THE PRIMARY STRUCTURE OF HISTONE H2B FROM THE MOLLUSC PATELLA GRANATINA

by

P.D. VAN HELDEN.

Submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy
in the
Faculty of Science,
University of Cape Town.

July, 1978.

The Mollusc Patella granatina.

ACKNOWLEDGMENTS.

I gratefully acknowledge the help of the following people in the preparation of this work;

Professor C. von Holt for his valuable help and guidance and for the provision of excellent facilities.

Associate Professor W.N. Strickland for his helpful criticisms and suggestions and most of all for the friendly atmosphere in his laboratory.

Dr. W.F. Brandt for his technical assistance and guidance in the sequencing work.

Mrs. M. Strickland for her help with many of the technical aspects of this work.

My parents, for enabling me to do this work, and especially my mother, Mrs. Nancy van Helden, for so expertly and patiently typing this manuscript.

The University of Cape Town and the Council for Scientific and Industrial Research for their financial support without which this work would not have been possible.

Certification of Supervisors.

In terms of paragraph eight of 'General regulations for the degree of Ph.D' we, as supervisors of the candidate, P.D. van Helden, certify that we approve of the incorporation in this thesis of material that has already been published or submitted for publication.

Signed by officials

Professor C. von Holt.

Head of the Department of Biochemistry.

Signed by officials

Associate Professor W.N. Strickland.

Associate Professor of Genetics.

Signed by officials

Dr. W.F. Brandt.

Senior Lecturer in the Department of Biochemistry.

SUMMARY.

Histones H2B were isolated from the gonads of a mollusc (*Patella granatina*) and from chicken (*Gallus domesticus*), crocodile (*Crocodylus niloticus*) and amphibian (*Xenopus laevis*) erythrocytes. The H2B's were purified by ion-exchange and gel exclusion chromatography.

The complete primary structure of the mollusc histone $H2B_{patella}$ has been deduced from the sequences established of adjoining and overlapping peptides by the Edman degradation procedure. The partial structure of H2B from chicken erythrocytes (87 residues), crocodile erythrocytes (75 residues) and *Xenopus* erythrocytes (63 residues) was also established.

The amino acid sequences are compared to those of other histones H2B. The effect of mutations on the predicted secondary structure of histone H2B is considered.

C O N T E N T S.

Chapter 1.

Introduction.

It has long been known that the DNA of eukaryotes is associated with a complement of histone protein and a variable amount of non-histone protein to form chromatin.

Control of gene expression is likely to result in part from specific interactions of these proteins with DNA in such a way as to permit transcription of a given set of genes in one tissue while restricting their expression in other tissues. It was suggested that the histones could play a major part in gene control (Stedmann, 1950) but this theory gave way to the current theory that the main role of the histones is to maintain the higher order structure of chromatin. The precise function of the histones has not been unequivocally established.

It is likely that the histones each have a slightly different association with DNA owing to the fact that they have different primary structures and may be dissociated differentially from the DNA, as evidenced by different extraction requirements (Review - Hnilica, 1972).

Kornberg (1974) proposed that chromatin structure was a repeating unit of histones and DNA. The repeating unit consisted of an octamer of histones ($(H4\text{-}H3)_2$ and $(H2A\text{-}H2B)_2$) and approximately 200 base pairs of DNA (Hewish and Burgoyne, 1973) wrapped around the histone octamer (Noll, 1974).

The histones not accounted for in the octamer, i.e. H1 or H5, are assigned to the task of binding to the linking DNA spacers and thus drawing the nucleosomes together causing further condensation of the fibre.

Early researchers, such as J. Bonner, showed that the histones seemed to show minimal variation in different animal species and this led to the widespread acceptance of certain beliefs, viz. that the histones had a constant structure (e.g.Fambrough and Bonner, 1966) and association pattern in different tissues and that there was a universal chromatin

structure, based on the nucleosome.

Primary structure studies on histone H3 showed that there was little sequence variation in H3. From the primary structures of chicken (Brandt et al. 1972), shark (Brandt et al. 1974) and bovine H3 (Delange et al. 1973) only one difference is evident. Between chicken and pea H3 (Patthy et al. 1973) 4 differences are evident, which still only indicates a 3% difference between vertebrate and plant H3.

Similarly, H4 is an even more conservative histone (Dayhoff, 1972, 1973) while even H2A shows only 9 differences between trout (Bailey and Dixon, 1973) and mammals (Sautiere et al. 1974).

Although it is unlikely that gene expression is solely histone controlled, it is nevertheless difficult to reconcile the structural invariance of the histone sequences known up to 1974 with the functional flexibility expressed by chromatin. With certain aspects of chromatin one expects constancy, such as in replication, but when it comes to evolution and regulation it is difficult to conceive of only four constant proteins (histones) in control.

Belief that histone variants would be found led to the search and discovery of H2B histone variants (Strickland et al. 1974) which gave rise to this project which aimed to investigate H2B from a number of different animal species, widely separated on the evolutionary scale. The H2B variants discovered in 1974 (Strickland et al.) were from sea-urchins which are deemed at a lower stage of evolution.

I isolated histone H2B from an animal at a similarly low stage of evolution viz. the mollusc Patella granatina, which was sequenced (van Helden et al. 1978a). In addition, H2B histones were isolated from three vertebrates viz. the domestic chicken Gallus domesticus, the Nile crocodile Crocodylus niloticus and an amphibian Xenopus laevis sometimes called the African clawed toad. The source of histones was nucleated erythrocytes which are transcriptionally inert. The H2B's were isolated

and partially sequenced (van Helden et al. 1978b). The sequence variation of the H2B's is compared and the implications of the mutations with regard to protein secondary structure is discussed.

Chapter 2.

2.1 Purification of H2B histones.

2.1.1 Patella granatina (mollusc).

Basic proteins were extracted from mature or immature male gonads or sperm of Patella granatina (mollusc) as described (4.2.2). The yield was approximately 2 g protein per 80 g mature gonads (wet weight) and 0,45 g per 20 g sperm. Gel electrophoresis of these extracted proteins is given (Fig. 2.1).

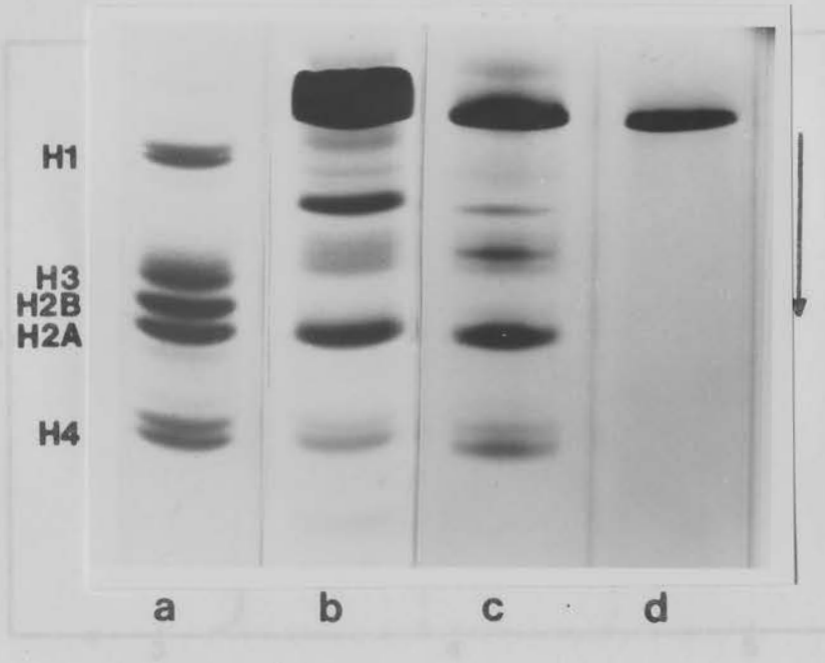


Fig. 2.1.

Gel electrophoresis of acid extractable proteins from chromatin of (a) calf thymus, (b) mature mollusc gonads, (c) immature mollusc, (d) motile sperm.

In comparison to total calf thymus histones, there is an additional major protein present in the mature gonads of mollusc. This additional protein is the only major protein that was extracted from sperm cells. If proteins from immature gonads are acid extracted, the quantity of this protein is drastically decreased. In order to increase the relative histone quantity, subsequent studies were done using immature gonads in

order to reduce the spermatid content of the tissue to be extracted. From mature gonads approximately 80% of the total acid extractable material consists of this protein and from resting phase (immature) gonads, approximately 50%.

The purification of the mollusc histone H2B from total gonad acid-extractable material was achieved in two stages, viz. ion-exchange chromatography followed by molecular sieve chromatography (4.2.2.2).

Initial experiments were performed using Bio-Gel P60 as a weak ion-exchanger with the following results (Fig. 2.2).

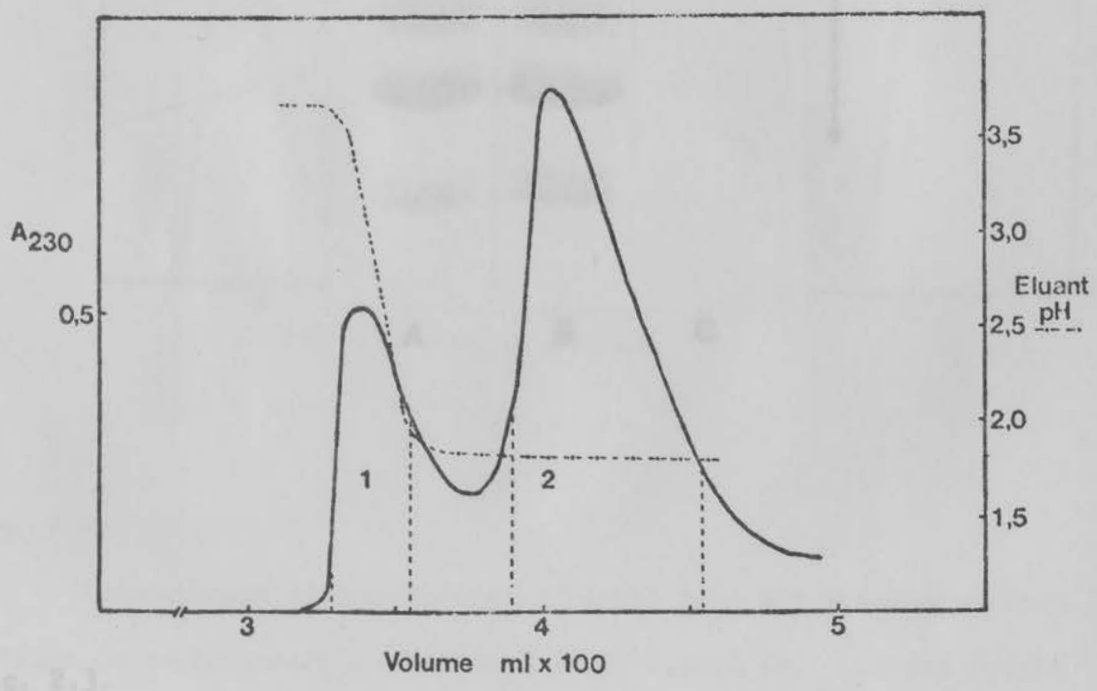


Fig. 2.2.

