Arch. Biol. Med. Exp. *24:* 67-74 (1991)
Printed in Chile

Sagaxine, the protamine from *Sardinops sagax*. Characterization, sequence and secondary structure prediction

Caracterización, secuencia y predicción de la estructura secundaria de la sagaxina, protamina del *Sardinops sagax*.

ALEJANDRO ARELLANO*; HUGO PEÑA-CORTÉS[1]; HILDA CID[2] and JÜRGEN HOPPE[3]

[1] Départamento de Química, Universidad de Concepción, Chile.
[2] Departamento de Biología Molecular, Universidad de Concepción, Chile.
[3] Institut für Physiologische Chemie, Universität Würtzburg, West Germany.

This paper describes the isolation, purification and full characterization of the protamine from *Sardinops sagax*. The protein was purified by both gel filtration and ion exchange chromatography and its amino acid composition, electrophoretic behaviour and sequence were determined. We postulate that the protamine exhibits microheterogeneity with a major and a minor component. The molecular weight for the major component is 4377, as calculated from the sequence. The N-terminal amino acid is proline and the C-terminal is arginine. The amino acid composition is:

$$Thr - Ser - Glx - (Pro)_2 - Gly - (Ala)_3 - (Val)_2 - (Lys)_2 - (Arg)_{20}$$

The secondary structure for the protamines was also predicted and a revision of our previously published results in this area is discussed. We have called this protamine Sagaxine and an overall comparison with the hitherto described fish protamines is also made.

INTRODUCTION

Protamines are the smallest and most basic known proteins and are found associated with the DNA in the sperm nuclei of several animal species (1). Since the first reported work done with protamines by Miescher on 1874 (2) and the start of their proper study by Kossel in 1986 (3), this protein material has been widely used for the study of sperm chromatin structure (4-7), on cardiological surgery (8) and as a factor against the angiogenesis produced by inflamatory and tumor processes (9).

Several fish protamines have been studied so far (1), all of them being extracted from fishes captured in the Northern Hemisphere. The geographic location of Chile, with more than 4000 km seaboard and a capture over 2.2 million tons of sardine during 1987 (10) provides sufficient interesting biological material for the study of protamines.

This work describes the isolation, purification, chemical characterization and sequence determination of the major basic sperm proteins extracted from *Sardinops sagax*. Its secondary structure is predicted and compared with other fish protamines, including a revision of our own previously published results (11). An overall comparison with the hitherto described protamines is also made.

MATERIALS AND METHODS

Isolation of sperm nuclei and protein extraction: Male gonads from mature *Sardinops sagax* captured in the Gulf of Arauco, Chile (Lat. 37°3'S, Long. 73°20'W), were frozen immediately after removal from the living fish. The pure sperm heads were obtained according to a method already described (12) and washed successively with acetic acid, distilled water, pure

* Correspondence address: Dr. Alejandro Arellano, Department of Microbiology and Infectious Diseases, Flinders Medical Centre, Bedford Park SA 5042, Australia.

ethanol and diethylether. The cells were then dried over P_2O_5 overnight before protein extraction through the use of 0.2 N sulfuric acid under homogenization with an Ultraturrax device, followed by centrifugation at 5000 r.p.m. in a JA 14 rotor. The extraction was repeated until no more protein was detected in the supernatants using the Biuret method (13). The proteins were then precipitated overnight with cooled ethanol (1:4 in volume) and after suitable centrifugation, were dissolved in 0.5 N acetic acid and desalted on Bio-Gel P-2. The protein fractions obtained were freeze-dried and stored.

