

# Analysis of *N*-Glycosyl Bond Length in Crystal Structures of Nucleosides and Nucleotides

ALBERT LO \*, ELI SHEFTER \*\*, and TODD G. COCHRAN †

**Abstract** □ A relationship was found between the *N*-glycosyl bond length in  $\beta$ -nucleosides and nucleotides and the spatial relationship of this bond with respect to the lone-pair orbitals on the furanose ring oxygen. This relationship may be relevant to the geometry involved in the hydrolysis of the glycosidic bond.

**Keyphrases** □ Nucleosides and nucleotides—analysis of *N*-glycosyl bond length in crystal structures □ Glycosyl bonds—analysis of bond length in crystal structures of nucleosides and nucleotides □ Hydrolysis—glycosyl bonds, analysis of bond length in crystal structures of nucleosides and nucleotides

The *N*-glycosyl bonds of nucleic acids are readily hydrolyzed by either enzymes (specific nucleosidases) or acid-catalyzed reactions. Although the mechanism of enzymatic cleavage of these glycosidic bonds has been poorly characterized, numerous kinetic studies on the acid-catalyzed hydrolysis of these bonds in nucleosides have produced some definitive ideas on the hydrolytic mechanism (1–3).

Two schemes have been proposed for the acid-catalyzed route. One involves the opening of the protonated furanose ring at the  $C_1'-O_1'$  bond, with subsequent water attack and the formation of the base and sugar products. In the other scheme, the glycosidic bond of the base-protonated species (mono- or dication) is directly cleaved. Evidence for the latter scheme appears to be strongest for most nucleosides (3). Although mechanisms have been developed, there is a dearth of any detailed structural information concerning the geometry of the minimum energy pathway followed by the reaction, *i.e.*, reaction coordinates.

Recently, crystal structure data were used to analyze the geometry of some nucleophilic addition reactions (4, 5). The general approach was to examine the atomic coordinates of a large number of molecules possessing similar chemical subunits for structural variations that might correspond to existing knowledge about these subunits in a reaction scheme. The geometrical parameters observed for any molecule in a crystal are a complicated function of inter- and in-

tramolecular forces. Therefore, one would expect to find differing geometries for a particular subunit in a series of molecular structures.

The hypothesis behind the use of these data is that the interplay of forces that displaces the subunit of interest from its geometrical minimum energy configuration in the crystal can provide a clue as to the path that may occur in the chemical breakdown of the entity. This hypothesis assumes that the initial stages of the chemical breakdown and the geometrical distortions in the crystal structures both follow similar minimum energy routes.

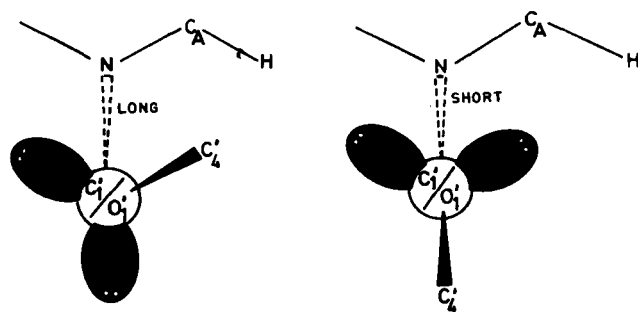
It was deemed worthwhile to utilize this approach to analyze the *N*-glycosyl linkage of nucleotides and nucleosides. For this purpose, published crystal structures whose accuracy was sufficient for such a correlation were selected. The structures examined are listed in Table I together with some pertinent parameters describing the glycosidic bond.

## DISCUSSION

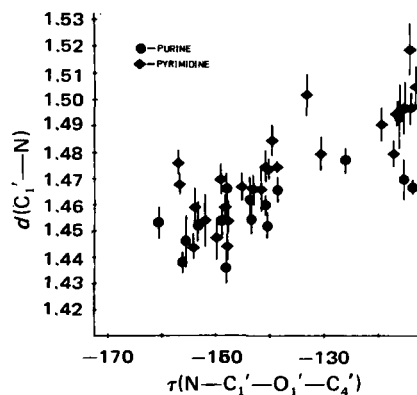
A scrutiny of the parameters in the table shows that there is a significant variation in the  $C_1'-N$  bond length among the tabulated structures. Since the range of C—N bond lengths is reflective of distortions that could occur at the initial stages of the hydrolytic sequence, one might glean some clues from them as to the beginning of the hydrolytic reaction.

