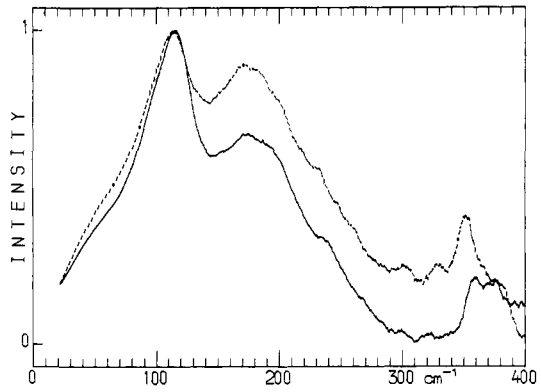


Figure 8. Full line: Smoothed $R(\nu)$ curve of a ca. 0.57 M aqueous solution of K_2-5′-GMP. Broken line: smoothed $R(\bar{\nu})$ curve of a ca. 0.55 M solution of deuterated K_2-5′-GMP in D_2O. Both solutions were investigated at pH ca. 8.

occurs. This does not imply that the binding of the sodium ions in the gel state is identical with the binding of the cations in the solution, because there is a difference in protonization of the phosphor group,[7] and this may play an important role for the self-association.

In comparison to the corresponding band in the slightly basic solutions of the potassium salt, the band at 112 cm^{-1} is sharper in the gels. The half-width is ca. 20 cm^{-1} in the gels and ca. 30 cm^{-1} in solutions at room temperature. This reflects most probably the fact that the number of purine rings in a stack is greater for the gel, and it seems reasonable to assume that this stacking of self-associated purine rings might result in the high viscosity (gel state) at this pH value.

Isotope Effects. In order to achieve a better knowledge of the physical origin of the modes giving rise to the low-frequency bands observed here, we compared the $R(\bar{\nu})$ curves of 5′-GMP dissolved in H_2O with the similar solutions of deuterated 5′-GMP in D_2O.

However, as shown in Figure 8 only small isotope effects are observed when comparing the $R(\bar{\nu})$ curves of K_2-5′-GMP in H_2O (0.57 M) with K_2-5′-GMP in D_2O (0.55 M). An identical conclusion could be drawn when comparing the similar solutions of Na_2-5′-GMP in H_2O and D_2O (not shown).

Although these observations might seem to contradict the interpretation that the modes involve hydrogen bonding, a similar isotopic independence have been reported for other hydrogen-bonded systems of which dimeric formic acid is a typical example.[21]

Thus a thorough discussion of the modes giving rise to these low-frequency bands shall wait until more experimental data has been gathered. In this context a study of other isotopic substituted species of 5′-GMP is of special interest.

Conclusion

This paper shows that it is possible to use the $R(\bar{\nu})$ representation in studying self-association of 5′-GMP in aqueous solutions as a function of cation (Na^+ or K^+) concentration, temperature, and pH. A band with a maximum at 110–120 cm^{-1} is assigned to self-associated 5′-GMP molecules with coplanar purine rings. Previous NMR results showing that this self-association is enhanced by potassium ions in solutions at pH ca. 8 are corroborated.[11-15]

The self-association of the sodium salt in the gel state at pH ca. 5 is more pronounced than in the corresponding solution of this salt at pH ca. 8. Although solutions of deuterium-substituted molecules (N–D and O–D) were investigated, a description of the modes giving rise to the low-frequency bands could not be achieved.

It is known that local changes in the environments of DNA molecules due to ions or proteins are likely to produce changes in conformation.[22] These conformational changes may give rise to changes in the low-frequency vibrational spectrum, and it is our hope that the $R(\bar{\nu})$ representation of the low-frequency Raman scattering in analogy with results in this paper can be used for more complicated systems of nucleic acids.

Acknowledgment. We wish to express our gratitude to the Danish Natural Science Research Council for partially financing the Raman equipment and to Dr. D. Christensen and Professor E. Praestgaard for helpful discussions.

Registry No. K_2-5′-GMP, 3254-39-5; Na_2-5′-GMP, 5550-12-9.