Protein purification and analysis: The separation of protamine from other basic sperm proteins and/or residual histones was performed both by gel filtration on Sephadex G-75 and by ion exchange chromatography on CM Sephadex. As criteria of purity as well as for the apparent molecular weight estimations, electrophoresis on 20% urea/acetic acid polyacrylamide gels (14) was used. Histones from *Tetrapygus niger* spermatozoa, whole clupein (Sigma) and a bull protamine crude extract were used as standards during the runs. Since protamine is very hygroscopic, the samples for amino acid analysis were weighed at different times after opening a dessicator and the weights extrapolated to time zero. The samples were then hydrolyzed with 6 N HCl at 110°C for 24, 48 and 72 hours under vacuum, according to the standard method of Moore and Stein (15). The analyses were performed on a Beckman 119 CL Analyser equipped with an integrator model 126 data system. The AACOM2 program was used to obtain the best fit of the number of residues for a given molecular weight of the protein. This program is available from us on request. The N-terminal analyses were carried out using the dansyl method according to Hartley (16) and the DNS-amino acids were identified on Cheng Chin polyamide sheets from Pierce. The C-terminal analyses were done by the enzymatic method described by Ambler (17), using carboxypeptidase B (Sigma, lot 74F-8125). The released amino acids were then identified as above, norleucine (Pierce) being used as the internal standard.

Sequence determination and secondary structure prediction: The sequence of the purified protein fractions were determined both by the solid phase and the gas phase methods. For the solid phase method, the protein was derivatized with ethylenediamine in the presence of EDC (1-Ethyl-3-(3-aminopropylethylene) carbodiimide) from Sigma to add an amino group at the carboxy end. The protein was then reacted with DITC-glass (18) and sequenced as described by Hoppe and Sebald (19). A gas phase sequencer model 470 (Applied Biosystems) with an updated standard program for the gas phase method was used. Conversion was done with 25% trifluoracetic acid. PTH amino acids were separated and quantified on a Spectra-Physics 8700 HPLC System, using a Dupont Zorbax CN column (5 μm, 0.45 x 25 cm). Elution was performed by applying a 18 min linear grandient from 16% B in A to 72% B in A (A: 70mM sodium acetate, 5% tetrahydrofuran pH 3.8 and B: 60% acetonitrile).

The secondary structure prediction was made using the method by Chou and Fasman (20), based on empirical probabilities. Conformational parameters $P\alpha$, $P\beta$ and P_t are defined for each one of the 20 natural amino acids, which represent the normalized frequency of occurrence of a particular amino acid residue in a certain type of secondary structure, as obtained from a data base of 29 fully determined protein structures. An average $< P_j >$ greater than 1.0, obtained for a group of amino acids taken in the sequence (6 for an helix, 5 for a β-strand and 4 for a β-turn) is an indication that the structure type **j** is likely to occur in that region of the sequence. A modification that substitutes the average of the conformational parameters with their product, in order to improve the sensibility of the method in the vicinity of the limit value 1.0 has been used in this work (11). Two other modifications of the method, one that allows a differentiation between β-strands participating in parallel or antiparallel β-sheets (21), and other that considers the fre-

quencies of occurrence of the amino acid on each of the four positions in a β-turn, were employed (22). According to this last technique, β-turn would occur if $< P_t >$ is greater than 1.0, and greater than $< P_\alpha >$ and $< P_\beta >$ and in addition to it, the product of the frequencies of occurrence of the amino acid participating in the β-turn, in their respective positions, is greater than 0.55×10^{-4} gives a more reliable prediction.

RESULTS

Protein extraction and purification: The gonads of *Sardinops sagax* during spawning are full of mature spermatozoa and occupy almost the whole abdominal cavity of the fish. Our extraction procedure yielded sperm heads free from tails and other structures. Acid extraction of these sperm heads yielded both protamines and histones. To separate them, gel filtration was performed (Fig. 1a.). Here the histones were eluted first and the protamine later in a well defined, symmetrical peak, suggesting an homogeneously sized molecular population. However, this result alone cannot be considered as a criterion for homogeneity. Therefore a separation was carried out using CM Sephadex, eluting with a salt gradient, similar to that used by us to resolve the three components of clupein (12). The elution profile shown in Fig. 1b suggests homogeneity as the protamine elutes in one symmetrical peak.