The change in the glycosidic bond length is not accompanied by any significant alteration in bonding parameters about  $C_1'$ , *i.e.*, bond lengths and angles. However, some subtle trends in these parameters are worth noting even though they are not *highly* significant. As the glycosidic bond lengthens, the  $N-C_1'-C_2'$  angle decreases slightly and the furanose ring angle ( $O_1'-C_1'-C_2'$ ) tends to enlarge.

The torsion angle about the  $C_1'-O_1'$  bond [ $\tau(N-C_1'-O_1'-C_4')$ ] provides an indication of the spatial relationship of the lone pair of orbitals on  $O_1'$  with regard to the glycosidic bond (Fig. 1). Geometric constraints imposed by the furanose ring limit the values obtained for the  $C_1'-O_1'$  torsion angle to those between synclinal ( $60^\circ$ ) and antiperiplanar ( $180^\circ$ ). The manner by which



**Figure 1**—Schematic drawing of the  $N-C_1'-O_1'-O_4'$  torsion angle. Key: left, synclinal; and right, antiperiplanar.



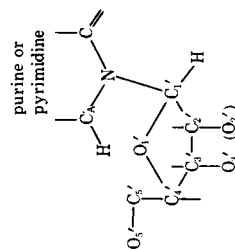
**Figure 2**—Relationship between  $\tau(N-C_1'-O_1'-C_4')$  and the glycosidic bond length.

Table I—Intramolecular Parameters about C<sub>1</sub>' in a Number of Nucleosides and Nucleotides

