Complete Amino Acid Sequence of Guanaco (*Lama guanicoe*) Cytochrome *c*†

R. L. Niece, E. Margoliash, and Walter M. Fitch*

ABSTRACT: The amino acid sequence of the guanaco, *Lama guanicoe*, a South American tylopod closely related to the llama, has been completely determined. The previously reported, not completely sequenced cytochromes *c* of the whale and camel are probably identical to that of the guanaco.

The comparative study of amino acid sequences from a collection of orthologous proteins has proven useful in many ways. These include: testing of methods for reconstructing phylogenies where there is good paleontological basis for knowing the correct answer, as is true for many vertebrates; utilization of the proven methods for obtaining probable phylogenies where no good paleontological material is available; examining the structural effects of changes in sequence; examining the functional effects of changes in sequences; recognizing antigenic determinants in molecules; studying evolutionary rates in structural genes; studying mechanisms of genetic change and variability. Such studies on cytochrome *c* have been particularly fruitful because of the large number of their known sequences. This work provides the complete amino acid sequence of guanaco (*Lama guanicoe*) cytochrome. Two previously examined, but not completely sequenced, cytochromes *c* (whale, *Rachianectes glaucus* (Goldstone and Smith, 1966) and camel, *Camelus dromedarius* (Sokolovsky and Moldovan, 1972)) are probably identical to that of the guanaco.

Materials and Methods

Cytochrome *c* was extracted and purified from homogenates of frozen guanaco heart muscle by the method of Margoliash and Walasek (1967). A 20-ml aqueous solution containing 164 mg of cytochrome *c*, 2 mmol of NH_4HCO_3, and 4 mg of TPCK[1] trypsin (Worthington Biochemicals) was incubated at 40 °C for 5 h and then frozen and lyophylized. The lyophylized material was dissolved in water and centrifuged. The material that did not dissolve was called the digest-insoluble fraction. The supernatant fluid was taken to pH 4.0 with acetic acid at which pH the solution tended to gel. It was readjusted to pH 3.0 and applied to a column identical to that described by Nolan et al. (1973). The elution procedure was also that of Nolan et al. (1973), except that the starting pH was 3.0, and 5.0-ml fractions were collected. Aliquots of 150 μl were taken from every second tube for direct ninhydrin analysis. The results are shown in Figure 1.

Preparative paper electrophoresis on Whatman 3MM paper at pH 6.5 and 50 V/cm was used for further purification of peptides when required. Peptide mobility determinations and electrophoretic separations were performed on thin layers of cellulose at pH 6.5 and 20 V/cm (Nolan et al., 1971). Ninhydrin-negative peptides were located by dipping the dried thin layer in 0.1% bromcresol green in acetone (Greenstein and Winitz, 1961); it is suggested here that the use of fluorescent pH indicators would provide an extremely sensitive and nondestructive assay for acidic and basic peptides on fingerprints.

Heme Removal. Heme removal was effected by iodination (Lederer and Tarin, 1971) or the sulfenyl choloride method (Fontana et al., 1973a,b). Cysteine was converted to (Konigsberg, 1972) and identified as carboxymethylcysteine or as hydroxypropylcysteine (or its derivatives) after coupling with propylene oxide (G. E. Tarr, personal communication) as follows. Dissolve dry peptide in 0.1 M *N*-methylmorpholine, 0.05 M acetic acid, pH 7–8; add excess 2-mercaptoethanol, then excess propylene oxide. After incubation at room temperature for 30 min dilute with an equal volume of water and extract excess reagent with water-saturated ethyl ether and ethyl acetate.

Analytical Procedures. Amino acid analyses were performed as previously described (Niece, 1975) using hydrochloric or methanesulfonic acid for hydrolysis, and calculated using algorithms described by Smithies et al. (1971).