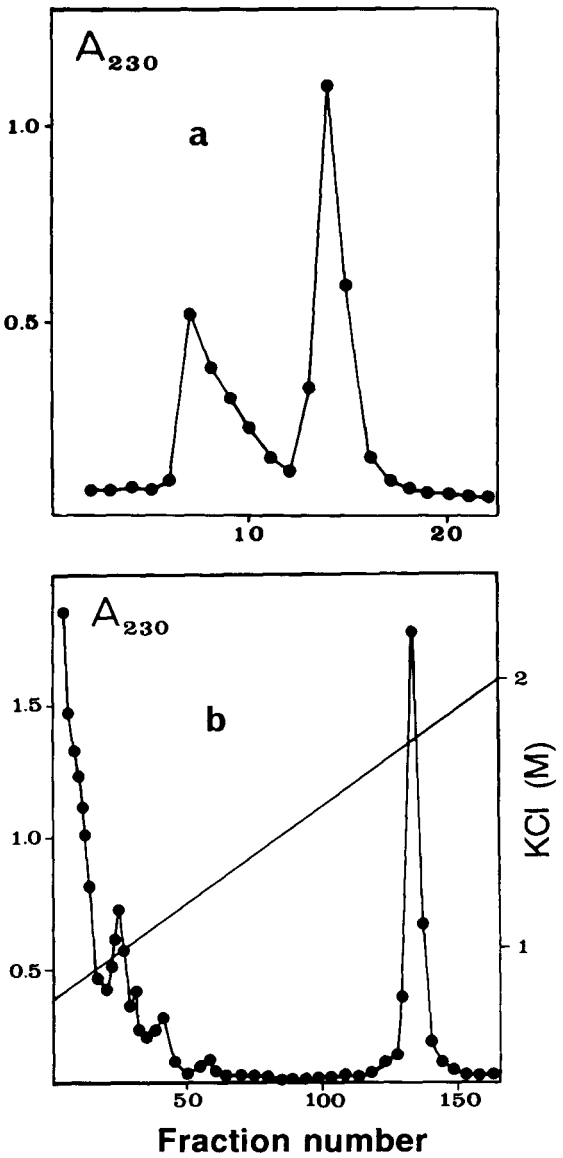

Fig. 1: Separation of histones and protamines from *Sardinops sagax* on:

a) Sephadex G-75 (1.5 x 40 cm), eluted with 0.15 M NH_4HCO_3 pH 7.5.

b) CM Sephadex (2.5 x 80 cm), eluted with a KCl gradient from 0.8 to 2.0 M in 50 mM sodium acetate, pH 6.8.

Electrophoretic behavior and amino acid composition: The electrophoretic mobility of proteins on polyacrylamide gels is an accepted criterion for purity and homogeneity, as well as a good estimation of molecular size. Since protamines precipitate in SDS gels, we used the Panyim and Chalkley (14) gel system commonly used for histones. The photograph in Fig. 2 shows, as expected, the high mobility of our protein is comparable to those of the clupein and bull protamine standards. If we assume that different protamine molecules exhibit similar axial positive charges as a result of their high content of arginyl residues, it is appropriate to use these gels to estimate apparent molecular weight, according to Rodbard and Chrambach (23). From a double logarithmic plot of mobility versus molecular weight, we obtained a value of 4300 Da for our protein. The amino acid composition was determined at three hydrolysis times and the rate of destruction of labile amino acids such as serine and threonine was calculated using the AACOM2 program by