(21) Carlson, G. L.; Witkowski, R. E.; Fateley, W. G.; *Spectrochim Acta* **1966**, *22*, 1117–1123.
(22) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; van der Marel, G.; van Boom, J. H.; Rich, A. *Science (Washington, D.C.)* **1981**, *211*, 171–176.

l-Canavanine, a Paradigm for the Structures of Substituted Guanidines

Alan Boyar and Richard E. Marsh*[1]

Contribution No. 6485 from the Arthur Amos Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91125. *Received July 16, 1981*

Abstract: A crystal-structure study of the plant amino acid *l*-canavanine, the δ-oxa analogue of arginine, has been carried out. The zwitterionic molecule is protonated on the α-amino nitrogen atom; the neutral guanidine grouping is in the amino form, which seems to be preferred by all guanidine groups with electron-withdrawing substituents. Revised values of the various p*K*'s of canavanine and of canaline are reported, derived by reanalysis of the titration curves reported in 1935.

l-Canavanine, the δ-oxa analogue of *l*-arginine, is the principal nonprotein amino acid found in many species of the Lotoideae, also known as Papilionaceae or Fabaceae, a subfamily of the Leguminosae.[2] It is synthesized in the leaves and in the pod wall (pericarp) by a series of reactions which is postulated to be analogous to the arginine biosynthetic pathway and is stored in the seed where it may account for as much as 13% of the total dry weight. Canavanine functions as a source of nitrogen during germination and as a toxin to rodents and insects which might otherwise destroy the seeds. Canavanine is a substrate for nearly every enzyme for which arginine is also a substrate; it can competitively inhibit enzymes of arginine metabolism and is known

(1) To whom correspondence should be addressed.
(2) Bell, E. A.; Lakey, J. A.; Polhill, R. M. *Biochem. Syst. Ecol.* **1978**, *6*, 201–212.

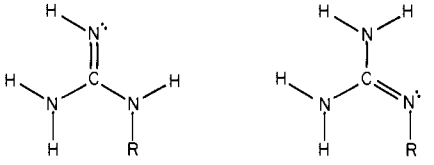
0002-7863/82/1504-1995$01.25/0

Table I. Crystal Data for *l*-Canavanine

$C_5O_3N_4H_{12}$	fw 176.17
monoclinic	$P2_1$
$a = 5.505$ (1) Å	$Z = 2$
$b = 8.419$ (3) Å	$F(000) = 188$ e
$c = 8.432$ (2) Å	$D_{measd} = 1.494$ (7) g cm^{-3}
$\beta = 93.06$ (1)°	$D_{calcd} = 1.499$ (1) g cm^{-3}
$V = 390.2$ (3) Å^3	λ(Cu Kα) = 1.5418 Å

to be a noncompetitive inhibitor for many other enzymes. Canavanine is a useful laboratory agent because it exhibits potent antimetabolic effects in many organisms and can be incorporated into polypeptides in place of arginine, thus modifying the structural and functional properties of proteins; it has been used in studies of mechanisms for degrading defective proteins, of amino acid toxicities, and of metabolic control in higher plants.[3]

Structural interest in *l*-canavanine focuses on its close relationship to *l*-arginine and, in particular, on the tautomeric form that the uncharged oxyguanidinium group assumes. A neutral monosubstituted guanidine group can exist in two forms, which have been designated as imino ((aminoiminomethyl)amino, I) and amino (*N*-(diaminomethyl)imino, II). Differentiation between



imino I amino II

these two tautomers may not be easily accomplished because of the stringent conditions often necessary to achieve an uncharged group; for example, the guanidine group in *l*-arginine is protonated at all pHs below about 12.5.[4] Canavanine, on the other hand, loses this proton at a pH of about 7; as a result, the zwitterionic form of canavanine in aqueous solution or in the solid state shows the extra proton on the α-amino group (pK = 9.2) and the guanidine group remains uncharged.

In the absence of definitive information, it appears that neutral guanidine substituents are usually assumed to exist in the imino form (I); canavanine, for example, is listed in Chemical Abstracts as the *O*-iminoguanidine derivative of homoserine, *O*-((aminoiminomethyl)amino)-*l*-homoserine. However, structural studies have shown that nitroguanidine,[5] sulfaguanidine,[6] and cyanoguanidine[7] all exist in the amino form.

