions (20 ml) were collected when the acetic acid elution began and fractions 4-60 (ninhydrin-negative, UV positive) were combined and evaporated to a glass in vacuo. This was triturated with ethyl acetate (10 ml) to yield 590 mg (85%) of colorless powder. A small portion of this material was crystallized from a mixture of methanol and ethyl acetate and then recrystallized from methanol to afford an analytical sample, mp 190 °C (sinters), after being dried at 0.1 mm at room temperature for 4 h: $[\alpha]^{25}D$ 68.2° (*c* 1, H_2O). Anal. Calcd for $C_{15}H_{20}N_6O_9$: C, 45.45; H, 5.09; N, 21.20. Found: C, 45.55; H, 5.40; N, 21.02.

The physical properties of these compounds were determined with the following instruments: melting point, Thomas-Hoover apparatus (uncorrected); specific rotation, Perkin-Elmer Model 141 polarimeter; UV spectra, Cary 15 UV spectrometer (see Table I). Elemental analyses were performed by Galbraith Laboratories, Garden City, Mich.

Data Collection and Structure Determination. Compound IV, $C_{15}H_{20}N_6O_7 \cdot 1.5H_2O$, was crystallized by evaporation from a water-methanol (9:1) solution. The crystals were very small; the largest was a needle 0.03 × 0.04 × 0.1 mm. The cell dimensions, as determined on a diffractometer, are $a = 17.535$ (6); $b = 4.879$ (1); $c = 22.414$ (10) Å; $\beta = 107.2°$ (3). The space group is $A2$ with one molecule in the asymmetric unit. Three-dimensional data were collected by the $\theta/2\theta$ scan method on a $P\bar{1}$ diffractometer, using Cu Kα ($\lambda = 1.5418$ Å) radiation. Because of the extremely small size of the crystal, only 613 of the 1582 reflections collected had intensities greater than $3\sigma(I)$.

A partial structure was derived utilizing the program, MULTAN.[12] The rest of the structure was determined using Fourier methods. The hydrogen atoms attached to C(2), C(8), N(6), O(2′), O(5′), and O(CBX1) were found in difference Fourier maps. The ones belonging to the methyl group, O(W1) and O(3′), were ambiguous. The other positions belonging to hydrogen atoms of the sugar ring and peptide were calculated. All of these were included in the structure factor calculations. The positional and isotropic thermal parameters of the heavy atoms were refined to an R of 0.135 using full-matrix least squares. The occupancy which is ideally 0.5 of O(W2) was refined to a value of 0.4. Weights were assigned from an analysis of variance: $9.0 < F_o < 23.0$, $\sigma = 3.3 + {}^1/_{13}(23.0 - F_o)$; $23.0 \leq F_o < 50.0$, $\sigma = 3.3$;

Table I. Ultraviolet Spectra

Buffer pH	λ max	ε max	λ min	ε min
pH 1	207	21.06	233	4.05
	264	17.01		
pH 11	265	18.63		
MeOH	212	14.99	229	1.62
	265	19.04		

Table II. Positional and Thermal Parameters of the Atoms in *N*-(9-β-D-Ribofuranosylpurin-6-yl)glycyl-L-alanine-Water (1/1.5)[a]