Ion-exchange chromatography of total acid extractable protein from mollusc gonads on Bio-Gel P60, 2,5x30 cm. Sample weight: 64 mg; Initial column pH: 4,5; Linear gradient: pH 3,20 to pH 1,70; Volume: 1000 ml; Flow rate: 19 ml/hr. Fraction 1: (12 mg) histones; fraction 2: sperm protein only (see Fig. 2.3 and Table 2.1).

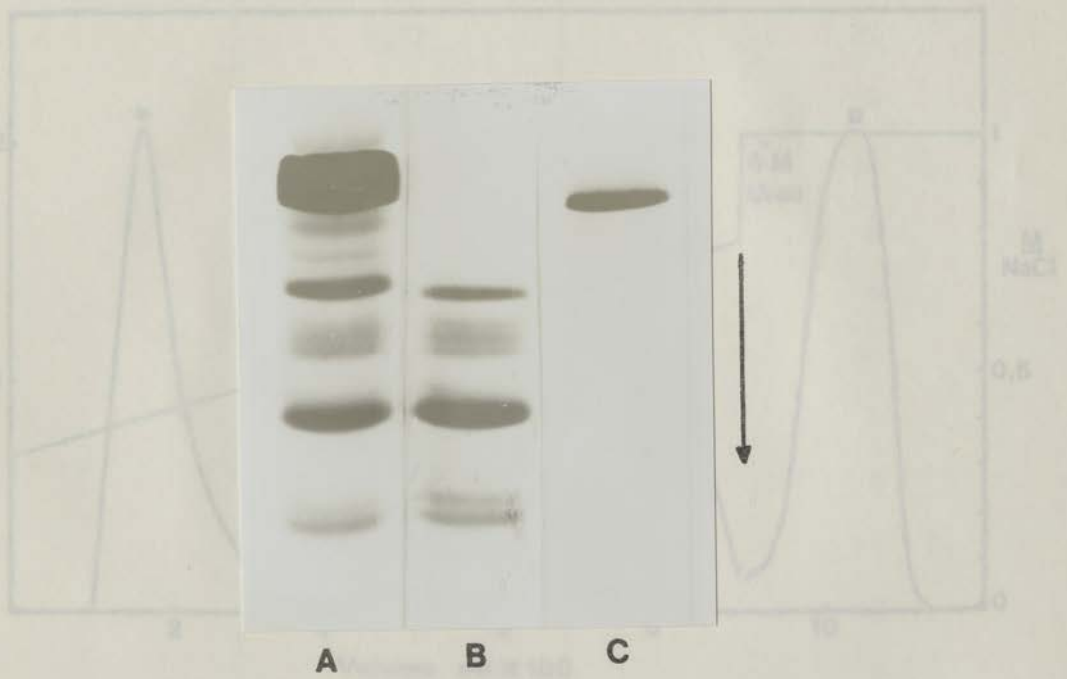


Fig. 2.3.

Gel electrophoresis of fractions eluted from Bio-Gel P60 (Fig. 2.2). A = Total extract from mature gonads; B = Fraction 1; C = Fraction 2.

Since the ion-exchange properties of Bio-Gel P60 were variable, further ion-exchange separation was done with carboxymethylcellulose (Whatman CM-52)(Fig. 2.4).

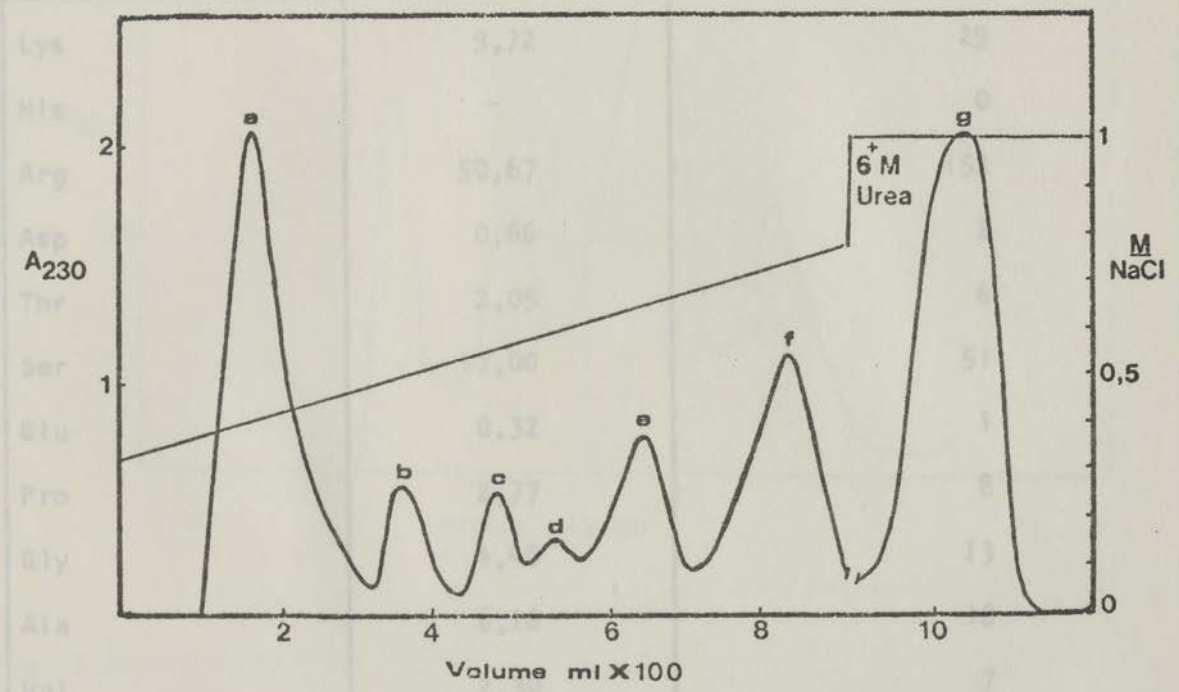


Fig. 2.4.

Ion-exchange chromatography of total acid extractable protein from mollusc on carboxymethylcellulose (CM-52) 2,5x30 cm. Sample weight: 300 mg; Buffer: 0,05 M sodium acetate/HCl pH 4,50; Linear gradient: sodium chloride 0,2-0,8 M; Volume: 900 ml. The final fraction was eluted by 0,05 M sodium acetate/HCl, 1 M NaCl, 6 M urea, pH 4,50.

Amino acid composition of the sperm-protein showed it to be a large molecule (Table 2.1) with 60,4 mole % basic amino acids.

Table 2.1.

Amino-acid composition of Patella granatina basic protein.

Amino acid	Mole %	Likely no. of residues
Lys	9,72	29
His	-	0
Arg	50,67	152
Asp	0,66	2
Thr	2,05	6
Ser	17,00	51
Glu	0,32	1
Pro	2,77	8
Gly	4,40	13
Ala	6,16	18
Val	2,30	7
Met	-	0
Ile	0,65	2
Leu	2,70	7
Tyr	0,35	1
Phe	0,34	1
Residues/mole		298

Since each of the peaks b-f contained H2B and none of these was pure, further experiments were done with stepwise elution. The sample was applied to a carboxymethylcellulose column in 0,05 M sodium acetate/HCl pH 4,50 and the histone fractions eluted with the same buffer containing 0,8 M NaCl. The final fraction was eluted as before.

The histone fraction from ion-exchange was dialysed and freeze-dried, subsequent to which it was dissolved in urea and applied to a Bio-Gel P30 column (Fig. 2.5).

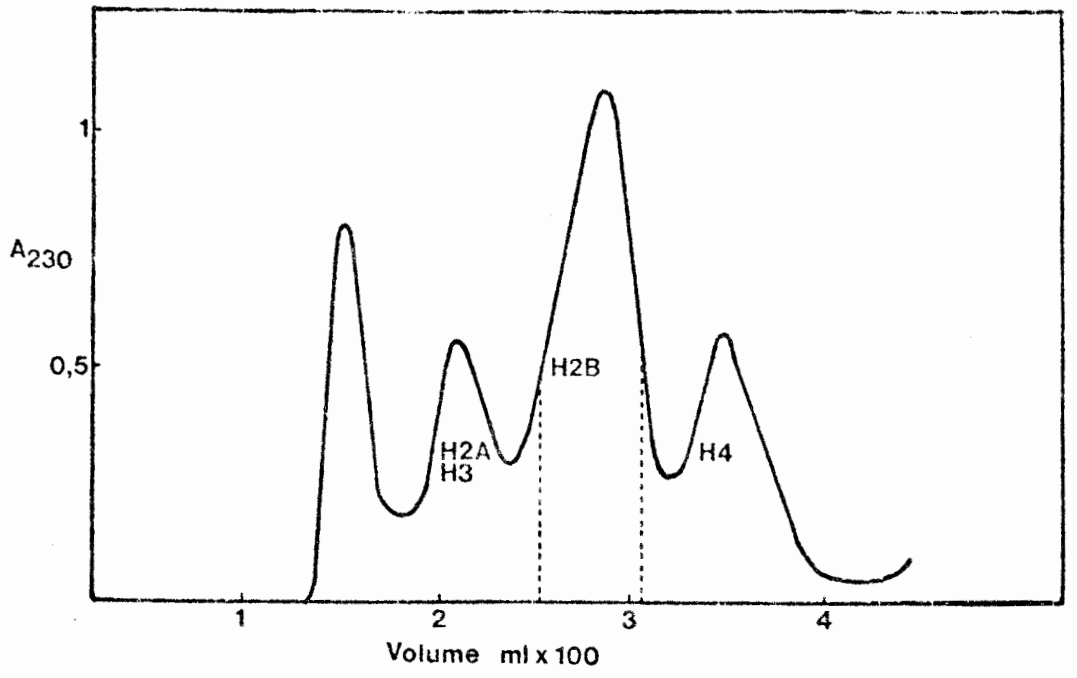


Fig. 2.5.

Gel exclusion chromatography of histone fraction from ion-exchange chromatography on Bio-Gel P30 (2,5x90 cm). Sample weight: 80 mg; Eluant: 0,02 N HCl 0,05 M NaCl; Flow rate: 33 ml/hr. Fraction 3 contains impure H2B.

Four fractions were eluted. N-terminal analysis of all fractions demonstrated the presence of H2B (based on proline as an N-terminus) in fraction 3 only. The presence of N-terminal alanine in fraction 3 indicated that this fraction also contained H3. In addition, electrophoresis of this fraction showed a slow moving band which disappeared after reduction of disulfide bonds, further demonstrating the presence of H3 (Fig. 2.6).

Fig. 2.6.

Mollusc whole histone (A) on gel electrophoresis and zones from fraction 3 of Fig. 2.5 (B).