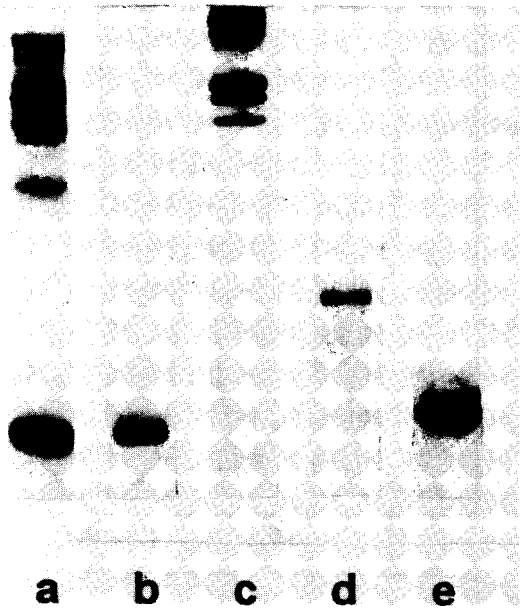

Fig. 2: 20% acetic acid-urea polyacrylamide gel according to Panyim and Chalkley (14).

A: Whole basic protein extract form the sperm heads of *Sardinops sagax*.
B: Sagaxine after purification.
C: *Tetrapygus niger* histones.
D: Bull protamine crude extract.
E: Whole clupeins.

extrapolating their concentration to time zero. The program also allows a fitting of the best number of residues for a given molecular weight. As can be seen on Table 1, our protein, like other fish protamines, lacks aromatic and cysteinyl residues. The arginine content is similar to that of typical protamines. However two unusual lysine residues, which are not commonly present in other fish protamines, were found. This finding classifies our protein as a diprotamine, according to Kossel (24). The molecular weight calculated from the amino acid composition is 4377, being in good agreement with that determined from the electrophoretic experiments. It was dificult to obtain constant values for serine throughout several analyses, the AACOM2 program indicating one or two residues for the same molecular weight and for different hydrolysates. N-terminal analysis clearly yielded proline, since no other DNS-amino acid spot was observed in repeated assays. Carboxypeptidase Y was used for the C-terminal analysis without positive results, probably due to the strongly basic character of the protein. Carboxypeptidase B was then used and a large amount of arginine was rapidly released.

Sequence: The sequence determined by the solid phase method was almost totally confirmed by the more accurate gas phase method (Fig. 3). Positions 8, 23 and 26, which were uncertain by the solid phase method, were identified and assigned to be serine, lysine and glycine, respectively. It is interesting to take into account the presence of a small amount of serine in addition to alanine detected by both methods at position 7. This fact, together with the finding of alanine traces at the N-terminal position given by gas phase sequence experiments, supports the

TABLE 1

Amino acid composition, in residue number, of the protamine from *Sardinops sagax* as compared to other fish protamines. Sagaxine has 33 residues and the molecular weight derived from its amino acid composition is 4377

	Sagaxine	Salmine A1	Iridine IA	Thynnin Y1	Clupein Z
Thr	1	–	–	1	–
Ser	1	4	4	2	3
Glx	1	–	–	1	–
Pro	2	3	3	2	2
Gly	1	2	2	–	–
Ala	3	1	–	3	3
Val	2	2	2	2	2
Lys	2	–	–	–	–
Arg	20	20	21	21	21
Tyr	–	–	–	1	–

Sagaxine 1:

1 10 20
Pro - Arg - Arg - Arg - Glx - Lys - Ala - Ser - Arg - Pro - Val - Arg - Arg - Arg - Arg - Arg - Ala - Arg - Val - Thr - Arg -

30
Arg - Lys - Arg - Ala - Gly - Arg - Arg - Arg - Arg - Arg - Arg - Arg.

Sagaxine 2:

1 10 20
Ala - Arg - Arg - Arg - Glx - Lys - Ser - Ser - Arg - Pro - Val - Arg - Arg - Arg - Arg - Arg - Ala - Arg - Val - Thr -

30
Arg - Arg - Lys - Arg - Ala - Gly - Arg - Arg - Arg - Arg - Arg - Arg - Arg.

Fig. 3: Sequences of Sagaxine 1 and 2, the major and minor components of the protamine from *Sardinops sagax*.

existence of microheterogeneity in this protamine. The number of arginine residues at the C-terminal extreme of the molecule was determined based only on the precision of amino acid analysis, it being impossible to follow the kinetics of amino acid release using Carboxypeptidase B, due to the long arginine tract at that end.