Sequencing using the Edman degradation was done in the liquid phase manually (Tarr, 1975) or automatically (Smithies et al., 1971) or by the solid-phase technique (Laursen, 1972) using diisothiocyanate (Laursen et al., 1972). Sulfonated isothiocyanates (Braunitzer et al., 1970, 1971), volatile buffers (Edman, 1970; Hermodson et al., 1972), or sequential degradation followed by dansylation (Gray, 1972a,b) were used where necessary. Identification of PhNCS-amino acids was by thin-layer chromatography (Tarr, 1975), or acid or alkaline hydrolysis (Smithies et al., 1971) followed by amino acid analysis. Hydrolysis was used to determine Asx or Glx; the acid/amide status was determined from TLE or net charge of the peptides on fingerprints.

Partial acid hydrolysis was in 0.03 N HCl 1 h at 100 °C followed by chromatography on Dowex-50 in 0.2 M pyridinium acetate, pH 3.1.

Thermolysin (Matsubara, 1970) was used at 1/2000 (mole of enzyme/mole of peptide) for 2 h (T-5) to 38 days (T-1) at 20–50 °C to cleave tryptic peptides. Leucine aminopeptidase (Light, 1967) was used to digest T-19 for amino acid analysis. Carboxypeptidase B digestion (Ambler, 1972a,b) followed by amino acid analysis identified the C-terminal residue of selected tryptic peptides.

†From the Department of Physiological Chemistry (R.L.N. and W.M.F.), University of Wisconsin Medical School, Madison, Wisconsin 53706, and the Department of Biochemistry and Molecular Biology (E.M.), Northwestern University, Evanston, Illinois 60201. *Received April 19, 1976.* This work was supported by grants from the United States Public Health Service (AM15282, GM19121, AI2001, and HL1119), and the National Science Foundation (GB32274 and BM575-20109).

[1]Abbreviations used are: TPCK, L-1-tosylamido-2-phenylethyl chloromethyl ketone; TLE, thin-layer electrophoresis; BPNS-skatole, 2-(2-nitrophenylsulfenyl)-3-methyl-3-bromoindolenine; amino acid > PhNCS, a phenylthiohydantoin derivative of the amino acid indicated.

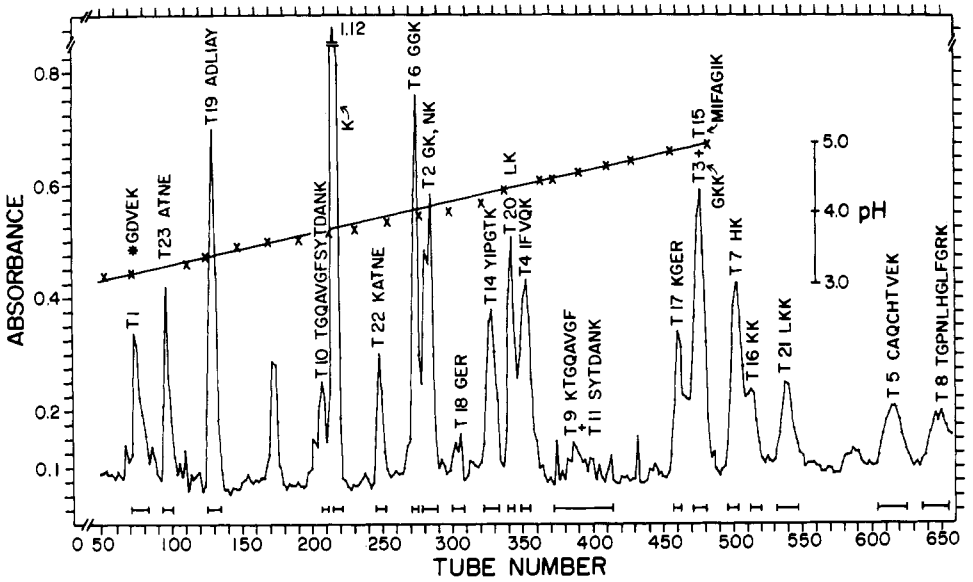

FIGURE 1: The soluble portion of the tryptic digest was chromatographed on Dowex AG 50W-X2 as described by Nolan et al. (1973). The abscissa gives the fraction number (5 ml/tube); (—) ninhydrin color (absorbance at 570 nm), determined on even-numbered tubes; (x-x) pH. Peaks are labeled according to their ultimate position in the sequence as shown in Figure 2. In addition, the sequence of that peptide is given in the IUB single letter code. The asterisk preceding the sequence of T1 is to indicate that its N terminus is blocked by an acetyl group. The fractions pooled for each peak are shown by the horizontal lines below them.