	X	*Y*	*Z*	*B*
O(1′)	318 (1)	−564 (5)	363 (1)	2.4 (2)
C(1′)	358 (1)	−412 (6)	416 (1)	2.3 (3)
C(2′)	426 (1)	−265 (5)	400 (1)	2.2 (3)
C(3′)	396 (1)	−234 (5)	332 (1)	2.1 (3)
C(4′)	353 (1)	−505 (5)	314 (1)	1.6 (3)
C(5′)	290 (1)	−511 (6)	251 (1)	3.8 (4)
O(2′)	493 (1)	−436 (5)	416 (1)	2.6 (2)
O(3′)	455 (1)	−182 (5)	303 (1)	2.9 (2)
O(5′)	230 (1)	−319 (5)	250 (1)	4.4 (3)
N(1)	321 (1)	268 (5)	579 (1)	2.3 (3)
C(2)	378 (1)	91 (6)	579 (1)	2.8 (4)
N(3)	382 (1)	−89 (5)	534 (1)	2.6 (3)
C(4)	316 (1)	−70 (5)	483 (1)	1.9 (3)
C(5)	253 (1)	108 (5)	479 (1)	3.0 (4)
C(6)	257 (1)	283 (*)[b]	528 (1)	2.7 (3)
N(7)	197 (1)	69 (5)	419 (1)	2.1 (3)
C(8)	230 (1)	−122 (5)	394 (1)	1.5 (3)
N(9)	302 (1)	−211 (5)	430 (1)	1.7 (2)
N(6)	198 (1)	464 (5)	527 (1)	2.7 (3)
C(A1)	196 (1)	650 (5)	579 (1)	2.5 (3)
C(11)	136 (1)	542 (6)	608 (1)	2.2 (3)
O(11)	134 (1)	303 (5)	625 (1)	4.1 (3)
N(PEP)	92 (1)	737 (5)	626 (1)	2.6 (3)
C(A2)	35 (1)	676 (6)	659 (1)	3.1 (4)
C(12)	67 (1)	585 (6)	725 (1)	3.5 (4)
C(CBX)	−31 (1)	495 (6)	616 (1)	3.3 (4)
O(CBX1)	−72 (1)	368 (5)	646 (1)	4.5 (4)
O(CBX2)	−48 (1)	505 (5)	561 (1)	5.5 (3)
O(W1)	401 (1)	566 (5)	691 (1)	3.6 (3)
O(W2)	0	1007 (23)	500	22.2 (33)
H(1′)	380	−537	453	3.0
H(2′)	436	−81	422	2.5
H(3′)	354	−78	321	2.5
H(4′)	395	−646	315	2.0
H(5′1)	266	−700	244	3.7
H(5′2)	316	−471	217	3.7
H(O2′)	517	−469	444	3.0
H(O3′)	442	−187	259	3.5
H(O5′)	180	−375	212	4.5
H(2)	426	86	618	3.2
H(8)	205	−176	349	2.0
H(N6)	152	467	487	3.0
H(A11)	181	841	559	3.0
H(A12)	251	655	608	3.0
H(PEP)	107	931	617	3.0
H(A2)	7	861	661	3.5
H(121)	23	556	745	4.0
H(122)	103	741	750	4.0
H(123)	100	422	728	4.0
H(OCBX1)	−108	305	633	5.0
H(W11)	464	558	671	4.0
H(W12)	420	525	740	4.0

[a] The fractional coordinates of all atoms are multiplied by 10^3. [b] This parameter was held constant during refinement.

$50.0 \leq F_o \leq 125.0, \sigma = 3.3 + {}^1/_{45}(F - 50.0)$; $F_o > 125.0, \sigma = 14.0$. The scattering factors for O, N, and C were from the International Tables for X-ray Crystallography[13] and for the H atoms from Stewart et al.[14] Attempts at anisotropic refinement were abandoned because the data to parameters ratio were too small (4.3). The coordinates of the atoms and their temperature parameters are given in Table II. The structure factors are deposited on microfilm. (See paragraph at end of paper regarding supplementary material.)