Compound <sup>a</sup>	Bond Length, Å				Bond Angles, degrees				Torsion Angles, degrees				$\sigma_d^b$ , Å	$\sigma_a^b$ , degrees
	C <sub>1</sub> '—N	C <sub>1</sub> '—O <sub>1</sub> '	C <sub>1</sub> '—C <sub>2</sub> '	C <sub>1</sub> '—H	N—C <sub>1</sub> '—O <sub>1</sub> '	N—C <sub>1</sub> '—C <sub>2</sub> '	N—C <sub>1</sub> '—H	O <sub>1</sub> '—C <sub>1</sub> '—H	O <sub>1</sub> '—C <sub>1</sub> '—C <sub>2</sub> '	O <sub>1</sub> '—C <sub>1</sub> '—H	C <sub>2</sub> '—C <sub>1</sub> '—H	C <sub>2</sub> '—C <sub>1</sub> '—O <sub>1</sub> '		
I Py	1.436	1.409	1.528	1.07	110.1	114.0	103	116	105.7	108	-135.0	-148.2	0.006	0.4
II Pu	1.438	1.422	1.520	1.02	109.8	114.7	101	116	104.1	111	73.0	-156.2	0.004	0.2
III Py	1.444	1.420	1.510	0.95	109.2	116.1	111	104	103.9	112	66.4	-154.2	0.005	0.5
IV Pu	1.445	1.415	1.516	1.04	109.2	115.3	111	99	105.4	115	57.8	-148.0		
V Py	1.446	1.412	1.499	0.97	108.4	117.3	100	114	103.5	114	54.6	-155.6	0.01	0.7
VI Pu	1.448	1.431	1.560	1.01	108.2	116.0	108	107	108.4	109	58.4	-149.9	0.008	0.7
VII Py	1.452	1.412	1.522	0.98	108.3	114.6	107	110	103.7	114	46.6	-160.6		
VIII Pu	1.454	1.413	1.528	0.99	108.4	113.0	107	110	105.8	113	122.5	-143.5		
IX Py	1.452	1.426	1.517	0.89	106.8	114.2	109	110	107.0	109	52.8	-152.0	0.01	1.0
X Pu	1.454	1.413	1.521	0.98	109.8	114.1	102	113	105.3	113	67.1	-148.0	0.008	0.4
XI Py	1.459	1.420	1.507	0.98	108.4	115.6	102	116	103.5	112	65.5	-153.9		
XII Pu	1.452	1.410	1.516	1.01	107.3	114.4	104	113	105.1	113	120.1	-153.4	0.009	0.5
XIII Py	1.460	1.416	1.525	0.99	108.9	113.4	109	107	105.1	114	49.1	-143.7		
XIV Pu	1.467	1.404	1.535	0.95	108.6	112.3	111	110	105.6	109	45.1	-148.4	0.005	0.3
XV Py	1.466	1.413	1.524	0.96	109.0	113.9	107	112	105.8	110	52.3	-145.2	0.006	0.3
XVI Pu	1.467	1.416	1.525	1.05	109.0	113.4	106	114	105.9	109	49.7	-143.1	0.006	0.3
XVII Py	1.466	1.413	1.531	0.97	108.6	114.4	106	113	103.9	112	42.1	-156.7	0.004	0.2
XVIII Pu	1.466	1.413	1.535	0.99	108.8	119.2	98	110	107.6	113	-96.6	-141.7	0.009	0.8
XIX Py	1.468	1.410	1.510	1.20	108.4	113.6	98	118	106.1	113	-143.7	-148.1	0.006	0.4
XX Pu	1.454	1.416	1.529	1.06	108.6	113.9	109	112	106.3	107	103.9	-138.5	0.005	0.3
XXI Py	1.460	1.415	1.545	0.98	107.9	115.8	104	111	105.9	110	-132.9	-149.1	0.006	0.3
XXII Pu	1.467	1.411	1.528	0.98	109.3	113.0	105	114	105.6	112	43.1	-140.6	0.005	0.3
XXIII Py	1.470	1.390	1.528	1.19	108.1	111.6	108	110	107.3	111	9.9	-113.7	0.003	0.2
XXIV Pu	1.474	1.403	1.536	1.03	106.9	114.3	109	110	105.9	107	28.9	-115.2	0.008	0.4
XXV Py	1.475	1.410	1.525	0.95	107.8	113.6	106	112	106.6	111	34.0	-149.0	0.006	0.4
XXVI Pu	1.475	1.418	1.513	0.91	107.7	112.9	100	119	105.9	112	51.4	-140.2	0.004	0.3
XXVII Py	1.475	1.407	1.540	0.90	108.5	112.1	106	109	106.5	114	41.8	-140.8	0.006	0.5
XXVIII Pu	1.477	1.409	1.526	1.01	107.2	113.5	106	112	108.6	109	26.7	-138.7	0.002	0.2
XXIX Py	1.476	1.420	1.512	0.95	108.2	114.5	104	117	104.4	109	3.7	-126.0	0.005	0.3
XXX Pu	1.480	1.434	1.515	0.96	109.2	114.5	106	113	106.5	109	41.8	-156.9	0.005	0.3
XXXI Py	1.486	1.418	1.525	1.03	108.0	113.1	104	107	106.7	118	39.1	-130.6	0.006	0.4
XXXII Pu	1.492	1.423	1.539	1.06	108.3	113.9	108	108	107.1	111	54.1	-139.7	0.006	0.5
XXXIII Py	1.494	1.444	1.508	1.12	107.3	112.6	100	126	106.6	104	19.6	-119.4	0.006	0.4
XXXIV Pu	1.494	1.405	1.536	1.00	108.7	112.2	109	107	110.2	102	25.7	-116.1	0.012	0.7
XXXV Py	1.496	1.410	1.543	1.00	108.4	111.4	108	104	107.7	116	20.3	-116.6	0.005	0.3
XXXVI Pu	1.497	1.412	1.527	1.04	108.7	111.4	110	112	106.9	108	18.1	-120.0	0.004	0.2
XXXVII Py	1.502	1.431	1.538	1.20	106.4	114.3	102	113	108.3	112	18.3	-114.1	0.006	0.4
XXXVIII Pu	1.505	1.387	1.528	0.97	108.6	112.4	100	106	107.7	121	57.4	-133.2	0.008	0.4
XXXIX Py	1.508	1.404	1.508	0.99	108.2	112.3	106	109	107.2	114	27.8	-113.27	0.003	0.2
XXXI Pu	1.510	1.404	1.528	0.96	107.7	113.6	105	102	108.1	120	-0.09	-111.55	0.003	0.2
XXXVI Py	1.519	1.397	1.534	1.05	108.3	112.2	110	102	108.6	115	-6.2	-114.36	0.003	0.2
Average	1.471	1.414	1.527	1.01	108.6	113.9	106	111	106.3	112	16.0	-137.2	0.001	0.6
(SD)	(0.02)	(-)	(0.01)	(0.07)	(1.6)	(1.6)	(5)	(5)	(1.4)	(4)	(92.9)	(15.6)		

<sup>a</sup> Name and reference are listed below. Pu and Py denote purine and pyrimidine analog, respectively. If more than one crystallographically unique molecule appears in a crystal, they are listed with the same identifying number.