The H2B was finally purified by rechromatography twice more through the same Bio-Gel P30 (Fig. 2.7) after which it was 96-98% pure by the criteria of gel electrophoresis (Fig. 2.8) and end-group analysis.

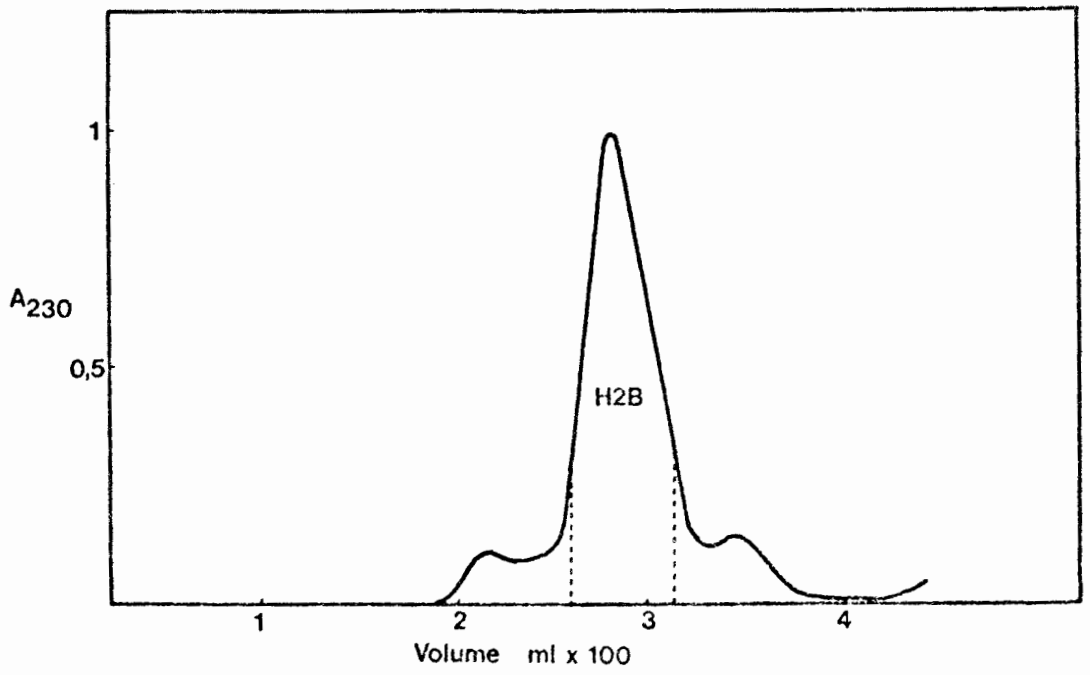


Fig. 2.7.

The second gel exclusion chromatographic separation of fraction 3 (containing impure H2B) from Fig. 2.5. Experimental conditions as in Fig. 2.5.

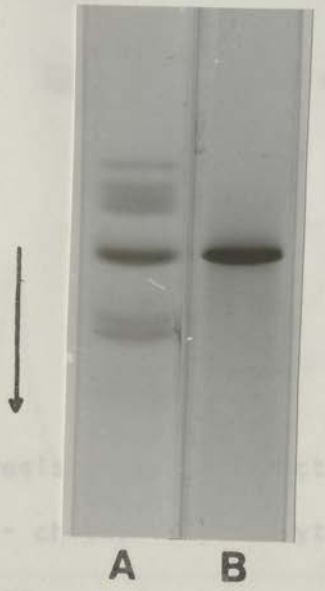


Fig. 2.8.

Gel electrophoresis of total mollusc gonad histones (A) and pure mollusc gonad H2B (B).

2.1.2. Gallus domesticus (chicken), Crocodylus niloticus (crocodile) and Xenopus laevis (amphibian).

Basic proteins were extracted from the erythrocytes of these animals. For comparison, gel electrophoresis and gel chromatography of total histone extracts were done (Figs. 2.9, 2.10, 2.11 and 2.12).

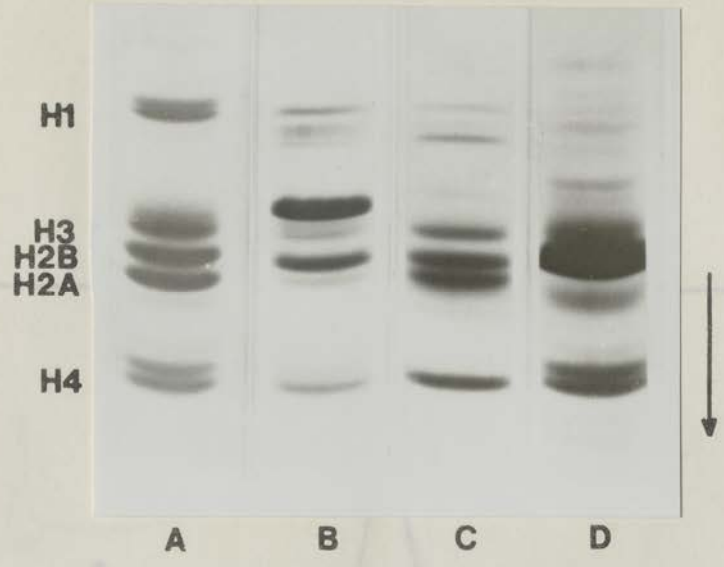


Fig. 2.9.

Gel electrophoresis of acid extractable proteins from chromatin of A - calf thymus, B - chicken erythrocytes, C - crocodile erythrocytes, D - Xenopus erythrocytes.

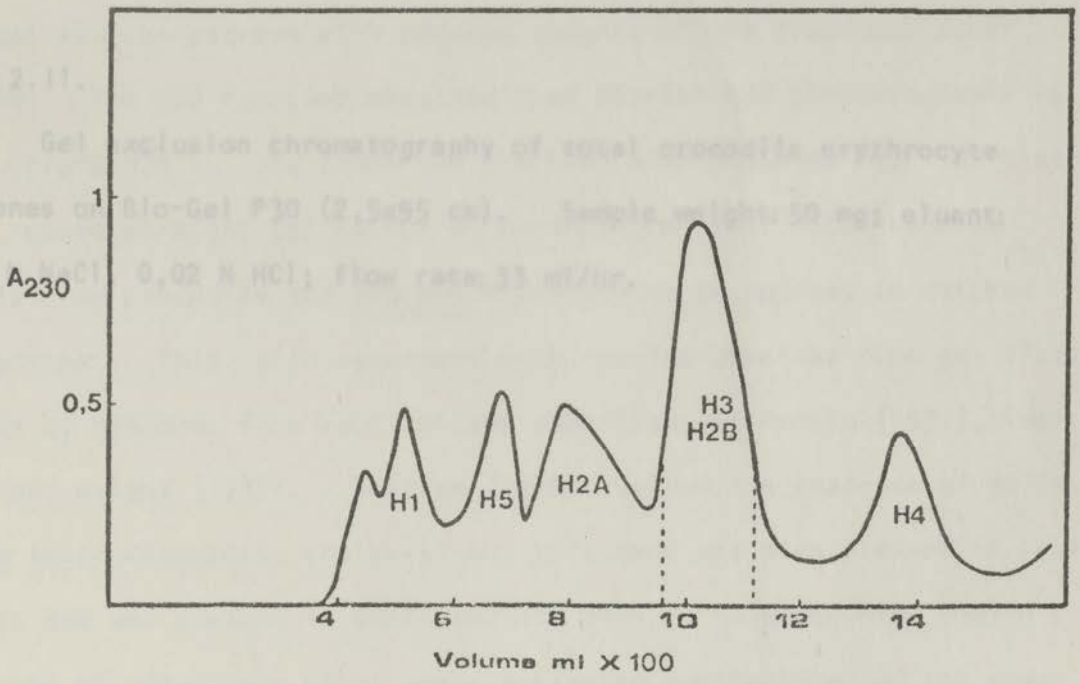


Fig. 2.10.

Gel exclusion chromatography of total chicken erythrocyte histones on Bio-Gel P30 (5x80 cm). Sample weight: 130 mg; eluant: 0,05 M NaCl, 0,02 N HCl; flow rate: 70 ml/hr.

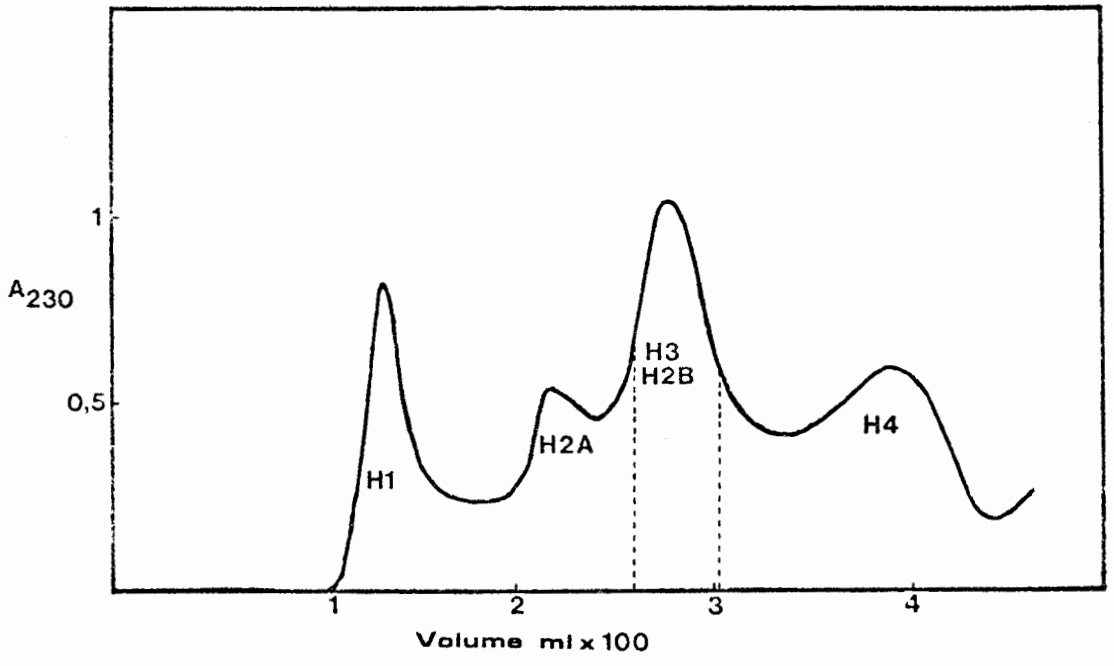


Fig. 2.11.

Gel exclusion chromatography of total crocodile erythrocyte histones on Bio-Gel P30 (2,5x95 cm). Sample weight: 50 mg; eluant: 0,05 M NaCl, 0,02 N HCl; flow rate: 33 ml/hr.