The tryptophanyl peptide bond was cleaved by the method of Fontana et al. (1973b) with BNPS-Skatole (Pierce Chemical Co.) with 6 daily additional aliquots of half the initial quantity of reagent. The resulting peptide mixture was sequenced directly.

Peptide Nomenclature. Tryptic peptides are numbered consecutively from the N terminus. Subfragments of tryptic peptides are similarly numbered to indicate their position in the tryptic peptide. Suffixes are also used to indicate special derivatives of residues of that tryptic peptide.

Results

The compositions of peptides isolated from a tryptic digest of the cytochrome *c* are shown in Table I. Peptides derived from those tryptic peptides and their compositions are included in the table. Tryptic peptides T-12 and T-13 were not found in the effluent of the Dowex-50 column, rather they precipitated from the digestion mixture when the pH was adjusted to 4 and were subsequently purified by preparative paper electrophoresis.

The amino acid sequence of the cytochrome (Figure 2) was determined from the sequences of the tryptic peptides. The ordering of the peptides was deduced by assuming homology with the structures of cytochromes *c* from other eukaryotic species. The amino acid sequences of the peptides were determined as described under Methods and as shown in Figure 2. To prove that no unsuspected problems remained after determining a sequence, in many cases, sequencing was continued beyond the expected C terminus of the peptide, although only those data necessary to establish the sequence are presented in Figure 2. The sequencing data pertaining to those peptides that presented special problems are summarized below.

Peptide T-1 (Residues 1–5). Although this peptide reacted with ninhydrin, it was expected to have a blocked amino terminus. Amino acid composition and net charge at pH 6.5 supported that and indicated that both Asx and Glx must be acids. Carboxypeptidase B digestion of T-1 released only lysine. Thermolysin peptide T-1-Th1 was ninhydrin negative, doubly negatively charged, and composed of glycine and aspartic acid. The C-terminal portion of T-1 was sequenced using T-1-Th2. The drop-through fraction from Dowex-50 of a partial acid hydrolysis of T-1 yielded three partially separated peptides. T-1-PA1 and T-1-PA2 were ninhydrin negative, while T-1-PA3 was free aspartic acid. Peptide T-1-Th1 was esterified with 1 N acetyl chloride in methanol by the method of Tarr (1975) and examined by mass spectroscopy. A molecular ion of 260 and other fragments confirmed the sequence (block)-Gly-Asp and established that the blocking group was acetyl.

Peptide T-2 (Residues 6–7, 54–55). Nonintegral values for the amino acid composition for T-2, a small peptide, indicated the presence of two peptides in a ratio of about 2:1. Two peaks eluting at about the position of diiodotyrosine on the amino acid analyzer (Niece, 1975) were observed in the unhydrolyzed sample. Sequence analysis by both Edman and dansyl-Edman degradation showed two residues in the first cycle. Only lysine occurred in the second cycle of manual degradation and nothing was observed from cycles three and four. Of the two peptides, Gly-Lys and Asn-Lys, the former was assigned to residues 6–7 and Asn-Lys assigned to positions 54–55 by homology. This assignment is supported by the isolation and sequencing of the T-3 Gly-Lys-Lys peptide.

Peptide T-5 (Residues 14–22). Initial attempts to sequence without cleavage of the thioether bonds were successful through seven cycles but left uncertainty with respect to Cys and Ser at positions 1 and 4 and Glx at position 3. Removal of the heme with 2-nitrophenylsulfenyl chloride and coupling the cysteine residues with propylene oxide permitted sequencing to the C terminus. The last three residues were also sequenced as a thermolytic peptide (T-5-Th2).