- I 6-Thiopurine riboside, E. Shefter, *J. Pharm. Sci.*, **57**, 1157(1968).
- II Tubercidin, J. Abola and M. Sundaralingam, *Acta Crystallogr.*, **B29**, 697(1973).
- III Dihydrouridine hemihydrate, D. Suck, W. Saenger, and K. Zeckmeister, *ibid.*, **B28**, 596(1972).
- IV  $\beta$ -Adenosine-2'- $\beta$ -uridine-5'-phosphoric acid tetrahydrate, E. Shefter, M. Barlow, R. A. Sparks, and K. N. Trueblood, *ibid.*, **B25**, 895(1969).
- V Triethylammonium cyclic uridine-3',5'-phosphate, C. L. Coulter, *ibid.*, **B25**, 2055(1969).
- VI Guanosine dihydrate, C. E. Bugg and R. E. Marsh, *ibid.*, **B26**, 1089(1970).
- VII Dihydrothymine, J. Konner and I. L. Karle, *ibid.*, **B26**, 770(1970).
- VIII Dihydrouridine, M. Sundaralingam, S. T. Rao, and J. Abola, *J. Amer. Chem. Soc.*, **93**, 1089(1970).
- IX Inosine dihydrate, U. Thewalt, C. E. Bugg, and R. E. Marsh, *Acta Crystallogr.*, **B26**, 1089(1970).
- X Disodium uridine 3'-phosphate, M. Viswamitra, B. Reddy, M. N. James, and G. Williams, *ibid.*, **B28**, 1108(1972).
- XI 3-Deazuridine, C. W. Schalte and W. Saenger, *ibid.*, **B29**, 61(1973).
- XII 2-O-Tetrahydropyranlyuridine, P. H. Stothart, I. D. Brown, and T. Neilson, *ibid.*, **B29**, 2237(1973).
- XIII 5-Hydroxyuridine, U. Thewalt and C. E. Bugg, *ibid.*, **B29**, 1393(1973).
- XIV 4-Thiouridine, W. Saenger and K. H. Scheit, *J. Mol. Biol.*, **50**, 153(1970).
- XV N-2-Dimethylguanosine, T. Brennan, C. Weeks, E. Shefter, S. T. Rao, and M. Sundaralingam, *J. Amer. Chem. Soc.*, **94**, 8548(1972).
- XVI 3'-O-Acetyladenosine, S. T. Rao and M. Sundaralingam, *ibid.*, **92**, 4963(1970).
- XVII 5'-Methylene adenosine 3',5'-cyclic monophosphate monohydrate, M. Sundaralingam and J. Abola, *ibid.*, **94**, 5070(1972).
- XVIII Adenosine hydrochloride, K. Shikata, T. Ueki, and T. Mitsui, *Acta Crystallogr.*, **B29**, 31(1973).
- XIX Adenosine, T. F. Lai and R. E. Marsh, *ibid.*, **B28**, 1982(1972).
- XX 1:1 Ethanol and 3'-deoxy-3'-(dihydroxyphosphinylmethyl)adenosine, S. M. Hecht and M. Sundaralingam, *J. Amer. Chem. Soc.*, **94**, 4314(1972).
- XXI Uracil- $\beta$ -D-arabinofuranoside, P. Tollin, H. R. Wilson, and D. W. Young, *Acta Crystallogr.*, **B29**, 1641(1973).
- XXII 5-Chlorouridine, S. W. Hawkinson and C. L. Coulter, *ibid.*, **B27**, 34(1971).
- XXIII Cytidylic acid b, M. Sundaralingam and L. H. Jensen, *J. Mol. Biol.*, **13**, 914(1965).
- XXIV 1- $\beta$ -D-Arabinofuranosyl cytosine hydrochloride (absolute configuration), J. S. Shetfinski and R. E. Marsh, *Acta Crystallogr.*, **B29**, 192(1972).
- XXV Adenosine-3'-phosphate dihydrate, M. Sundaralingam, *ibid.*, **21**, 495(1966).
- XXVI 5-Chloro-2'-deoxyuridine, D. W. Young and E. M. Morris, *ibid.*, **B29**, 1259(1973).
- XXVII Thymidine, D. W. Young, P. Tollin, and H. R. Wilson, *ibid.*, **B25**, 1423(1969).
- XXVIII 3'-O-Acetyl-4-thiothymidine, W. Saenger and D. Suck, *ibid.*, **B27**, 2105(1971).
- XXIX 2,4-Dithiouridine hydrate, G. Lin and M. Sundaralingam, *ibid.*, **B27**, 961(1971).
- XXX Adenosine-5'-phosphate monohydrate, J. Kraut and L. H. Jensen, *ibid.*, **16**, 79(1963).
- XXXI 2-Thiocytidine dihydrate, G. H.-Y. Lin, M. Sundaralingam, and S. K. Akora, *J. Amer. Chem. Soc.*, **93**, 1235(1971).
- XXXII 2,4-Dithiouridine hydrate, W. Saenger and D. Suck, *Acta Crystallogr.*, **B27**, 1178(1971).
- XXXIII Cytidine, S. Furberg, C. Petersen, and C. Romming, *ibid.*, **18**, 313(1965).
- XXXIV 5-[1-(2'-Deoxy- $\alpha$ -D-ribofuranosyl)uracilyl] disulfide, E. Shefter, M. D. Kotick, and T. J. Bardos, *J. Pharm. Sci.*, **56**, 1293(1967).
- XXXV 2-Deoxycytidine hydrochloride, E. Subramanian and D. J. Hunt, *Acta Crystallogr.*, **B26**, 303(1970).
- XXXVI Deoxycytidine-5'-phosphate monohydrate, M. S. Viswamitra, B. Reddy, G. H.-Y. Lin, and M. Sundaralingam, *J. Amer. Chem. Soc.*, **93**, 4565(1971).