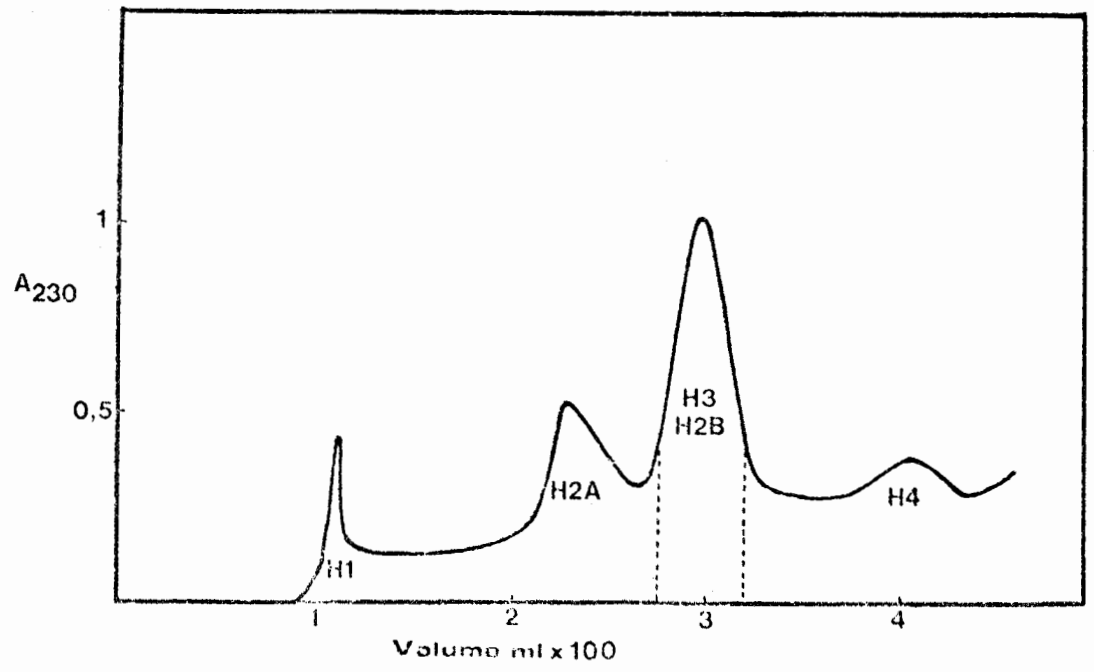


Fig. 2.12.

Gel exclusion chromatography of total Xenopus erythrocyte histones on Bio-Gel P30. Conditions as in Fig. 2.11.

A protein fraction enriched in H2B obtained as described in 4.2.3.2 and chromatographed on Bio-Gel P30 as described gave a qualitatively identical elution pattern with reduced amounts of the fractions other than H2B. The H2B fraction obtained from Bio-Gel P30 chromatography is termed 'crude H2B'. The crude H2B fraction also contains H3. No histone H5 was recovered after gel exclusion chromatography of total histone extracts from crocodile and Xenopus erythrocytes as opposed to chicken erythrocytes. This is in agreement with results obtained from gel electrophoresis of histones from reptiles and amphibians by Panyim (1971), Tsai (1975) and Wright (1975). Destree (1972) claimed the presence of H5 in Xenopus (electrophoretic analysis) but this band was also present in liver extracts and was present in small amounts only. To quote Miki (1977) ... "The scope of occurrence of H5 among nucleated erythrocytes of the lower vertebrates is still uncertain ...". Dimerization of H3 (Brandt, 1971) by the formation of disulfide bonds increased the molecular size after which separation of H2B and H3 dimer was achieved by chromatography on Sephadex G100 (Fig. 2.13).

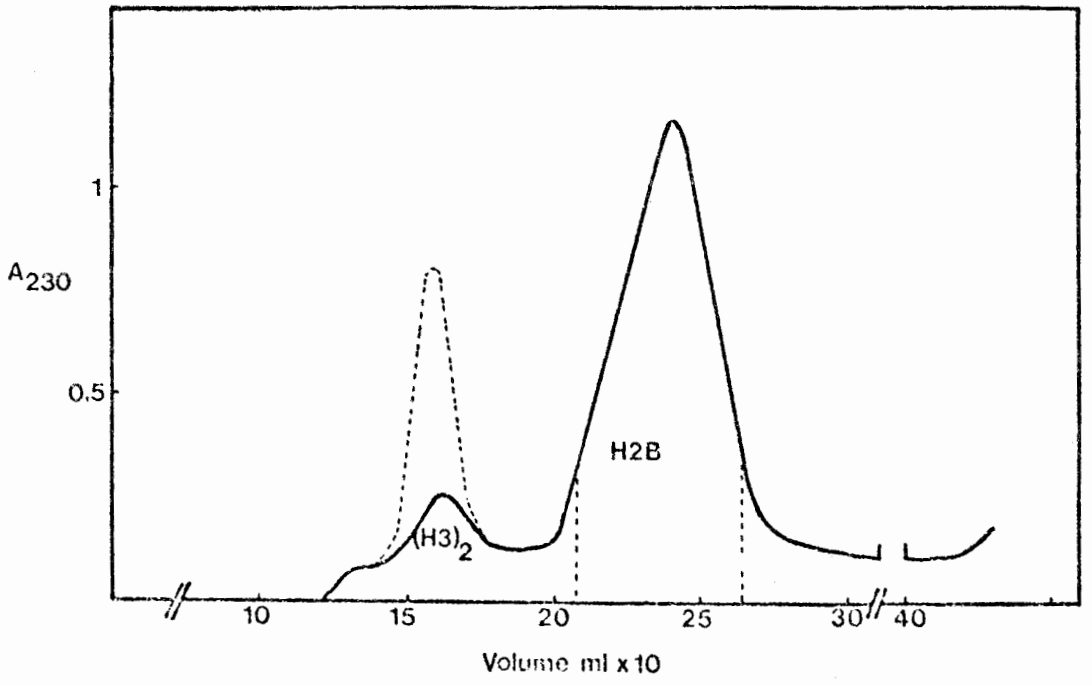


Fig. 2.13.

Gel exclusion chromatography of Crude H2B from chicken, crocodile or Xenopus erythrocytes after iodosobenzoate treatment (4.2.3.3) on Sephadex G100 (2,5x95 cm). Sample weight: 30 mg; eluant: 0,01 N HCl; flow rate: 30 ml/hr.

The dimerization of H3 does not go to completion and to obtain 97% pure H2B this process was repeated two to four times. The dotted line in Fig. 2.13 indicates the variable amount of H3 dimer.

The stages of chicken erythrocyte H2B purification are shown below. The corresponding stages of crocodile and Xenopus purification were identical (not shown).

Fig. 2.14.

Histone H2B purification from chicken erythrocytes. Gel 1; (1% 2-mercaptoethanol) H2B after Biogel P30 chromatography (deliberately overloaded to make H3 contamination visible), Gel 2; same as (1) but not overloaded (no mercaptoethanol; H3 dimer visible), Gel 3; H2B after further purification by dimerization of H3 and gel chromatographic separation of H3 dimer (no mercaptoethanol). End-group determination by manual Edman degradation (4.4.4.2) and quantitation by gas chromatographic identification (4.5.3): Gels 1 and 2, 15% H3: Gel 3, 2,5% H3.

Two H2B histones from sea-urchin (Parenchinus angulosus - Strickland et al. 1977) and histones H2B from calf thymus, chicken, crocodile and Xenopus erythrocytes and also from mollusc gonads were electrophoresed on polyacrylamide gels as described (4.4.2). Ten micrograms of each histone were electrophoresed alone and in combination with calf thymus H2B (Fig. 2.15).

Fig. 2.15.

Gel electrophoresis of H2B histones. A - calf thymus H2B, B - calf thymus + chicken erythrocyte H2B, C - calf thymus + crocodile erythrocyte H2B, D - calf thymus + Xenopus erythrocyte H2B, E - calf thymus + mollusc gonad H2B. 10 µg of calf thymus H2B as a standard in each gel plus 10 µg of each H2B to be tested.

Where calf thymus H2B was mixed with sea-urchin and mollusc H2B, two bands were seen, but where calf was mixed with chicken, crocodile and Xenopus H2B's only one band was visible (Fig. 2.15, Table 2.2). Gel scans were done at 615 nm on a Vitatron densitometer and all calculations were based on measurements from these scans, e.g. Figs. 2.16, 2.17 and 2.18. All of the densitometer scans were obtained from gels made with the same batch of reagents. These gels were electrophoresed, stained and destained in a single experiment under identical conditions.

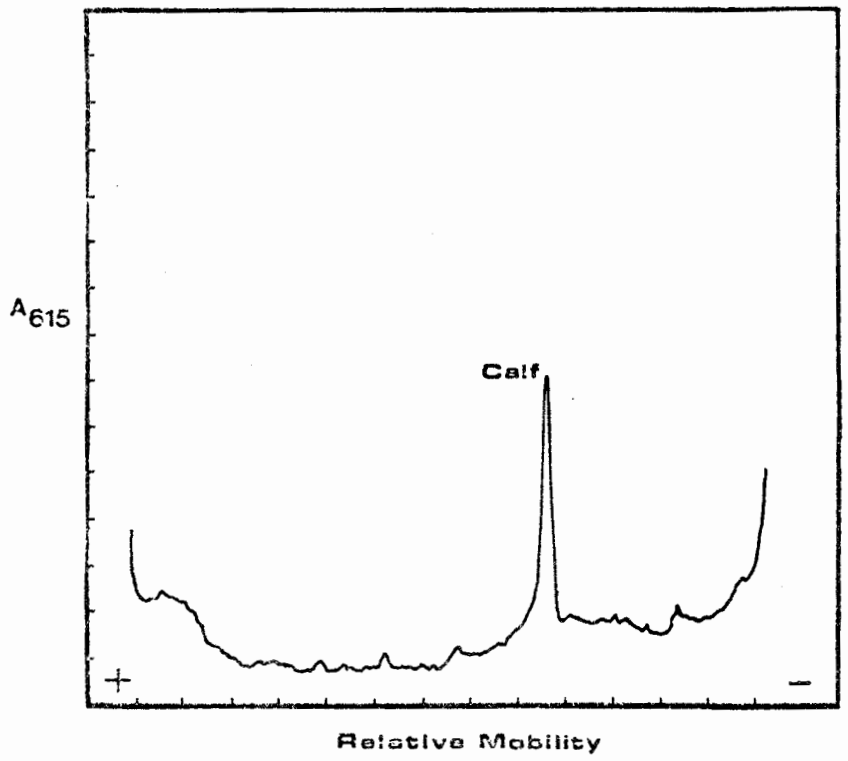


Fig. 2.16.

Densitometer scan of calf thymus H2B from gel electrophoresis.