We have now carried out an X-ray diffraction study on single crystals of *l*-canavanine and find that the molecules exist in the amino form (II). We also report a low-resolution ^{15}N NMR spectrum of *l*-canavanine which shows the chemical shifts for the three nonequivalent nitrogen atoms. During the course of this work we became interested in the pK values for canavanine and grew dissatisfied with the values reported nearly 50 years ago.[8] Accordingly, we have recalculated the pK values of both *l*-canavanine and *l*-canaline (*O*-amino-*l*-homoserine; the guanidine group of canavanine is replaced by $-NH_2$), using the published titration curves, and report revised values.

Experimental Section

A sample of *l*-canavanine, obtained from CalBiochem Inc., was crystallized by evaporation from a water–ethanol solution. The small, transparent crystals grew in clusters. Oscillation and Weissenberg photographs showed Laue symmetry $2/m$ and systematic absence of reflections $0k0$ with k odd, characteristic of the monoclinic space group $P2_1$. Cell dimensions were obtained from 2θ values for 20 reflections centered on a General Electric two-circle diffractometer; the density was determined from neutral buoyancy in an ethanol–chloroform solution. Crystal data are listed in Table I.

Table II. Atomic Coordinates for *l*-Canavanine (×10^4 for C, N, and O; ×10^3 for H)

atom	x	y	z
C(1)	−837 (5)	−1176 (4)	220 (3)
C(2)	1888 (5)	−907 (4)	550 (3)
C(3)	2248 (6)	309 (5)	1888 (4)
C(4)	4852 (6)	795 (5)	2252 (4)
C(5)	8185 (5)	2071 (5)	5042 (3)
N(1)	2995 (5)	−340 (4)	−934 (3)
N(2)	7370 (5)	2572 (4)	3653 (2)
N(3)	6865 (6)	1281 (5)	6086 (4)
N(4)	10568 (5)	2322 (5)	5481 (4)
O(1)	−1866 (4)	−421 (4)	−894 (3)
O(2)	−1861 (4)	−2089 (*a*)	1159 (3)
O(3)	4863 (4)	2056 (4)	3373 (3)

atom	bonded to	x	y	z
H(1)	N(1)	234 (6)	−99 (5)	−194 (5)
H(2)		246 (7)	75 (6)	−118 (5)
H(3)		454 (8)	−37 (6)	−90 (5)
H(4)	C(2)	270 (6)	−192 (5)	86 (4)
H(5)	C(3)	171 (6)	−11 (5)	288 (4)
H(6)		124 (6)	121 (5)	153 (4)
H(7)	C(4)	583 (7)	−1 (6)	272 (5)
H(8)		544 (8)	119 (6)	125 (6)
H(9)	N(3)	771 (8)	68 (7)	705 (6)
H(10)		551 (8)	96 (6)	572 (5)
H(11)	N(4)	1100 (6)	249 (5)	655 (5)
H(12)		1143 (6)	279 (6)	480 (5)

[a] Held fixed to define the origin.

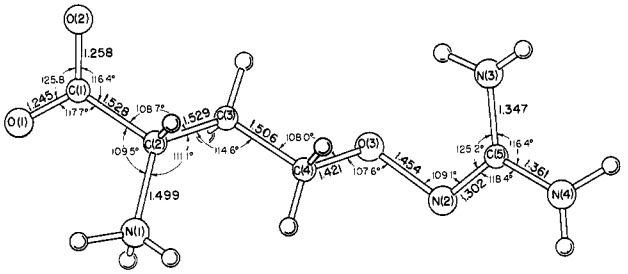


Figure 1. Bond distances and angles in *l*-canavanine. Esd's are 0.004–0.005 Å and about 0.2°.