Secondary structure prediction: The secondary structure prediction for the two components of the protamine obtained from *Sardinops sagax*, according to the Chou and Fasman's method with all the modifications included, is presented on Table 2. For each sequence analysed, the averaged conformation parameters $<P_\alpha>$ and $<P_\beta>$, as well as the conformational products Π_α and Π_β, are shown. Also, in the prediction of β-strand, the values of the conformational parameters obtained when the amino acid residue participates on an antiparallel β-strand ($P_{\beta AP}$), are also shown. In the β-turn prediction, in addition to $<P_\alpha>$, $<P_\beta>$ and $<P_t>$, the parameter P_t (equal to the product of the frequencies of the first, second, third and fourth residues in the β-turn) is also shown.

DISCUSSION

Protamine polypeptides, very similar to each other, have been separated by ion exchange chromatography (25) and the separation has succeeded in spite of their small differences in amino acid composition. We have also used this method to separate cleanly the three components of the herring protamine, namely clupeine YI, YII and Z (12). When the same technique was applied to the protamine studied here, no positive results were obtained. This led us to think that this particular protamine was not heterogeneous. However, when sequencing was done, two findings supported the idea that a microheterogeneity does occur. Firstly, no constant amount of serine was obtained throughout several analyses and secondly, both sequencing methods indicated serine in addition to alanine in position 7, and also traces of alanine were found at the N-terminal position, where proline is the major residue. We postulate, therefore, that this protamine presents a major and a minor component, the amount of the first being much larger than the second (Fig. 3). The existence of the minor component could occur due to single base changes at the structural gene level, as postulated by Black and Dixon (26). In fact, the codons for proline, alanine and serine are closely related, so that a single base change can explain the replacement of proline by alanine and alanine by serine in the sequence of the minor component as compared to the major one.

An interesting fact should be considered when the secondary structure prediction of protamines is made by using the Chou and Fasman method: arginine, the major component of protamines, presents the following values for the conformational coefficients: $<P_\alpha> = 0.98$; $<P_\beta> = 0.93$; $<P_{\beta AP}> = 1.02$; $P_t = 0.95$. Arginine is then an "indifferent" amino acid; by itself

TABLE 2

Predicted secondary structure for the major and minor component of the protamine obtained from *Sardinops sagax*. The numbers in parenthesis correspond to Lifson and Sander's (21) conformational coefficients for amino acid residues involved in antiparallel β-strands ($P\beta_{AP}$)

Structure	Protamine	Sequence	$<P_\alpha>$	$<P_\beta>$	Π_α	Π_β	Observation
α-helix	major component	2-7	1.105	0.91	1.721	0.544	
		3-8	1.070	0.88	1.352	0.438	helical region
		4-9	1.070	0.88	1.352	0.438	
	major and minor component	12-17	1.053	0.91	1.284	0.578	
		13-18	1.053	0.91	1.284	0.578	helical region
		14-19	1.067	0.04	1.388	1.056	
		15-20	1.042	1.085	1.176	1.351	
		16-21	1.042	1.085	1.176	1.351	helical region
		17-22	1.042	1.085	1.176	1.351	or β-strand?
		20-25	1.058	0.925	1.287	0.588	helical region
	minor component	1-6	1.105	0.91	1.721	0.544	helical region
β-strand	major and minor component	11-15	0.996	1.084 (1.116)	0.987	1.272 (1.624)	β-strand
		16-20	1.054	1.116 (1.148)	1.200	1.452 (1.826)	β-strand or helical region?
		17-21	1.054	1.116 (1.148)	1.200	1.452 (1.826)	
		18-22	0.966	1.136 (1.172)	0.828	1.627 (2.069)	β-strand
		19-23	1.002	1.098 (1.116)	0.981	1.295 (1.501)	
			$<P_t>$	$<P_\alpha>$	$<P_\beta>$	$P_t \times 10^{-4}$	
β-turns	major component	1-4	1.093	0.877	0.668	0.910	β-turn
		7-10	1.140	0.935	0.765	0.567	β-turn?
	minor component	4-8	1.213	0.952	0.835	1.128	
		6-9	1.205	0.920	0.793	0.812	β-turn or
		7-10	1.333	0.733	0.745	1.123	helix 3_{10}?
	major and minor component	26-29	1.103	0.878	0.885	0.910	β-turn