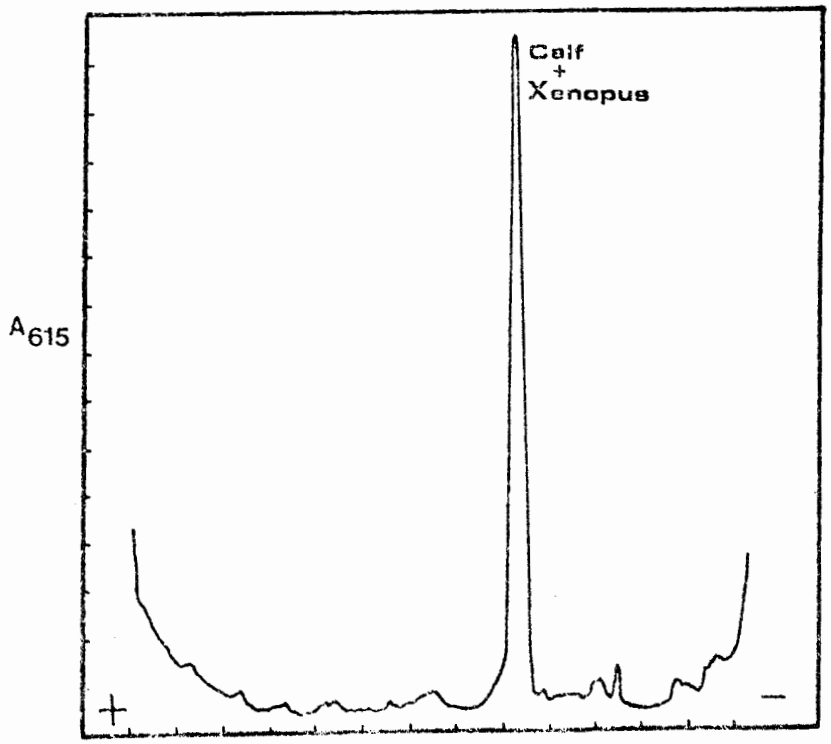


Fig. 2.17.

Densitometer scan of calf thymus + Xenopus erythrocyte H2B from gel electrophoresis.

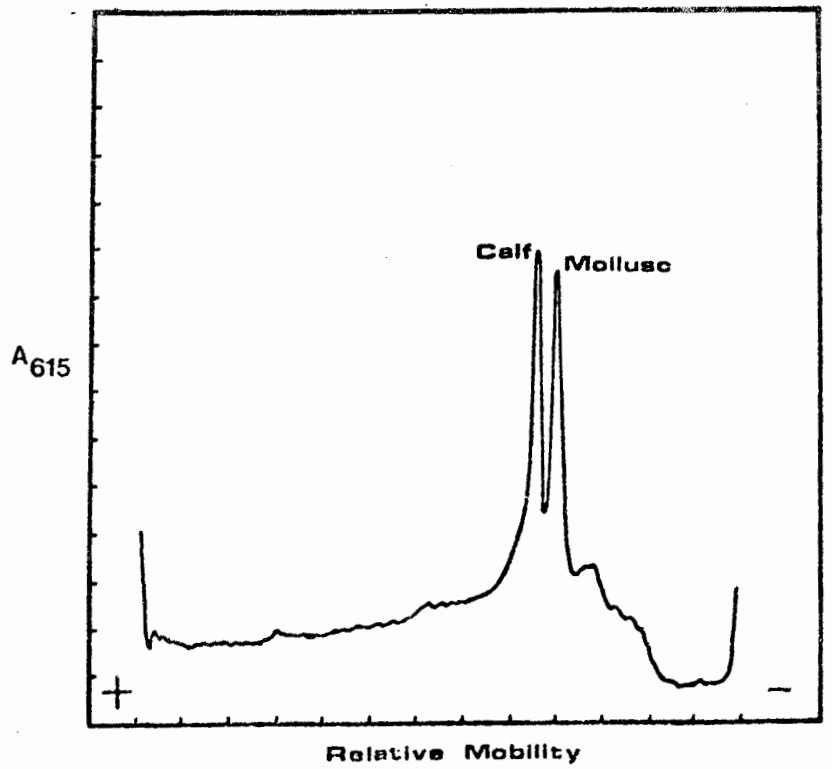


Fig. 2.18.

Densitometer scan of calf thymus + mollusc H2B from gel electrophoresis.

Table 2.2.

Source	Migration distance (cm) of H2B from top of gel measured from scan (+ terminal)
Calf thymus	8,3
Chicken erythrocyte	8,3
Crocodile erythrocyte	8,3
Xenopus erythrocyte	8,3
Mollusc gonad	8,8
Sea-urchin $H2B_1$ (sperm)	7,8
Sea-urchin $H2B_2$ (sperm)	8,1

The increased mobility of mollusc H2B is in accordance with its amino acid composition (Table 3.1.).

2.2 Purification of peptides.

2.2.1 Patella granatina.

2.2.1.1 Cyanogen bromide peptides.

The amino acid composition of H2B (Table 2.4.) shows the presence of two methionine residues. Therefore cleavage by cyanogen bromide should yield three peptides. The freeze-dried peptides from cyanogen bromide cleavage were chromatographed on a Bio-Gel P30 column (Fig. 2.19).

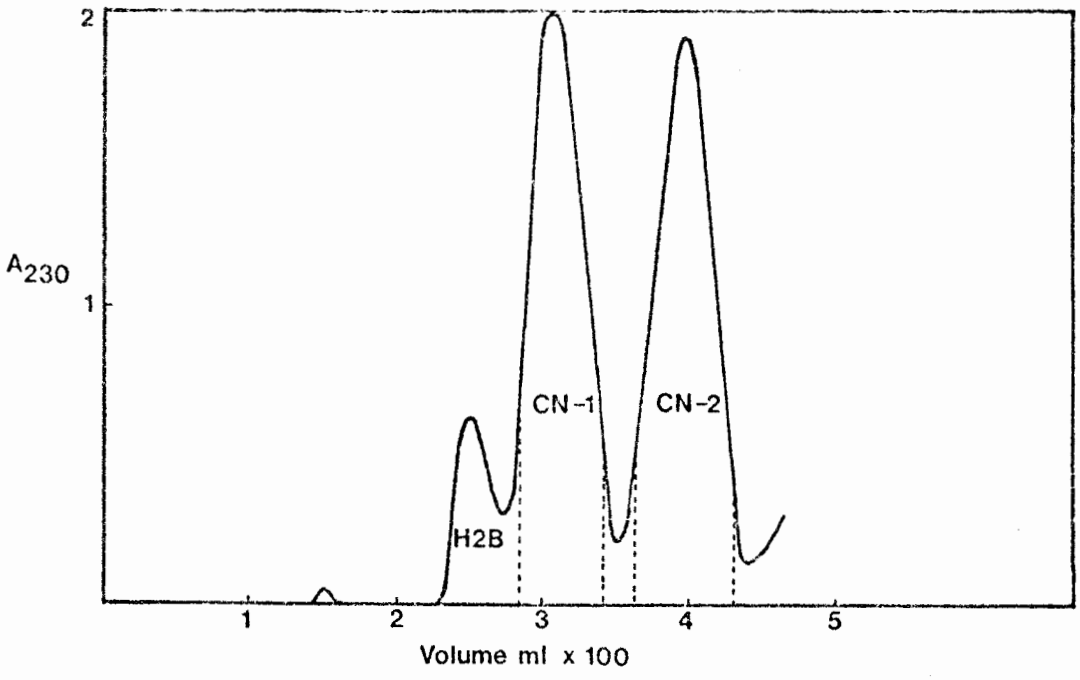


Fig. 2.19.

Gel exclusion chromatography of cyanogen bromide cleaved H2B on Bio-Gel P30 (2,5x90 cm). Sample weight: 50 mg; eluant: 0,05 M NaCl 0,02 N HCl; flow rate: 33 ml/hr.

The first minor fraction (V_o) eluted at the outer volume of the column, the next fraction was uncleaved H2B and the following fractions were peptides CN-1 and CN-2. The last fraction was urea. Three peptides were expected, but only two were recovered.

It was possible that the third peptide could have been very small, in which case it might have eluted together with the urea fraction (Fig. 2.19). The large volume of the urea fraction precluded easy manipulation. Therefore, the small peptide (CN-3) was recovered by chromatography of cleaved H2B on a Sephadex G15 column (Fig. 2.20).

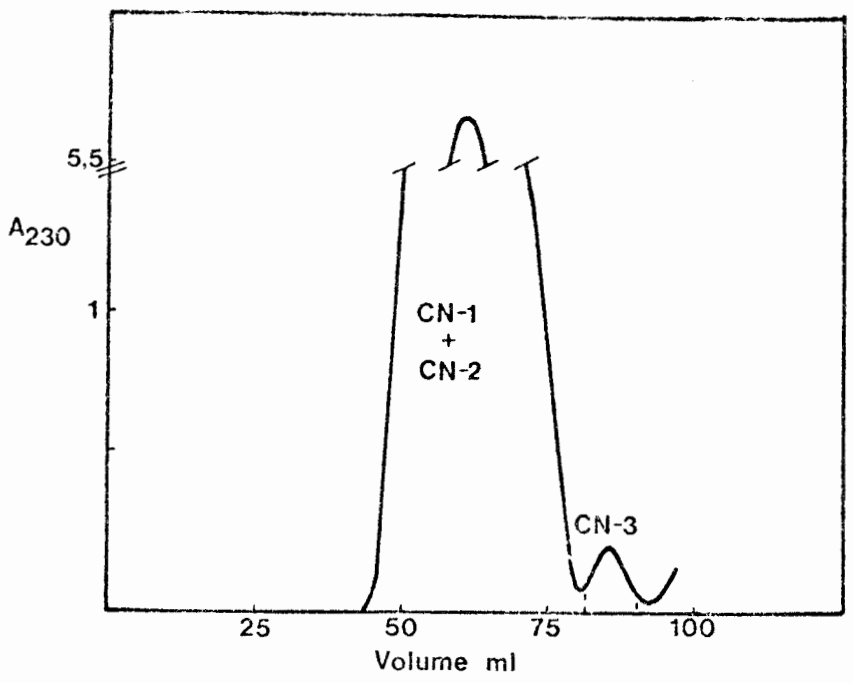


Fig. 2.20.

Gel exclusion chromatography of cyanogen bromide cleaved H2B on Sephadex G15 (1,5x90 cm). Eluant: 0,01 N HCl, flow rate: 20 ml/hr.

2.2.1.2 N-bromosuccinimide peptides.

From the amino acid analysis of CN-2 (Table 2.4.) two tyrosine residues were noted. N-bromosuccinimide (NBS) can be used to cleave tyrosyl-peptide bonds selectively.

1,66 micromoles of peptide CN-2 were cleaved by NBS (4.4.1.2). The experiment was monitored spectrophotometrically by measuring the absorption at 260 nm (Fig. 2.21). The end point of the reaction as calculated for a consumption of 3 moles of NBS per mole of tyrosine or histidine was expected to be 44,3 µl. The observed end point is 42,5 µl. The freeze-dried solution was dissolved in 8 M urea 0,01 N HCl and chromatographed on Sephadex G50 (fine). Two fractions (CN-2 NBS-1 and CN-2 NBS-2) were eluted and freeze-dried (Fig. 2.22).