An initial set of intensities was collected, and direct phasing methods quickly led to a solution of the structure. Least-squares refinement was unsatisfactory, however, in that a relatively large number of reflections showed persistently poor agreement between F_{obsd} and F_c, perhaps due to misalignment of the crystal. Since our original supply of crystals was exhausted, we repeated the recrystallization process; unfortunately, the new crystals were invariably twinned, even after several attempts. The twinning was macroscopic, across the *ab* plane, and the twinning index was 6, such that the $6kl$ reflections of one twin coincided with the $6,\bar{k},\bar{l} - 1$ reflections of the other. For the crystal we eventually used for intensity measurements, the volume ratio of the twins was 0.72:1.

Intensities were collected on a Datex-automated General Electric diffractometer using Ni-filtered Cu Kα radiation, 2θ scans at a rate of 2°/min, 30-s background counts at each extreme, and scan widths varying linearly from 1.95° at $2\theta = 10°$ to 2.65° at $2\theta = 150°$. Two check reflections showed no indication of intensity loss during the 42 h of X-ray exposure.

Data were collected for 883 independent reflections, of which only four had net intensities less than zero. For the reflections with $h = 6$ we at first attempted to assign intensities by prorating the measured values for the twin pairs $6kl$ and $6,\bar{k},\bar{l} - 1$, but the results were unsatisfactory so we assigned weights of zero to these 52 reflections. Reflections with $h = 0$ were treated successfully by multiplying their measured intensities by 0.58, the relative size of the larger twin. Observational variances $\sigma^2(I)$ were based on counting statistics plus an additional term (0.02 × scan count)^2.

(3) For reviews of canavanine biology, see: Rosenthal, G. A. *Q. Rev. Biol.* **1977**, *52*, 155–178. Fowden, L.; Lea, P. J.; Bell, E. A. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1979**, *50*, 117–175.

(4) Schmidt, C. L. A.; Kirk, P. L.; Appleman, W. K. *J. Biol. Chem.* **1930**, *88*, 285–293.

(5) Bryden, J. H.; Burkhardt, L. A.; Hughes, E. W.; Donohue, J. *Acta Crystallogr.* **1956**, *9*, 573–578.

(6) Alleaume, M.; Gulko, A.; Herbstein, F. H.; Kapon, M.; Marsh, R. E.; *Acta Crystallogr., Sect. B* **1976**, *B32*, 669–682.

(7) Hughes, E. W. *J. Am. Chem. Soc.* **1940**, *62*, 1258–1267. See also Hirshfeld, F. L.; Hope, H. *Acta Crystallogr., Sect. B* **1980**, *B36*, 406–415 and references therein.

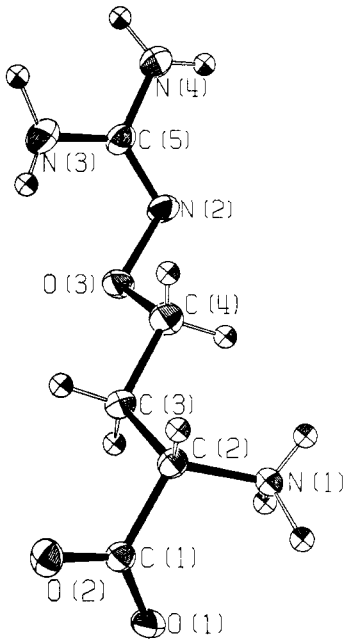


Figure 2. A stereoscopic view[9] of the *l*-canavanine molecule.

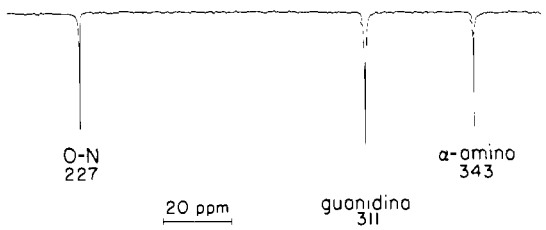


Figure 3. Natural-abundance ^{15}N NMR spectrum of *l*-canavanine sulfate in 90% H_2O and 10% D_2O with proton decoupling. The chemical shifts are in ppm upfield from 1 M NO_3^-.