Peptide T-9 (Residues 39–46), Peptide T-10 (Residues 40–53), and Peptide T-11 (Residues 47–53). No peptide covering residues 39–53 was available and T-10 yields fell drastically at about cycle seven of the Edman degradation, presumably because of a peptide N → O acyl shift. The overlap of T-9 and T-10 established the sequence from residue 39 through 49, except for the amide status of Glx-42. Ser-47 of peptide T-10 was identified by amino acid analysis: the yield of alanine in cycle eight after acid hydrolysis was much higher

TABLE I: Amino Acid Compositions of Tryptic Peptides.[a]

	T-1 (Res 1–5)						T-2[f] (Res 6–7, 54–55)	T-3 (Res 6–8)	T-4 (Res 9–13)
		–Th1[e]	–Th2	–PA1[e]	–PA2[e]	–PA3			
Asx	0.96	1.00	0.03		0.58	1.00	0.66	0.01	0.03
Thr	0.02								0.05
Ser	0.01	0.01	0.01		0.04		0.02	0.01	
Glx	0.99		0.73	0.01	0.07		0.02		1.08
Pro			0.05						
Gly	1.01	0.99	0.04	1.00	1.00		0.34	1.04	0.08
Ala		0.03	0.01						
Half-Cys									
Val	0.90		1.05		0.06				0.89
Met									0.06
Ile									0.88
Leu									0.05
Tyr									
Phe									0.89
Lys	1.00		1.00				1.00	2.00	1.00
His									0.05
Trp									
Arg									
Other									
Net charge[b]	–2	–2	0	–	–	–	+1	+2	+1
Color reactions[c]									
Purification procedure[d]		TLE	TLE	CC	CC	CC			

	T-14 (Res 74–79)	T-15 (Res 80–86)	T-16 (Res 87–88)	T-17 (Res 88–91)	T-18 (Res 89–91)	T-19 (Res 92–97)	–LAP	T-20 (Res 98–99)	T-21 (Res 98–100)	T-22 (Res 100–104)	T-23 (Res 101–104)
Asx	0.01	0.07	0.01		0.05	1.26	0.63	0.02		1.15	0.98
Thr	0.95	0.04						0.01		1.01	0.98
Ser		0.09	0.14	0.01	0.11	0.05	0.01	0.02	0.02	0.02	
Glx	0.01	0.05		1.00	1.52	0.07		0.06		0.98	1.00
Pro	1.09										
Gly	1.00	1.20	0.12	1.01	1.48	0.07		0.05	0.10	0.11	0.03
Ala		1.04				2.38	1.32		0.01	1.10	0.99
Half-Cys											
Val											
Met		0.59									
Ile	0.90	1.79				1.26	0.72				
Leu						1.33	0.63	1.05	1.09		
Tyr	0.91					1.00	1.00				
Phe		0.95									
Lys	1.00	1.00	2.00	1.00	0.09			1.00	2.00	1.00	
His											
Trp											
Arg				0.97	1.00						
Other											
Net charge[b]	+1	+	+	+1	0	–1		+1	+2	0	–1
Color reactions[c]				PQ	PQ						
Purification procedure[d]											

[a] The values are relative to the number of basic residues, if they are present. ND, specifically not determined. [b] Net charge determined as described under Methods. Integral values are presented with their sign. If the value was not near integral or if the peptide was too large or complex for calculations, only the sign is given. [c] PQ, positive phenanthrenequinone stain for arginine; Ehrlich, positive test for tryptophan. [d] Repurification of subfragments: TLE, thin-layer electrophoresis; CC, column chromatography; PPE, preparative paper electrophoresis. [e] No free amino group. [f] Mixture of 2 peptides. [g] Derivatives produced after removal of heme group: CMC, carboxymethylcysteine; HPC = hydroxypropylcysteine.

than after alkaline hydrolysis, whereas in cycle four the alanine yields were the same from both hydrolyses. An anomalous tryptic cleavage at Phe-46 resulted in peptide T-11. A three-residue overlap with the sequenced portion of T-10, the composition and sequence of T-11, the approximate equimolar yields of T-9 and T-11, and the usual specificity of trypsin permitted the assignment of the entire sequence from residues 39 to 53. The low level of serine (0.5 mol) and the presence of 0.5 mol of glycine remains unexplained. Valine, not proline (Dayhoff and Eck, 1967), is clearly the amino acid at position 44, as it is in the camel (Sokolovsky and Moldovan, 1972).