$b$   $\sigma_d$  and  $\sigma_a$  are the average standard deviations reported for bond lengths and angles, respectively, of parameters not involving hydrogen. Those for hydrogen are much greater.

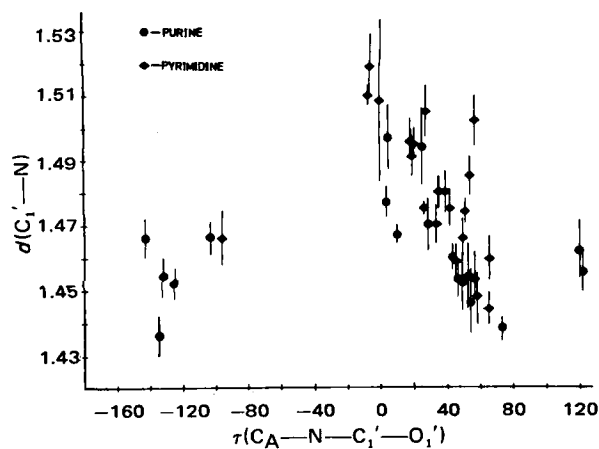


Figure 3—Correlation of  $\tau(C_A-N-C_1'-O_1')$  and glycosidic bond length.

the furanose ring puckers influences the value of this torsion angle.

A relationship between  $\tau(N-C_1'-O_1'-C_4')$  and the glycosidic bond length appears to exist (Fig. 2). The *N*-glycosyl linkage is shortest when this constellation of atoms is antiperiplanar and longest when synclinal. The orientation of a lone-pair orbital on the oxygen becomes antiperiplanar as this angle tends toward a synclinal value. Thus, the *N*-glycosyl bond length appears to be longest when it is antiperiplanar to a lone-pair orbital on the oxygen.

The following regression equation ( $r = 0.80$ ) was derived for the relationship shown in Fig. 2:

$$d(C_1'-N) = 1.625 - 0.00112\tau(N-C_1'-O_1'-C_4') \quad (\text{Eq. 1})$$

This equation predicts the shortest value possible for the glycosidic bond (occurring at  $\tau = 180^\circ$ ) to be 1.43 Å. This length is much shorter than an  $N(sp^2)-C(sp^2)$  single bond (1.485 Å) (6) or the  $N-C$  bonds of alkylated nucleic acid bases (7).

In a structural analysis of tetrahedral dioxo compounds and orthoesters, a similar type of relationship was found (5). The  $C-O'$  bond in these compounds was a function of the  $O'-C-O''-H$  torsion angle. The  $C-O'$  bond length was longest when a lone-pair orbital on  $O''$  was antiperiplanar to it. This observation is supported by studies carried out on the decomposition of conformationally rigid hemi-orthoesters (8) and hemi-amide acetals (9), where the bond most readily cleaved appeared to be the one with two lone-pair orbitals antiperiplanar to it.

Jeffrey *et al.* (10), using nonempirical molecular orbital calculations, found a similar relationship between bond length and conformation for methylenediol. The difference in bond length for the synclinal and antiperiplanar arrangements is also reflected in the stretching force constants (11). The calculated values are 5.8 and 6.0 mdynes/Å for the longer and shorter bonds, respectively. Therefore, the reaction that cleaves the longer bond (antiperiplanar conformation with respect to a lone-pair orbital) would be energetically favored, at least in the initial phase.