Refinement, starting with the parameters from the earlier work, was by full-matrix minimization of the quantity $\sum w(F_o^2 - F_c^2)^2$, where $w = 1/\sigma^2(F_o^2)$. In the final cycles, 158 parameters were adjusted: coordinates for 24 atoms, anisotropic *B*'s for the 12 C, N, and O atoms, isotropic *B*'s for the 12 H atoms, a scale factor and an extinction parameter. The goodness-of-fit, $[\sum(w(F_o^2 - F_c^2)^2)/(831 - 158)]^{1/2}$, was 2.2 for 831 reflections of nonzero weight; R ($= \sum|\Delta F|/|F_o|$) was 0.040 for the 827 having positive net intensities.

Final atomic parameters are given in Table II. Bond distances and angles involving the C, N, and O atoms are given in Figure 1; a stereoscopic view of the molecule is shown in Figure 2.

The natural-abundance ^{15}N NMR spectrum (Figure 3) was recorded on a WM 500 spectrometer in FT mode with broad-band decoupling. The magnetic field strength was 11.7 T, the temperature was 296 K, the pulse angle was 45°, the cycle time was 6.5 s and 6865 scans were collected. Acetonitrile was the external standard. The pH, measured before and after data collection, was 1.4 and the concentration of canavanine sulfate was 1 M.

The only p*K* values for *l*-canavanine that we could find in the literature are those reported by Tomiyama in 1935.[8] These values are rather confusing and somewhat at odds with the published titration curve; moreover, the p*I* value calculated from the reported p*K*'s, $^1/_2((14 - 7.4) + 9.25) = 7.92$, does not agree with the value of about 8.15 which can be clearly read directly from the curve. Accordingly, we have recalculated the p*K* values for both canavanine and canaline, using the titration curves published by Tomiyama.[8] Our results are given in Table III, and the computational details are included in the supplementary material.

Results

As predicted by the p*K* values, *l*-canavanine (in the crystalline state) is a zwitterion with the proton transferred from the carboxyl group to the α-amino nitrogen N(1); the guanidine group is uncharged and has the amino form. In contrast, in crystals of *l*-arginine dihydrate the proton is transferred to the guanidinium nitrogen N(2). This difference is clearly reflected in the bond distances for the two compounds: whereas all three C–N distances in the guanidinium group of *l*-arginine are approximately equal, at 1.34 (1) Å,[10] in *l*-canavanine the double bond is more localized in the C(5)–N(2) bond which is 0.05 Å shorter than C(5)–N(3) and C(5)–N(4). The additional proton on N(1) of *l*-canavanine results in only a marginally longer C(2)–N(1) bond, 1.499 (5) Å, than in *l*-arginine, 1.480 (9) Å. Other differences between the two molecules can be related to the replacement of the δ-carbon atom in arginine with the oxygen atom O(3) in canavanine.

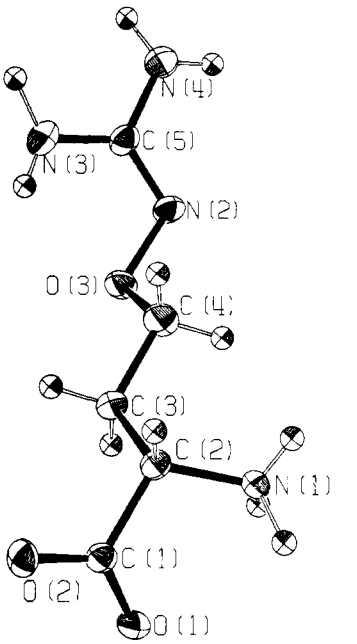


Table III. p*K* Values for *l*-Canavanine and *l*-Canaline, As Recalculated from the Titration Curves Published by Tomiyama[8] [a]

	previously reported[8]		revised values	
group	p*K*	p*I*	p*K*	p*I*
	Canavanine			
COOH	2.50		2.35	
guanidine	6.60	7.92	7.01	8.12
NH_2	9.25		9.22	
	Canaline			
COOH	2.40		1.93	
ONH_2	3.70	6.45	3.96	6.56
NH_2	9.20		9.14	

[a] We estimate standard deviations of about 0.05–0.1 unit.