Peptide T-12 (Residues 56–72) and Peptide T-13 (Residues 56–73). No tryptophan-containing peptides were found in the soluble fraction of the tryptic digest. Two negatively

T-5g				T-6	T-7	T-8	T-9	T-10	T-11	T-12	T-13
Res 14–22	Res 14–22	Res 14–22	–Th2	(Res 23–25)	(Res 26–27)	(Res 28–38)	(Res 39–46)	(Res 40–53)	(Res 47–53)	(Res 56–72)	(Res 56–73)
	0.13		0.02		0.01	1.02	0.12	1.77	2.09	1.04	1.20
1.17	0.84	1.24	0.05			0.92	0.94	1.70	1.03	1.67	2.20
	0.08				0.01		0.14	0.89	0.52	0.33	0.22
1.89	1.82	1.81	0.98			0.05	1.09	0.95	0.19	4.17	5.02
						1.32				0.58	0.76
0.04	0.13	0.07	0.19	1.82	0.01	2.50	2.04	1.77	0.45	2.25	2.05
1.00	0.85	1.03					1.00	1.79	1.21		0.64
0.68	0.02										
0.94	1.05	1.08	1.19				1.10	0.88	0.15		0.50
										0.96	0.94
						0.02	0.06		0.02	1.04	0.94
0.04						2.00	0.05		0.07	1.96	2.00
								0.80	0.85	0.29	1.06
						1.08	0.90	0.91			
1.00	1.00	1.00	1.00	1.00	1.00	0.02	1.00	1.00	1.00	1.00	2.00
1.02	0.84	0.77			0.87	1.05					
										ND	0.74
						1.00					
	CMC 0.84	HPC ND									
ND	–	–	0	+1	+1.5	+1.5	+1	0	0	–	–
heme color			TLE			PQ				Ehrlich PPE	Ehrlich PPE