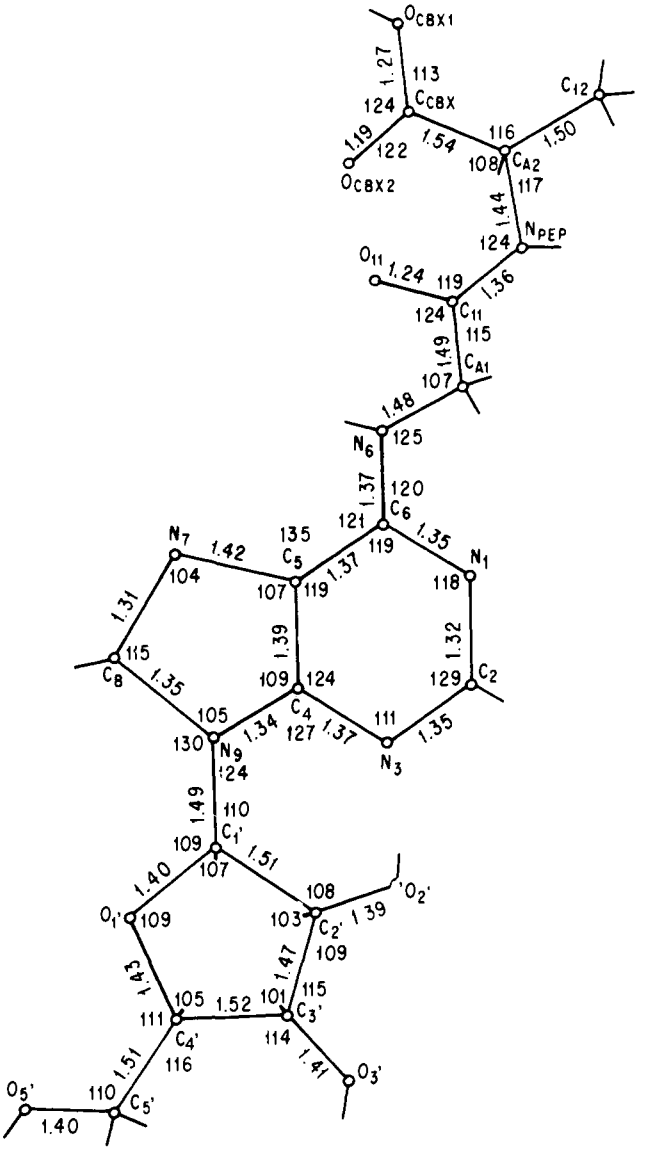


Figure 1. Distances, angles, and atomic numbering of compound IV, *N*-(9-β-D-ribofuranosylpurin-6-yl)glycyl-L-alanine. The average standard deviation for the bond distances is 0.025 Å and for the angles 1.5°.

Discussion

Description of the Molecule. The compound, IV, is the dipeptide glycylalanine covalently linked to the N(6) position of adenosine (Figure 1). Each part of the molecule will be described as well as the manner in which it is linked. The distances and angles are shown in Figure 1. The torsion angles and molecular planes are given in Table III. The ribose ring is in the C(3′)-endo-C(2′)-exo conformation ($^3T_2$);[15] the atom C(3′) is displaced by 0.6 Å from the plane defined by atoms C(4′), O(1′), C(1′), C(2′). The C(4′)-C(5′) bond is in the g-g conformation. The value for the C(2′)-C(3′) bond length is small (1.47 Å), but it does not deviate more than 3σ from the average value found in ribose sugars (1.52 Å). The adenine ring is planar and has "normal" bond lengths and angles.[16] It is "anti" to the ribose ring.[17] As shown in Table III, the conformation of the glycylalanine portion of the molecule is qualitatively the same as that found for glycylalanine hydrochloride;[18] there are quantitative differences. The ψ_1 value in this structure (−142°) deviates significantly from the value of

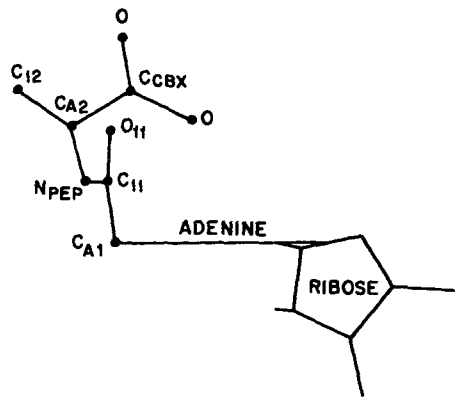
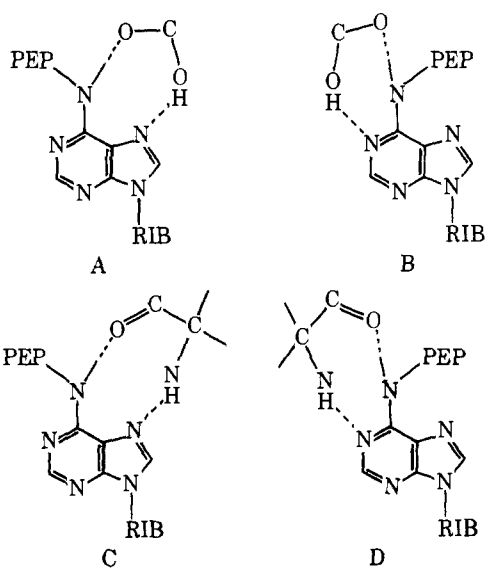


Figure 2. The conformation of the molecule emphasizing the relationship among the molecular planes.

−174° which resulted in a completely extended conformation for glycylalanine. The dihedral angle (defined as the angle between the normals) between the carboxyl and peptide planes is 76.5° as compared with the value of 99.2° for glycylalanine. The dihedral angle between the adenine ring and peptide plane is 91°; that is, the adenine ring is almost perpendicular to the peptide plane. This arrangement is in contrast to the structure of *N*-(purin-6-ylcarbamoyl)-L-threonine,[19] where an intramolecular hydrogen bond allows the amino acid side chain to be coplanar with the adenine. Figure 2 shows the relationship of the different parts of the molecule.