Other investigators (12, 13) indicated that the observed differences in glycosidic bonds could be attributed to nonbonded interactions between the base and the sugar. The conformation of the base relative to the sugar can be described by the torsion angle,  $\tau(C_A-N-C_1'-O_1')$ . This angle was correlated by these investigators to the glycosidic bond length. With the tabulated data, the correlation appears to break down (Fig. 3). However, this correlation is valid if no differentiation is made between the sides of the purine or pyrimidine rings, *i.e.*, between the *syn*- and *anti*-confor-

mations (14). If  $180^\circ$  is added to the  $C_A-N-C_1'-O_1'$  torsion angle of the six nucleosides with a *syn*-conformation that show a marked deviation from the linear trend in Fig. 3, a linear correlation is observed ( $r = 0.65$ ).

Since the torsion angles about the  $C_1'-O_1'$  and  $C_1'-N$  bonds can both be related to the glycosidic bond length, the conformations about these bonds appear to be dependent upon one another. Regression analysis of the tabulated values for these two torsion angles is suggestive of this, showing a linear correlation ( $r = 0.67$ ) for the 43 nucleosides and nucleotides in Table I. This result is not unexpected in light of work (15) showing that the freedom of rotation about the glycosidic bond in nucleosides and nucleotides is closely related to the puckering in the furanose ring.

In terms of reaction coordinates, the tabulated data show that a conformation in which the  $N-C_1'-O_1'-C_4'$  torsion angle tends toward a synclinal value would be favorable for a direct cleavage mechanism of hydrolysis, since this conformation is associated with an increasing glycosidic bond length and an antiperiplanar orientation between the glycosidic bond and a lone-pair orbital on the oxygen. While this cannot be construed as direct evidence that such a geometry occurs during hydrolysis, it is easily envisioned that a catalytic species (nucleosidase) could induce such a conformational change in a nucleoside or nucleotide, thus facilitating the hydrolysis.

## REFERENCES

- (1) B. Capon, *Chem. Rev.*, **69**, 407(1969).
- (2) N. K. Kochetkov and E. I. Budovskii, "Organic Chemistry of Nucleic Acids," part B (English translation), Plenum, New York, N.Y., 1972, p. 425.
- (3) R. Shapiro and M. Danzig, *Biochemistry*, **11**, 23(1972).
- (4) H. B. Bürgi, J. D. Dunitz, and E. Shefter, *J. Amer. Chem. Soc.*, **95**, 5065(1973).
- (5) H. B. Bürgi, J. D. Dunitz, and E. Shefter, *Acta Crystallogr.*, **B30**, 1517(1974).
- (6) A. Camerman, *Can. J. Chem.*, **48**, 179(1970).
- (7) D. Voet and A. Rich, *Progr. Nucleic Acid Res. Mol. Biol.*, **10**, 183(1970).
- (8) P. Deslongchamps, P. Atlani, D. Fréhel, and A. Malaval, *Can. J. Chem.*, **50**, 3405(1972).
- (9) P. Deslongchamps, C. Lebreux, and R. Taillefer, *ibid.*, **51**, 1665(1973).
- (10) G. A. Jeffrey, J. A. Pople, and L. Radom, *Carbohydr. Res.*, **25**, 117(1972).
- (11) J. M. Lehn, G. Wipff, and H. B. Bürgi, *Helv. Chim. Acta*, **57**, 493(1974).
- (12) G. H. Lin, M. Sundaralingam, and S. K. Arora, *J. Amer. Chem. Soc.*, **93**, 1235(1971).
- (13) J. Abola and M. Sundaralingam, *Acta Crystallogr.*, **B29**, 697(1973).
- (14) J. Donohue and K. N. Trueblood, *J. Mol. Biol.*, **2**, 363(1960).
- (15) A. E. V. Haschemeyer and A. Rich, *ibid.*, **27**, 369(1967).

## ACKNOWLEDGMENTS AND ADDRESSES

Received May 23, 1974, from the \*Department of Pharmaceutics, School of Pharmacy, State University of New York at Buffalo, Buffalo, NY 14214, and the †School of Pharmacy, Duquesne University, Pittsburgh, PA 15219

Accepted for publication February 21, 1975.

The authors are indebted to Mrs. Phyllis Sackman for her assistance, Dr. H. B. Bürgi and Dr. J. D. Dunitz for their interest, and the National Cancer Institute for support (Grant CA-10104).

\* To whom inquiries should be directed.