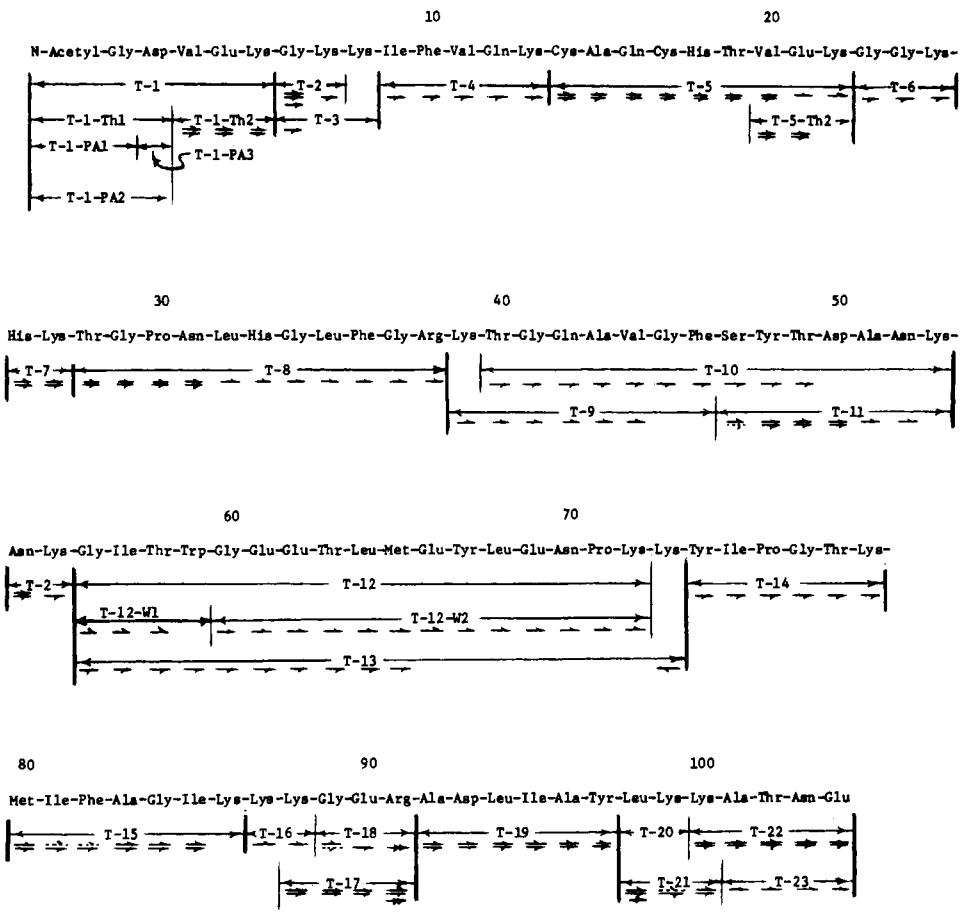

FIGURE 2: The tryptic peptides, indicated by the double-headed arrows, and the proposed amino acid sequence of the guanaco cytochrome *c* sequence: (→) indicates residue placed by Edman degradation and PhNCS > derivative identified by thin-layer chromatography; (⇀) Edman degradation and identification by amino acid analysis after hydrolysis of PhNCS > derivative; (↠) Dansyl-Edman; (⇻) denotes $n + 1^{th}$ residue after n cycles of Edman degradation identified by amino acid analysis without hydrolysis; (···· and ····) denote residues for which the PhNCS derivative residue was not identified; (⇀) indicates identification by amino acid analysis after carboxypeptidase digestion. The heavy vertical lines separate fragments for which there are no overlapping peptides. The order of these fragments was assumed by homology to other cytochromes *c*.

charged tryptophan (Ehrlich stain) peptides were purified by preparative paper electrophoresis from the insoluble portion. The amino acid compositions were not integral because of some visibly contaminating heme peptide on the paper, but represent two peptides differing in lysine content (T-12 being more negatively charged). The tryptophylglycine bond of T-12 was cleaved and manually sequenced without separation. Two different residues were seen in cycles two and three only. Based on the sequence of T-13 (see below), the shorter sequence (T-12-W1) represented the N terminus of the original peptide. The longer fragment (T-12-W2), which was sequenced to the terminal lysine, corresponded to the post-tryptophan sequence. Peptide T-13 was automatically sequenced successfully for ten cycles and provided an overlap of six residues with T-12-W2. Carboxypeptidase B released lysine in greater than 75% yield from T-13, showing the presence of two lysine residues after proline.

Peptide T-16 (Residues 87–88). T-16 was an oligopeptide composed only of lysine. Manual sequencing and direct identification and quantitation (alkaline hydrolysis) showed lysine in the first two cycles and less than one-sixth as much in the third cycle. Tryptic specificity indicates that the presence of three consecutive basic residues is necessary to produce lysyllysine from the interior of a protein.

Peptide T-19 (Residues 92–97). An efficient chymotryptic-like split, as has been observed by others (see Nolan et al., 1973, for several references), produced this peptide and no peptide extending beyond tyrosine 97 was observed. Leucine aminopeptidase was used, in addition to net charge, to confirm that position 93 was acidic.