Hydrogen Bonding. Example of an Amino Acid–Nucleic Acid Base Interaction. One rationale for studying structures of nucleoside peptides is to elucidate patterns of interaction between the peptide of one molecule and the nucleic acid base of another crystallographically related molecule. If we consider the *N*-glycylalanine in this structure, there are one nitrogen [N(PEP)] and three oxygen atoms [O(11), O(CBX1), O(CBX2)] capable of forming hydrogen bonds. The sites available for hydrogen bonding on the adenine are N(7), N(6), and N(1). However, there are only two possible *double* hydrogen bonds.[20] It is possible for the N(PEP) to donate and the O(CBX2) to accept hydrogen atoms to and from two sites simultaneously (C and D). The same is true of the protonated carboxyl group (A and B). Thus, we have the following possible geometries.



Unlike N^6-(purin-6-ylcarbamoyl)threonine,[19] no atom of the adenine ring is shielded by an intramolecular hydrogen bond. Close contact between N(7) and the C(A1) eliminates B and

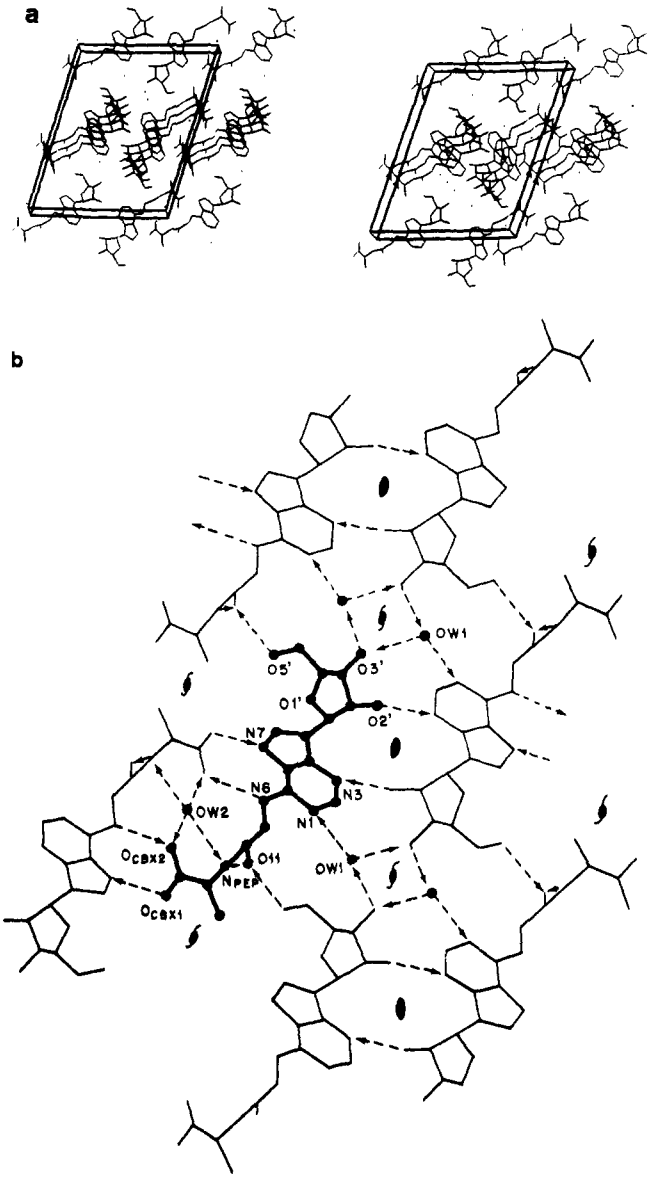


Figure 3. The hydrogen bonding: (a) a stereoview; (b) the surroundings of one molecule.