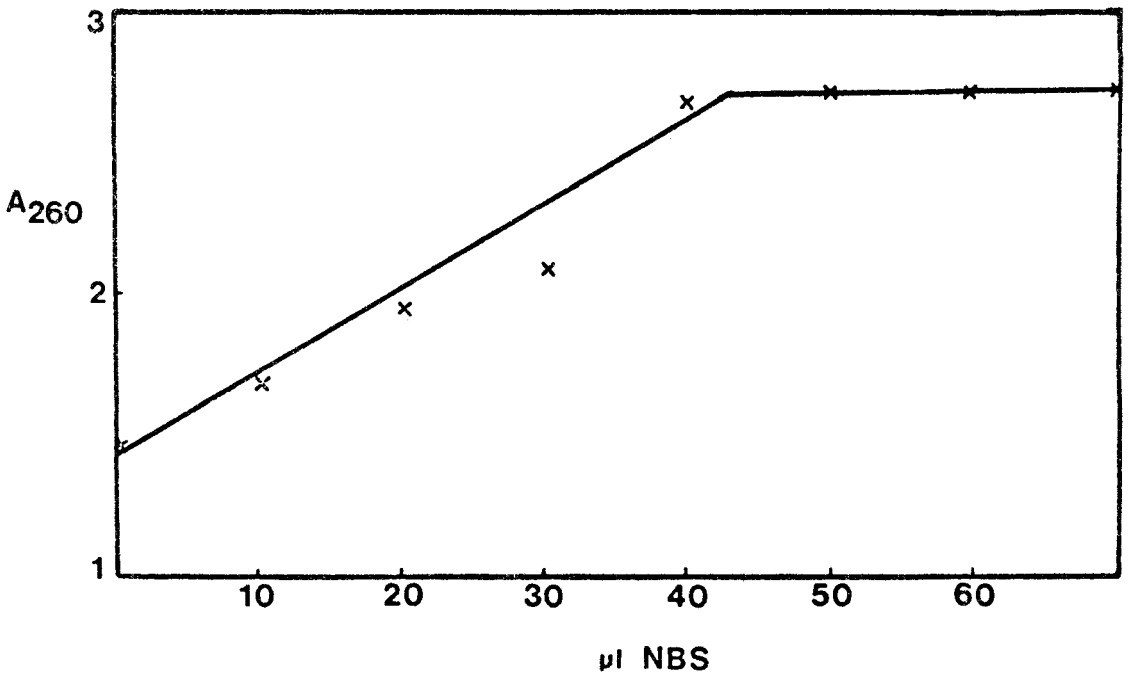


Fig. 2.21.

Titration of mollusc CN-2 by NBS (0,089 M). Optical path-length: 1,0 cm.

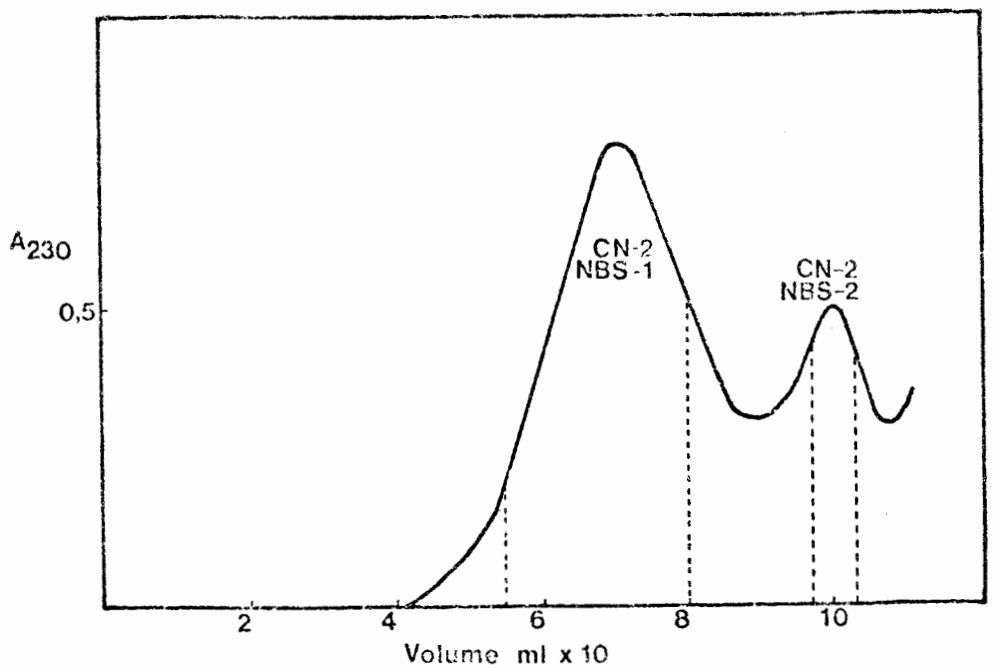


Fig. 2.22.

Gel exclusion chromatography of N-bromosuccinimide cleaved mollusc CN-2 on Sephadex G50 (fine) (1,6x95 cm). Eluant: 0,01 N HCl, flow rate: 25 ml/hr.

Three peptides were expected. In homology to calf thymus H2B it was expected that the third peptide was small and eluted in the urea fraction. This was confirmed when a peptide of 4 amino acids was recovered from NBS cleavage of CN-2 MT-1.

CN-2 MT-1 was dissolved in 50% (v/v) acetic acid and NBS added as described in 4.4.1.2 (Fig. 2.23).

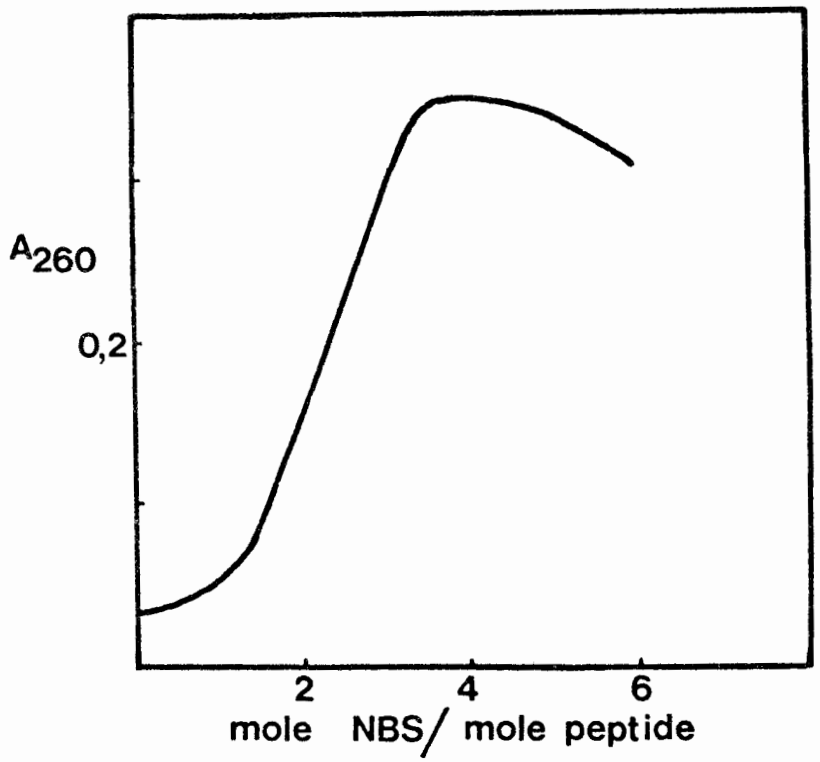


Fig. 2.23.

Titration of mollusc CN-2 MT-1 by NBS (0,089 M). Optical pathlength: 0,2 cm.

The peptide mixture was applied directly to a Sephadex G15 column, without the addition of urea (Fig. 2.24).

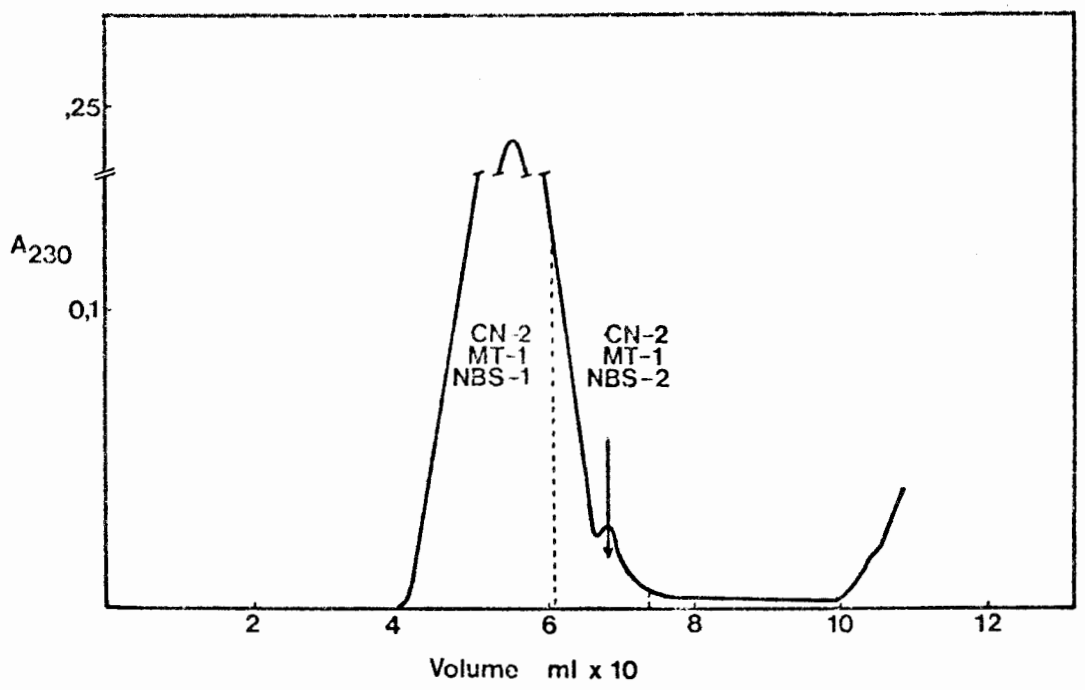


Fig. 2.24.

Gel exclusion chromatography of NBS cleaved mollusc CN-2 MT-1 on Sephadex G15 (1,5x95 cm). Eluant: 0,01 N HCl, flow rate: 30 ml/hr.

Separation of the peptides was poor, so the fraction CN-2 MT-1 NBS-2 (Fig. 2.24) was redissolved in 5% acetic acid 8 M urea and applied to a Sephadex G50 column (Fig. 2.25).