it would not nucleate or disrupt any secondary structure under physiological conditions, but the presence of a strong "helix former" like alanine ($< P_\alpha > = 1.42$), or even a "medium-strong helix former" like lysine ($< P_\alpha > = 1.16$), into an arginine sequence would immediately favour a helix structure. Also the presence of serine or proline would initiate a β-turn, or the presence of valine would start a β-strand. Thus, just by looking at the sequences one should expect no helical zones in Salmine 1, Iridine IA and Iridine IB, and only one helix in the middle of the sequence of Iridine II. Helical regions could also occur near the N-terminal of Thynnines Y1 and Y2, but not for Thynnine Z1 and Z2, in agreement with Toniolo's predictions (27). The presence of helical regions was not detected in our previous publication (11),

because the prediction was made with the original Chou and Fasman's parameters, based on 15 fully determined protein structures (20), which assigned to arginine a helical conformation parameter $< P_\alpha > = 0.79$. The presence of β-strand at the C-terminal region of all four Thynnins predicted by us (11), is also explained by the presence of 2 valines ($< P_\beta > = 1.70$) in the middle of a long arginine strand. Stellin A and the protamine described here are expected to present the longer helical region of all the protamines analyzed by us in our previously reported work (11), due to their alanine content combined with the presence of lysine that substitutes arginine in some sites.

The secondary structure predicted for this protamine (Table 2) indicates the presence of turns and zones of random coiled structure similar to those found for other protamines (11, 27). β-turns are proposed in the sequences 1-4, 7-10 and 26-29. Three helical regions are also possible, two of them, including amino acid residues 12-19 and 20-25, are common to the two components of the protamine. Due to the differences at position 1 and 7 of the sequence, the major component shows helical probability in the sequence 1-6. Two regions with clear preferences for β-strands (antiparallel) are located in the sequence 11-15 and 18-22. The first β-strand is present in all protamines analyzed by us (11), because of the presence of a valine residue at position 11, 12 or 13. The only exceptions are Sturine B and Stellin A, where valine has been replaced by glutamine or lysine respectively, which do not show any β-strand preference. There are regions, such as 15-22 (Table 2) where the average conformational parameters $< P_\alpha >$ and $< P_\beta >$ are both greater than 1. In this case it is not sufficient just to see which one is greater to decide if a helix or a β-strand would be generated. Probably, if a helix has been nucleated, it will continue, even though $< P_\beta >$ could be a little higher than $< P_\alpha >$.

The facility to change the secondary structure of a protamine region, just by a single change in one amino acid residue, can also be found in changes in the environment of the molecule. It is well known that polyarginine tends to build helices, specially if the amount of water molecules is less than five per amino acid residue (28). This property could be important considering the physiological role of protamines, which is the DNA packing into a very condensed state, where water molecules would be strongly displaced. Arginine is the only one of the 20 amino acids which does not show any preference or dislike for any type of secondary structure, that is probably why it has been chosen to build the protamine molecules.

The fact that both this protamine and Stelline A present a more stable helix structure due to the presence of lysine in their sequence is interesting. *Acipenser stellatus*, the Northern Hemisphere sturgeon, source of Stelline A, is a rather bizarre fish being classified as midway between bony and cartilaginous fishes, while *Sardinops sagax* belongs to Family Clupeidae and is a well evolved bony fish. At present it is not possible to ascertain a relationship, but this similarity at the protamine molecular level could be important from an evolutionary point of view.