Table IV. Details of the Hydrogen Bonds, D-H···A

			D···A,[g]	D–H,[h]	H···	dev,[a] Å	
D	H	A	Å	Å	A,[h] Å	H	A
N(1)	H(1)	N(2)[b]	2.887	1.06	1.90		
	H(2)	O(2)[c]	2.812	0.99	1.85		
	H(3)	O(1)[d]	2.828	0.85	1.97		
N(3)	H(9)	O(1)[e]	2.973	1.04	1.97	0.22	0.35
	H(10)			0.84		0.17	
N(4)	H(11)	O(2)[f]	2.926	0.93	2.00	−0.44	−1.44
	H(12)	O(3)[d]	3.041	0.86	2.37	0.03	1.59

[a] Deviations from the least-squares plane of the guanidine group C(5), N(2), N(3), and N(4). [b] $-x + 1, y - 1/2, -z$. [c] $-x, y + 1/2, -z$. [d] $x + 1, y, z$. [e] $x + 1, y, z + 1$. [f] $-x + 1, y + 1/2, -z + 1$. [g] $\sigma \cong 0.004$ Å. [h] $\sigma \cong 0.05$ Å.

(8) Tomiyama, T. *J. Biol. Chem.* **1935**, *111*, 45–49.
(9) Johnson, C. K. "ORTEP", U. S. Atomic Energy Commission Report ORNL-3794; Oak Ridge National Laboratory: Oak Ridge, Tenn., 1965.
(10) Karle, I. L.; Karle, J. *Acta Crystallogr.* **1964**, *17*, 835–841.

In the crystal structure of *l*-canavanine, hydrogen bonds are formed by six of the seven available protons, as detailed in Table IV. Even the ether-like oxygen atom O(3) is pressed into service as an acceptor although, at 3.04 Å, this N···O bond is quite weak.

The guanidine atoms N(2), N(3), N(4), and C(5) are coplanar within experimental error; O(3) lies 0.11 Å from their plane. Out-of-plane deviations of the hydrogen atoms on N(3) and N(4) range up to 0.44 Å, in directions such as to make the N–H···O hydrogen bonds more linear (see Table IV)—an effect that has been noted several times before.[11] Both C–NH_2 groupings are appreciably pyramidal, in opposite senses; in addition, the N(4) amino group is rotated about the C–N bond so that one proton, H(12), lies essentially in the guanidine plane while the other, H(11), is far out of plane.

The carboxylate atoms C(1), C(2), O(1), and O(2) are coplanar within 0.016 Å, which is somewhat larger than experimental error (about 0.005 Å); the amino nitrogen N(1) lies 0.57 Å from this plane, a not unusual value.[12]

Discussion

Our finding that the guanidine group in crystals of *l*-canavanine exists in the amino form (II) continues a parallel with a number of other substituted guanidines, all of which were originally proposed to exist in the imino form (I) before proof to the contrary was found. The story begins with cyanoguanidine (also called dicyandiamide), for which the imino form was favored[13] until a crystal structure study in 1940[7] indicated the amino form. In the case of nitroguanidine, chemical studies in 1951[14] cast first doubt on the imino assumption, and a crystal structure study in 1956 clearly established the amino form.[5] Sulfaguanidine was thought to exist in the imino form at least until 1962, when Schwenker[15] proposed the amino form on the basis of IR and proton NMR spectra; it was later confirmed by crystal structure studies[6] and ^{15}N NMR spectroscopy.[16]

For *l*-arginine the situation is more complicated, since the guanidine group remains protonated to a pH of about 12.5, effectively precluding crystal structure studies; moreover, proton NMR data are not useful at high pH because of rapid proton exchange and peak broadening due to quadrupole effects. In 1974 Suzuki and co-workers[17] reported solution ^{15}N NMR studies which seemed to show that the single resonance for the terminal nitrogen atoms of the guanidinium group observed in the pH range 2.1–11.8 splits into two peaks at pH 13.5, and they interpreted this splitting in terms of two nonequivalent terminal nitrogens and hence the imino form. However, in two later studies[18,19] it was pointed out that the observed splitting was due to the reappearance of the α-amino resonance, which had presumably been quenched at lower pH by trace amounts of paramagnetic ions; it was also noted that rapid proton exchange at high pH would probably result in a single resonance for the two terminal nitrogen atoms, whatever the tautomeric form. Very recently, J. D. Roberts and K. Kanamori (private communication) have estimated, from the chemical shifts for *l*-arginine and related model compounds, that the amino-to-imino equilibrium constant in solution of pH 14.0 is about 2, and hence that the imino form is slightly favored.