D.[21,22] In this structure, the conformation of the side chain is such that the N(7) and N(6) sites of the adenine are available for hydrogen bonding. While positions of the hydrogen atoms in this structure are ambiguous, consideration of the bond angles and distances of the carboxyl and adenine groups as well as of the hydrogen bonding scheme leads to the conclusion that the adenine is not protonated and the carboxyl has a hydrogen atom on O(CBX1). N(6) of the adenine donates to one carboxyl oxygen O(CBX2) and N(7) receives a hydrogen bond from the other carboxyl oxygen atom O(CBX1). Thus we have observed alternative A. This interaction is significant because it presents a model for the way in which an adenine, base paired via N(6) and N(1) to a uracil in double-stranded RNA or not base paired at all as in parts of tRNA or in ATP alone, might hydrogen bond with a carboxyl group of glutamic or aspartic acid. Another example of this kind of hydrogen bond is shown in the structure of the glycine–cytosine complex.[23]

In addition to these hydrogen bonds which are in the *ac* plane, there is an N(PEP) → O(11) hydrogen bond which links the dipeptides so they form infinite sheets in the *b* direction. The crystal structure is also held together by hydrogen bonds between the O(2') of a ribose and the N(3) of an adenine, be-

Table III. Torsion Angles and Planes

	Atoms	Degree
	Ribose Sugar	
ψ_{OO}	O(1′)-C(4′)-C(5′)-O(5′)	+59
ψ_{OC}	C(3′)-C(4′)-C(5′)-O(5′)	−61
χ	O(1′)-C(1′)-N(9)-C(8)	−16
τ_0	C(4′)-O(1′)-C(1′)-C(2′)	−4
τ_1	O(1′)-C(1′)-C(2′)-C(3′)	+28
τ_2	C(1′)-C(2′)-C(3′)-C(4′)	−38
τ_3	C(2′)-C(3′)-C(4′)-O(1′)	+36
τ_4	C(3′)-C(4′)-O(1′)-C(1′)	−20
	Dipeptide[b]	
	N(1)-C(6)-N(6)-C(A1)	0
ϕ_1	C(6)-N(6)-C(A1)-C(11)	−107
ψ_1	N(6)-C(A1)-C(11)-N(PEP)	−142 (−173.9)[a]
ω	C(A1)-C(11)-N(PEP-C(A2)	−176 (+169.3)
ϕ_2	C(11)-N(PEP)-C(A2)-C(CBX)	−62 (−72.4)
ψ_T^1	N(PEP)-C(A2)-C(CBX)-O(CBX2)	−28 (−1.0)
ψ_T^2	N(PEP)-C(A2)-C(CBX)-O(CBX1)	+161 (+179.8)

Planes

Ribose sugar		Adenine ring	
Atom	Distance from plane, Å	Atom	Distance from plane, Å
O(1′)	−0.02	N(1)	+0.01
C(1′)	+0.02	C(2)	+0.01
C(2′)	−0.01	N(3)	−0.01
C(4′)	+0.02	C(4)	−0.01
C(3′)	+0.59	C(5)	−0.02
C(5′)	+0.78	C(6)	+0.02
		N(7)	−0.01
Equation of plane:		C(8)	+0.01
$0.543x - 0.783y + 0.303z - 6.21 = 0$		N(9)	+0.02
		Equation of plane:	
		$0.602x + 0.697y - 0.390z + 2.85 = 0$	

Peptide linkage		Carboxyl group	
Atom	Distance from plane, Å	Atom	Distance from plane, Å
C(A1)	+0.03	C(A2)	+0.01
C(11)	−0.07	C(CBX)	−0.05
O(11)	+0.02	O(CBX1)	+0.02
N(PEP)	+0.00	O(CBX2)	+0.02
C(A2)	+0.01	Equation of plane:	
Equation of plane:		$0.639x - 0.766y + 0.067z + 3.97 = 0$	
$0.457x + 0.142y + 0.878z - 11.12 = 0$			

[a] Numbers in parentheses correspond to values for glycylalanine.[18] [b] Nomenclature conforms to IUPAC-IUB rules.[24]

tween the O(5′) of the sugar and the O(11) of the peptide, and between the water molecules and several atoms in the structure. There is an infinite hydrogen bonding helix formed by O(3′) and O(W1); the direction of the hydrogen bonding is not clear. The O(W2) is on the twofold axis and hydrogen bonds with six atoms—two O(CBX2) atoms, one N(PEP) atom, and their twofold related atoms. It is possible that this atom is an artifact. However, its absence would create a large hole in the structure.