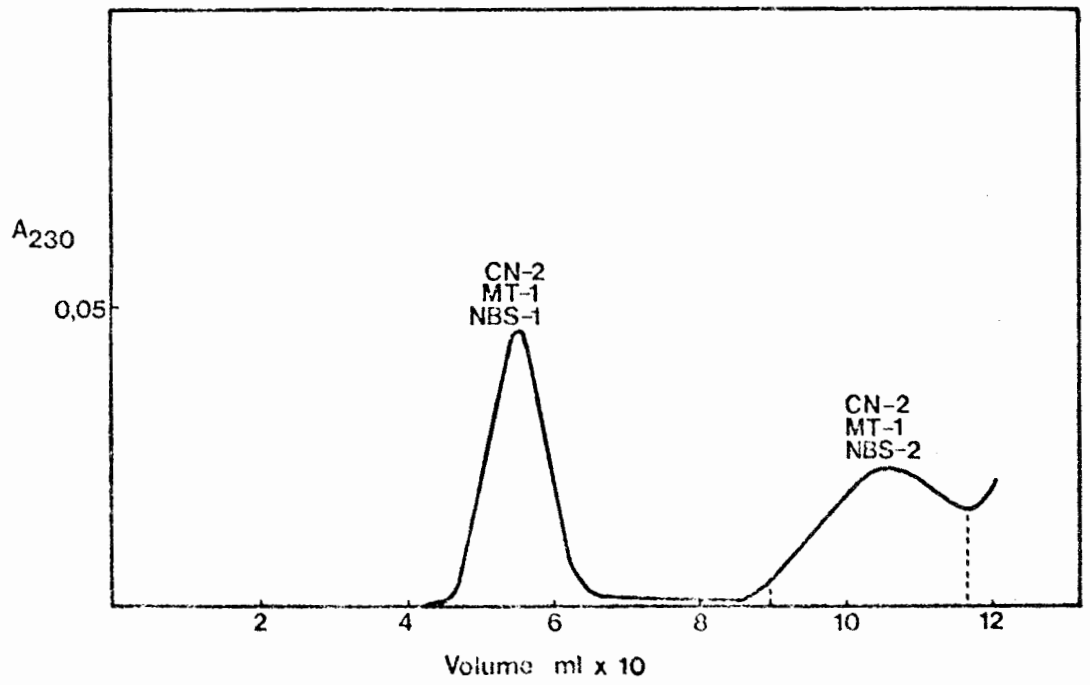


Fig. 2.25.

Gel exclusion chromatography of CN-2 MT-1 NBS-2 from Fig. 2.24 rechromatographed on Sephadex G50 (fine) (1,6x95 cm). Eluant: 0,01 N HCl, flow rate: 24 ml/hr. CN-2 MT-1 NBS-2 was recovered.

2.2.1.3 Thermolysin peptides.

The cleavage procedure has been described (4.3.3.3). The peptide mixture was chromatographed on Sephadex G15 (Fig. 2.26). Two fractions were eluted. Because only the larger peptides were of interest only the first was pooled, freeze-dried and subjected to amino acid analysis (Table 2.4.).

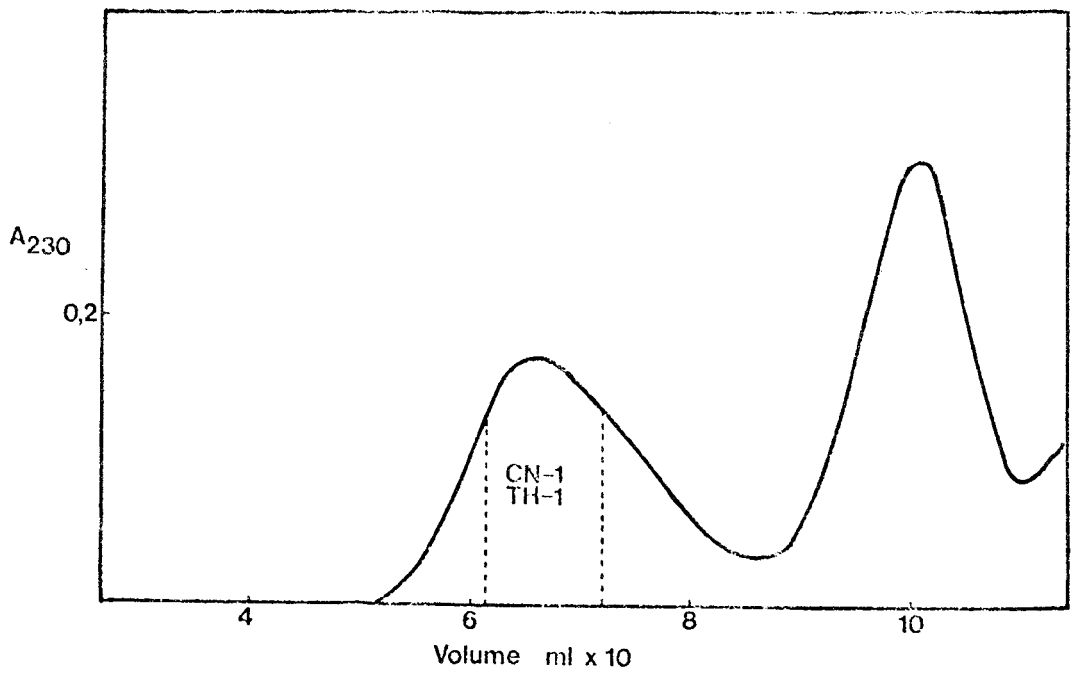


Fig. 2.26.

Gel exclusion chromatography of thermolysin cleaved mollusc CN-1 on Sephadex G15 (1,5x95 cm). Eluant: 0,01 N HCl, flow rate: 30 ml/hr.

2.2.1.4 Staphylococcus aureus protease peptides.

The cleavage method was described in 4.3.3.2. The peptide mixture was chromatographed on a Sephadex G50 column and three fractions eluted (Fig. 2.27).

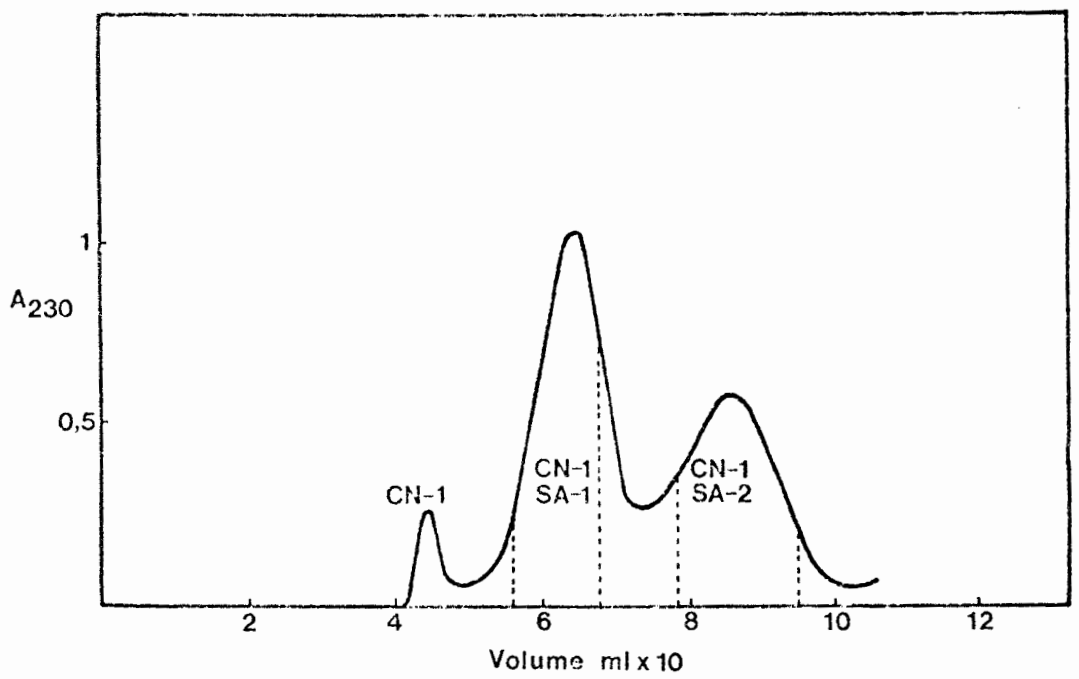


Fig. 2.27.

Gel exclusion chromatography of Staphylococcus aureus protease cleaved mollusc CN-1 on Sephadex G50 (fine)(1,5x90 cm). Sample weight: 9 mg; eluant: 0,01 N HCl, flow rate: 20 ml/hr.

Fraction CN-1 SA-2 was further purified by ion-exchange chromatography (Fig. 2.28). The quantity of peptides recovered in fractions I, II and III was too low for further analysis.

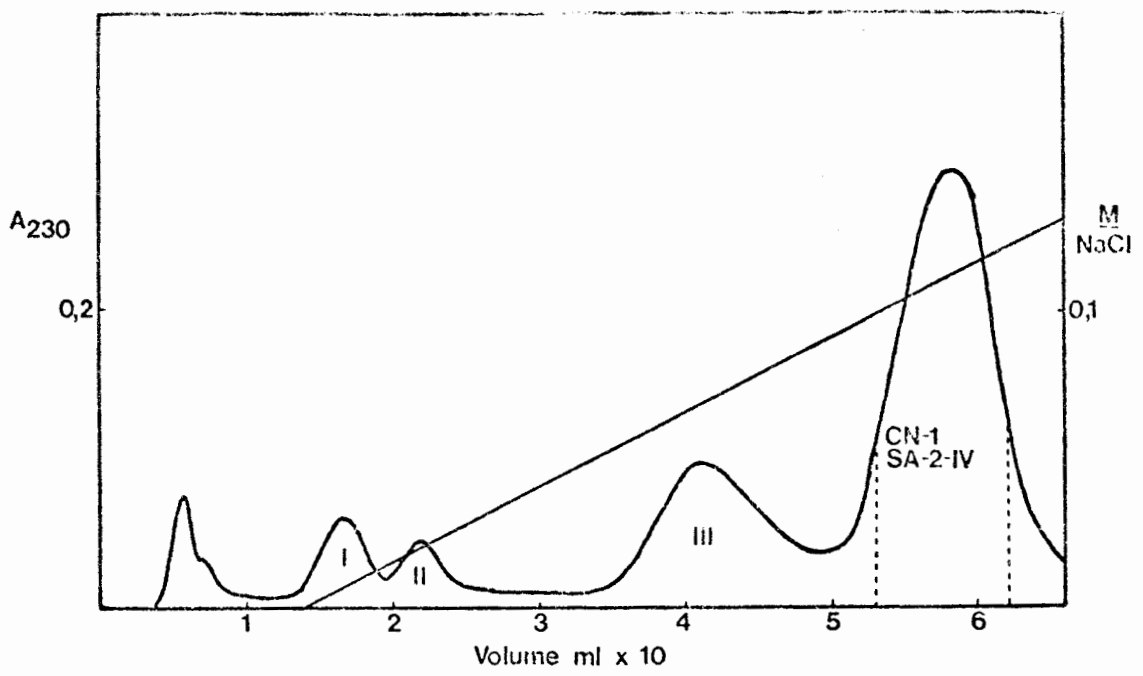


Fig. 2.28.