If we extend the series of substituted guanidines to guanidine itself (where the substituent is a hydrogen atom), the imino:amino ratio must be exactly 2:1 from symmetry. Replacement of a proton with an alkyl group, as in arginine, apparently has little effect on this equilibrium; however, replacement by an electron-withdrawing group as in canavanine, nitroguanidine, and cyanoguanidine clearly shifts the equilibrium so as to favor the amino form. We suggest, then, that guanidine compounds containing electron-withdrawing substituents be considered a priori to exist in the amino form (II) rather than the imino form (I). In addition, the structures commonly assigned to some disubstituted guanidines in which one substituent is strongly electron withdrawing, such as canavanosuccinic acid and desaminocanavanine, should probably be revised.

A related question concerning the nature of the oxyguanidine grouping in canavanine is the site of protonation in acid solution. On the one hand, stabilization due to a symmetric group would favor protonation at the nonterminal nitrogen; on the other hand, the destabilizing effect of the oxygen atom would favor protonation at a terminal nitrogen. We can get some indication of the relative importance of the two effects by comparing, in turn, the pK of the guanidine group in arginine—12.48—and that of the O–NH_2 group in canaline—3.95—with the pK of the ε-NH_2 group in lysine, 10.52. The implication is that the destabilizing effect of the oxygen atom is more important than the stabilizing effect of equivalent amino groups and hence that protonation at a terminal nitrogen atom would be favored. But proof must await experiment.

Acknowledgment. We are grateful to Dr. W. Croasmun and the Southern California Regional NMR facility (NSF Grant No. CHE-7916324) for preparing the ^{15}N NMR spectrum of canavanine sulfate, to Drs. J. D. Roberts and K. Kanamori for their help in interpreting the arginine spectra, and to Dr. G. A. Rosenthal for helpful advice. This investigation was supported in part by Public Health Service Research Grant No. GM 16966 from the National Institute of General Medical Sciences, National Institutes of Health.

Registry No. *l*-Canavanine, 543-38-4; *l*-canaline, 496-93-5.

Supplementary Material Available: Tables of structure factor amplitudes, anisotropic Gaussian parameters for C, N, and O, isotropic B's for H, and computational details for pK values of *l*-canavanine and *l*-canaline (10 pages). Ordering information is given on any current masthead page.

(11) Kistenmacher, T. J.; Rand, G. A.; Marsh, R. E. *Acta Crystallogr., Sect. B* **1974**, *B30*, 2573–2578. Marsh, R. E.; Ramakumar, S.; Venkatesan, K. *Ibid.* **1976**, *B32*, 66–70.
(12) Marsh, R. E.; Donohue, J. *Adv. Protein Chem.* **1967**, *22*, 235–256.
(13) Hale, W. J.; Vibrans, F. C. *J. Am. Chem. Soc.* **1918**, *40*, 1046–1063.
(14) McKay, A. F.; Picard, J. P.; Brunet, P. E. *Can. J. Chem.* **1951**, *29*, 746–758.
(15) Schwenker, G. *Arch. Pharm. (Weinheim, Ger.)* **1962**, *295*, 753–758.
(16) Sullivan, G. R.; Roberts, J. D. *J. Org. Chem.* **1977**, *42*, 1095–1096.
(17) Suzuki, T.; Yamagushi, T.; Imanari, M. *Tetrahedron Lett.* **1974**, *20*, 1809–1812.
(18) Blomberg, F.; Maurer, W.; Rueterjans, H. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1409–1413.
(19) Kanamori, K.; Cain, A. H.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 4979–4981.