Table IV. Hydrogen Bond Distances

Donor	Acceptor	Distance, Å	Symmetry
N(6)	O(CBX2)	2.78	$-x,y,-z+1$
O(CBX1)	N(7)	2.68	$-x,y,-z+1$
O(2′)	N(3)	2.74	$-x+1,y,-z+1$
O(5′)	O(11)	2.86	$x,y-\frac{1}{2},z-\frac{1}{2}$
N(PEP)	O(11)	2.86	$x,y+1,z$
O(W1)	N(1)	2.87	x,y,z
O(W1)	O(3′)	2.78	$-x+1,y+1,-z+1$
O(3′)	O(W1)	2.69	$x,y-\frac{1}{2},z-\frac{1}{2}$
O(W2)	N(PEP)	3.10	$x,y,z;\ -x,y,-z+1$
O(W2)	O(CBX2)	3.03	$x,y,z;\ -x,y,-z+1$
O(W2)	O(CBX2)	3.02	$x,y+1,z;\ -x,y+1,-z+1$

Finally, it is noted that the methyl side chain of the alanine is in close contact with two oxygens, O(CBX1) and O(5′) of symmetry-related molecules (Figure 3, Table IV).

Acknowledgments. We thank Roland K. Robins for his help and interest. The help of Deborah Marcu is gratefully acknowledged.

Supplementary Material Available: a listing of observed and calculated structure factors (9 pages). Ordering information is given on any current masthead page.

References and Notes

(1) Address correspondence to this author at The Institute for Cancer Research, Philadelphia, Pa. 19111. This research was supported by U.S. Public Health Service Grants GM-21589, CA-06927, and RR-05539 from the National Institutes of Health, Grant AG-370 from the National Science Foundation, and by an appropriation from the Commonwealth of Pennsylvania.
(2) R. Hall, "The Modified Nucleosides on Nucleic Acids", Columbia University Press, New York and London, 1971, pp 317–345.
(3) C. E. Carter and L. H. Cohen, *J. Am. Chem. Soc.*, **77**, 499 (1955).
(4) A. Ballio, S. Barcellona, and V. DiVittorio, *Arch. Biochem. Biophys.*, **101**, 311 (1963).
(5) D. N. Ward, J. Wade, E. F. Walborg, Jr., and T. S. Osdene, *J. Org. Chem.*, **26**, 5000 (1961).
(6) A. Ballio and V. DiVittorio, *Gazz. Chim. Ital.*, **90**, 501 (1960).
(7) J. S. Fleming and J. E. MacNintch, *Annu. Rev. Med. Chem.*, **9**, 75 (1975).
(8) H. M. Berman, W. C. Hamilton, and R. J. Rousseau, *Biochemistry*, **12**, 1809 (1973).
(9) H. M. Berman, D. E. Zacharias, H. L. Carrell, and A. J. Varghese, *Biochemistry*, **15**, 463 (1976).
(10) P. Narayanan and H. M. Berman, *Acta Crystallogr.*, submitted for publication.
(11) Glycyl-L-alanine was synthesized by a modified method of J. P. Greenstein and M. Winitz, "Chemistry of the Amino Acids", Vol. II, Wiley, New York, N.Y., 1961, p 822.
(12) G. Germain, P. Main, and M. M. Woolfson, *Acta Crystallogr., Sect. A*, **27**, 368 (1971).
(13) "International Tables of X-ray Crystallography", Vol. III, Kynoch Press, Birmingham, England, 1962, p 202.
(14) R. F. Stewart, E. R. Davidson, and W. T. Simpson, *J. Chem. Phys.*, **42**, 3175 (1965).
(15) M. Sundaralingam, *Jeru. Symp. Quantum Chem. Biochem.*, **5**, 417 (1973).
(16) D. Voet and A. Rich, *Prog. Nucl. Acid Res. Mol. Biol.*, **16**, 183 (1970).
(17) J. Donohue and K. N. Trueblood, *J. Mol. Biol.*, **2**, 363 (1960).
(18) P. S. Naganathan and K. Venkatesan, *Acta Crystallogr., Sect. B*, **28**, 552 (1972).
(19) R. Parthasarathy, J. Ohrt, and G. B. Chheda, *Biochem. Biophys. Res. Commun.*, **57**, 649 (1974).
(20) N. C. Seeman, J. M. Rosenberg, and A. Rich, *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 804 (1976).
(21) C. E. Bugg and U. Thewalt, *Biochem. Biophys. Res. Commun.*, **46**, 779 (1972).
(22) R. K. McMullan and M. Sundaralingam, *J. Am. Chem. Soc.*, **93**, 7050 (1971).
(23) C. Tamura, T. Hata, S. Sato, and W. Sakarai, *Bull. Chem. Soc. Jpn.*, **45**, 3254 (1972).
(24) IUPAC-IUB Commission on Biochemical Nomenclature, *J. Mol. Biol.*, **52**, 1 (1970).