Discussion

The sequence of the guanaco cytochrome *c* is rather more firmly established than sequences usually are when there are previously determined and closely related homologous sequences known. While the major fragments are aligned assuming homology, every amino acid has two, and frequently more, pieces of evidence to support the assignment. Only three amino acids were not identified directly, namely, lysines 8, 53, and 86. In each case, however, there are two supports for the assignment: (1) every other residue in the peptide's composition had already been specifically identified from Edman degradation; (2) the specificity of tryptic cleavage strongly supports the lysine being C terminal in the peptide. In no case were the last two residues of a peptide assigned using the tryptic specificity to assign the C-terminal residue and the composition to assign the penultimate residue as has been done, albeit without error, for whale and camel (Goldstone and Smith, 1966; Sokolovsky and Moldovan, 1972) sequences of cytochrome *c*. Moreover, in both the whale and camel cytochromes *c*, the heme peptides were largely or completely determined by homology to other cytochromes *c*. However, the compositions of all peptides determined were identical to those found in the guanaco, so that it is reasonable to assume that whale and camel cytochromes *c* are identical to that of the guanaco.

Acknowledgments

The authors are grateful to Mr. T. D. Evans of Abbott Laboratories for arranging the collection, frozen storage, and shipping of the guanaco hearts during his stay in Argentina, to Dr. O. Walasek for the initial purification of the cytochrome *c*, to Dr. G. E. Tarr for his invaluable advice and assistance, to Dr. K-T Lin for technical assistance, and to Dr. H. K. Schnoes for the mass spectral analyses.

References

Ambler, R. P. (1963), *Biochem. J. 89,* 349.
Ambler, R. P. (1972a), *Methods Enzymol. 25B,* 143.
Ambler, R. P. (1972b), *Methods Enzymol. 25B,* 262.
Braunitzer, G., Schrank, B., and Ruhfus, A. (1970), *Hoppe-Seyler's Z. Physiol. Chem. 351,* 1589.
Braunitzer, G., Schrank, B., Ruhfus, A., and Petersen, S. (1971), *Hoppe-Seyler's Z. Physiol. Chem. 352,* 1730.
Dayhoff, M. O., and Eck, R. V. (1967), Atlas of Protein Sequence and Structure, Silver Spring, Md., National Biomedical Research Foundation, p 88.
Edman, P. (1970), in Protein Sequence Determination, Needleman, S. B., Ed., New York, N.Y., Springer-Verlag, p 211.
Fontana, A., Veronese, F. M., and Boccu, E. (1973a), *FEBS Lett. 32,* 135.
Fontana, A., Vita, C., and Toniolo, C. (1973b), *FEBS Lett. 32,* 139.
Goldstone, A., and Smith, E. L. (1966), *J. Biol. Chem. 241,* 4480.
Gray, W. R. (1967), *Methods Enzymol. 11,* 469.
Gray, W. R. (1972a), *Methods Enzymol. 25B,* 121.
Gray, W. R. (1972b), *Methods Enzymol. 25B,* 333.
Greenstein, J. P., and Winitz, M. (1961), Chemistry of the Amino Acids, New York, N.Y., Wiley, p 1349.
Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H., and Walsh, K. A. (1972), *Biochemistry 11,* 4493.
Konigsberg, W. (1972), *Methods Enzymol. 25B,* 185.
Laursen, R. A. (1972), *Methods Enzymol. 25B,* 344.
Laursen, R. A., Horn, M. J., and Bonner, A. G. (1972), *FEBS Lett. 67,* 67.
Lederer, F., and Tarin, J. (1971), *Eur. J. Biochem. 20,* 482.
Light, A. (1967), *Methods Enzymol. 11,* 426.
Margoliash, E., and Walasek, O. F. (1967), *Methods Enzymol. 10,* 339.
Matsubara, H. (1970), *Methods Enzymol. 19,* 642.
Niece, R. L. (1975), *J. Chromatogr. 103,* 25.
Nolan, C., Fitch, W. M., Uzzell, T., Weiss, L. J., and Margoliash, E. (1973), *Biochemistry 12,* 4052.
Nolan, C., Margoliash, E., Peterson, J. D., and Steiner, D. F. (1971), *J. Biol. Chem. 246,* 2780.
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L. (1971), *Biochemistry 10,* 4912.
Sokolovsky, M., and Moldovan, M. (1972), *Biochemistry 11,* 145.
Tarr, G. E. (1975), *Anal. Biochem. 63,* 361.