We have given the name Sagaxine to the protamine obtained from *Sardinops sagax* described in this work.

ACKNOWLEDGEMENTS

We thank M. Imschenetzky for providing us with *Tetrapygus niger* histones and M. Fornells for the bull protamine crude extract, and Mrs. Anke Johnsen for tidying up the manuscript. This work was partially supported by research grants 20.13.41 from the Dirección de Investigación, Universidad de Concepción, and 5041/85 from FONDECYT Chile, and a DAAD overseas travel grant to A. Arellano.

REFERENCES

1. ANDO, T.; YAMASAKI, M. and SUZUKI, K. (1973) in "Protamines". Springer Verlag, Berlin.
2. MIESCHER, F. (1874) *Ber. Deut. Chem. Ges. 7:* 376-379.
3. KOSSEL, A. (1896) *Z. Physiol. Chem. 22:* 176-187.
4. OTTENSMEYER, F.P.; WHITING, R.F. and KORN, A.P. (1975) *Proc. Nat. Acad. Sci.* USA *72:* 4953-4955.

5. WILLMITZER, L.; BODE, J. and WAGNER, K.G. (1977) *Nucleic Acids Res. 4:* 163-176.
6. WARRANT, R.W. and KIM, S.H. (1978) *Nature 271:* 130-135.
7. WILLMITZER, L. and WAGNER, K.G. (1980) *Biophys. Struct. Mech. 6:* 95-110.
8. GOLDMAN, B.S.; JOINSON, J. and AUSTIN, W.G. (1969) *Ann. Thorac. Surg. 7:* 459-471.
9. TAYLOR, S. and FOLKMAN, J. (1982) *Nature 297:* 307-312.
10. ANUARIO ESTADISTICA DE PESCA (1988) Fishing National Service, Ministry of Economy, Chilean Government.
11. CID, H. and ARELLANO, A. (1982) *Int. J. Biol. Macromol. 4:* 3-8.
12. ARELLANO, A. (1980) Ph.D. Thesis, University of Braunschweig, F.R.G.
13. ZAMMENHOF, S. (1957) *Methods Enzymol. 3:* 696-704.
14. PANYIM, S. and CHALKLEY, R. (1969) *Arch. Biochem. Biophys. 130:* 337-346.
15. MOORE, S. and STEIN, W.H. (1963) *Methods Enzymol. 6:* 819-831.
16. GRAY, W.R. (1967) *Methods Enzymol. 11:* 139-151.
17. AMBLER, R.P. (1972) *Methods Enzymol. 25:* 143-154.
18. HOPPE, J.; MONTECUCCO, C. and FRIEDL, P. (1983) *J. Mol. Biol. 258:* 2882-2885.
19. HOPPE, J. and SEBALD, W. (1980) *Eur. J. Biochem. 107:* 57-65.
20. CHOU, P.Y. and FASMAN, G.D. (1974) *Biochemistry 13:* 222-244.
21. LIFSON, S. and SANDER, C. (1979) *Nature 282:* 109-111.
22. CHOU, P.Y. and FASMAN, G.D. (1978) *Ann. Rev. Biochem. 47:* 251-276.
23. RODBARD, D. and CHRAMBACH, A. (1971) *Anal. Biochem. 40:* 95-134.
24. KOSSEL, A. and SCHENCK, E.G. (1928) *Z. Physiol. Chem. 173:* 278-308.
25. ARELLANO, A.; WEHLING, K. and WAGNER, K.G. (1984) *Int. J. Biol. Macromol. 6:* 249-254.
26. BLACK, J.A. and DIXON, G.H. (1967) *Nature 216:* 152-154.
27. TONIOLO, C. (1980) *Biochem. Biophys. Acta 624:* 420-427.
28. SUWALSKY, M. and TRAUB, W. (1972) *Biopolymers 11:* 623-632.