Ion-exchange chromatography of mollusc CN-1 SA-2 (Fig. 2.27) on carboxymethylcellulose (CM-52)(6x110mm). Sample weight approx. 4 mg, buffer: 0,05 M sodium acetate/HCl pH 4,50, linear gradient: sodium chloride 0-0,2 M, volume: 100 ml, flow rate: 4 ml/hr.

2.2.1.5 Trypsin peptides.

After digestion of CN-1 SA-2-IV, the cleavage mixture was applied to a Sephadex G15 column, from which two peptide fractions were recovered, plus a salt-containing fraction i.e. T-3 (Fig. 2.29).

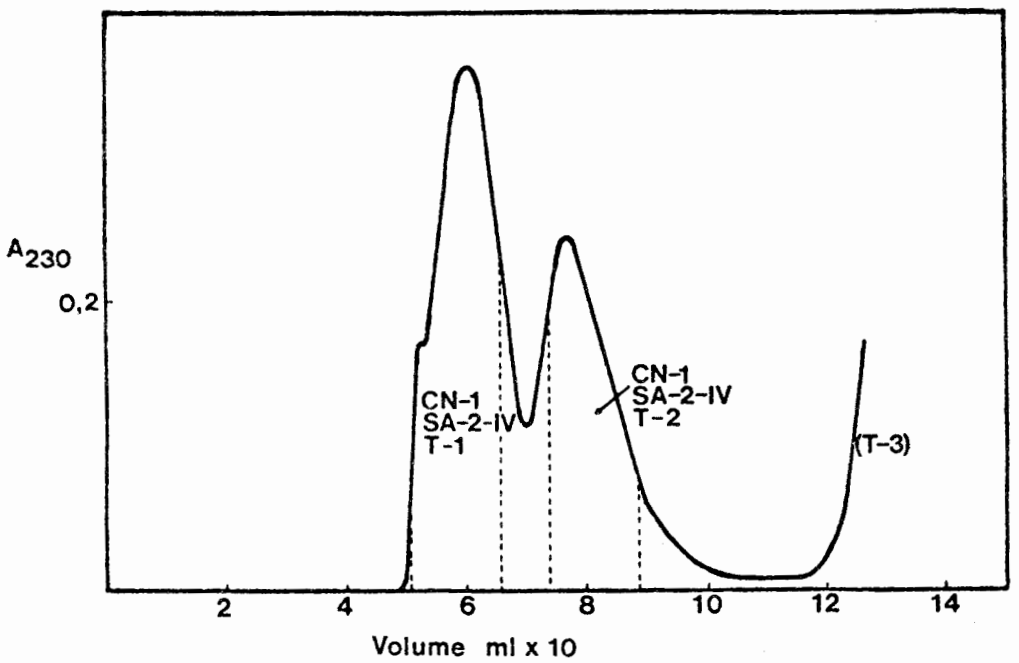
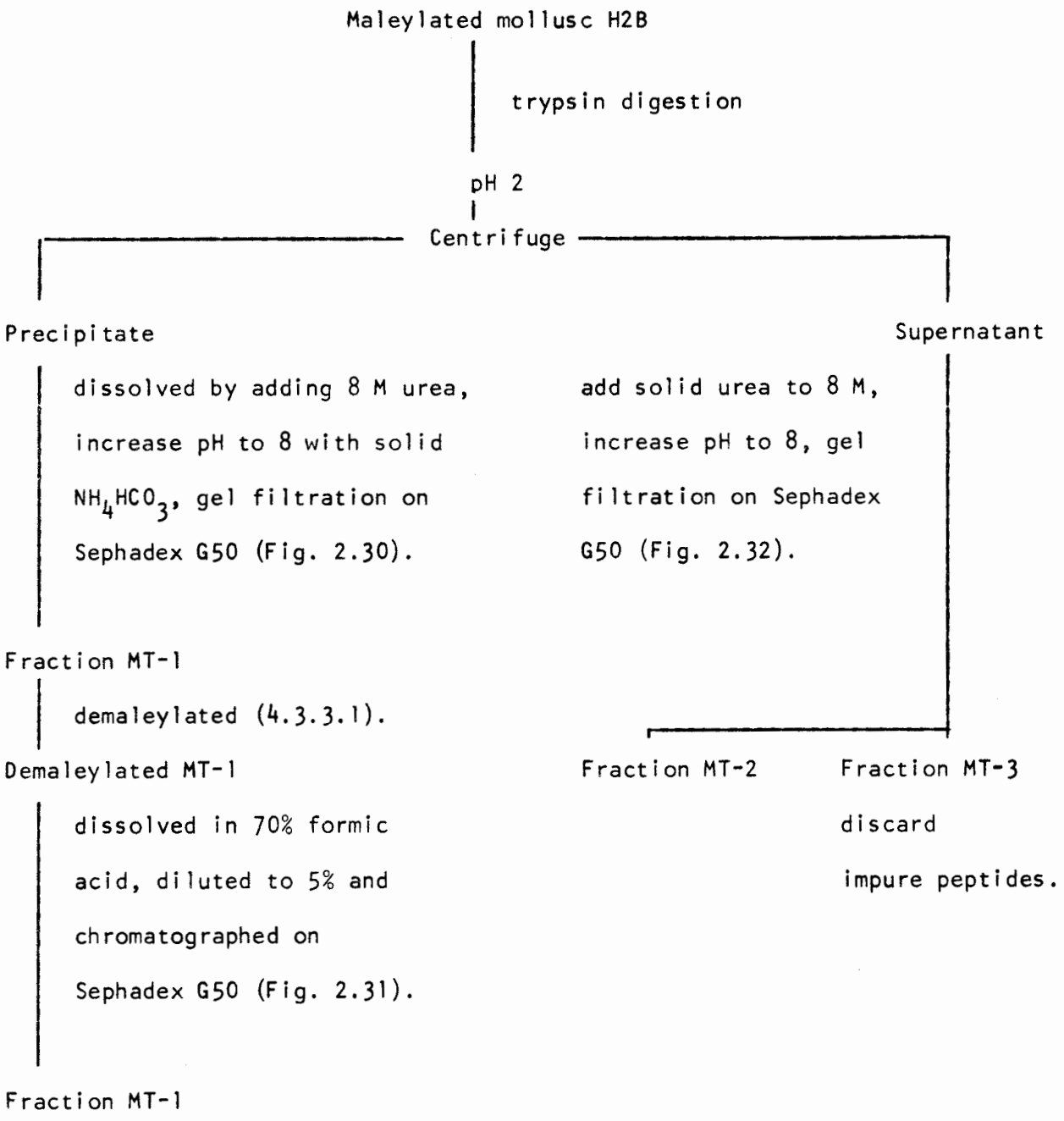


Fig. 2.29.

Gel exclusion chromatography of trypsin cleaved CN-1 SA-2-IV on Sephadex G15 (1,5x95 cm). Eluant: 0,01 N HCl, flow rate: 30 ml/hr.

In addition to the above, whole mollusc H2B was maleylated and then digested with trypsin (4.3.3.1). The recovery of peptides is illustrated schematically.


Maleylated mollusc H2B
trypsin digestion
pH 2
Centrifuge
Precipitate
dissolved by adding 8 M urea, increase pH to 8 with solid NH_4HCO_3, gel filtration on Sephadex G50 (Fig. 2.30).
Fraction MT-1
demaleylated (4.3.3.1).
Demaleylated MT-1
dissolved in 70% formic acid, diluted to 5% and chromatographed on Sephadex G50 (Fig. 2.31).
Fraction MT-1
Supernatant
add solid urea to 8 M, increase pH to 8, gel filtration on Sephadex G50 (Fig. 2.32).
Fraction MT-2
Fraction MT-3
discard impure peptides.

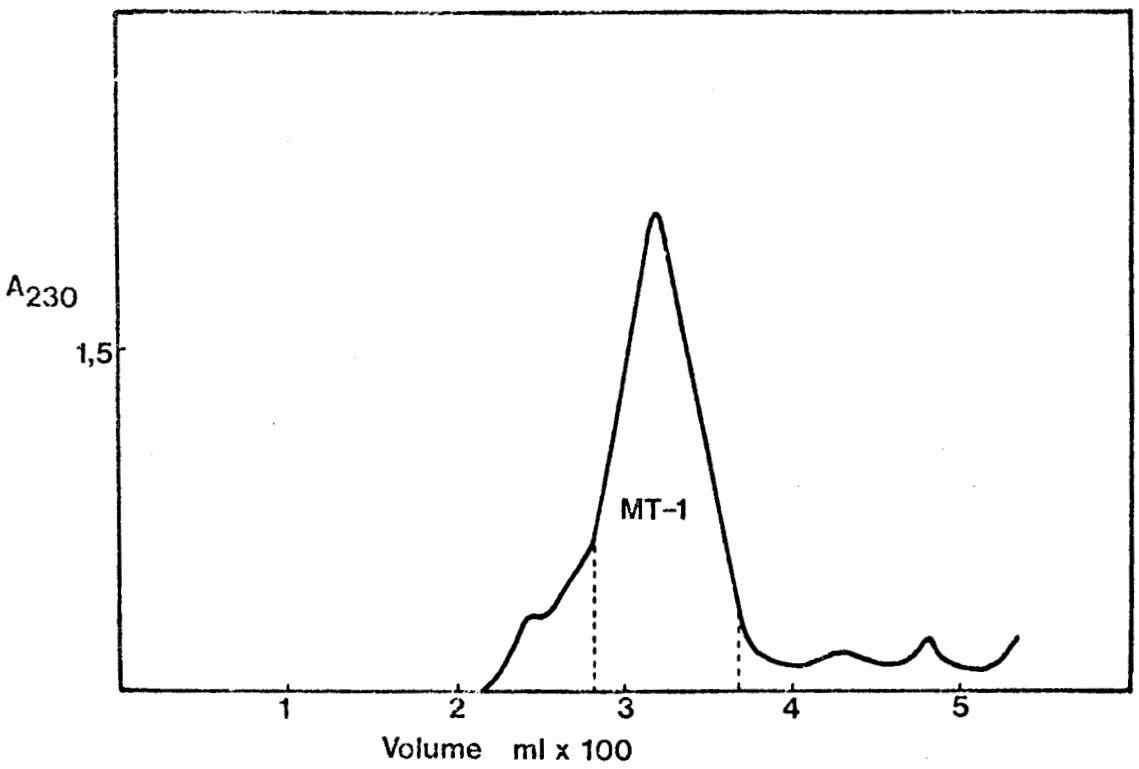


Fig. 2.30.

Gel exclusion chromatography of precipitate from trypsin cleavage of mollusc H2B on Sephadex G50 (fine) (2,5x95 cm). Eluant: NH_4OH, pH 8,0, flow rate: 100 ml/hr.

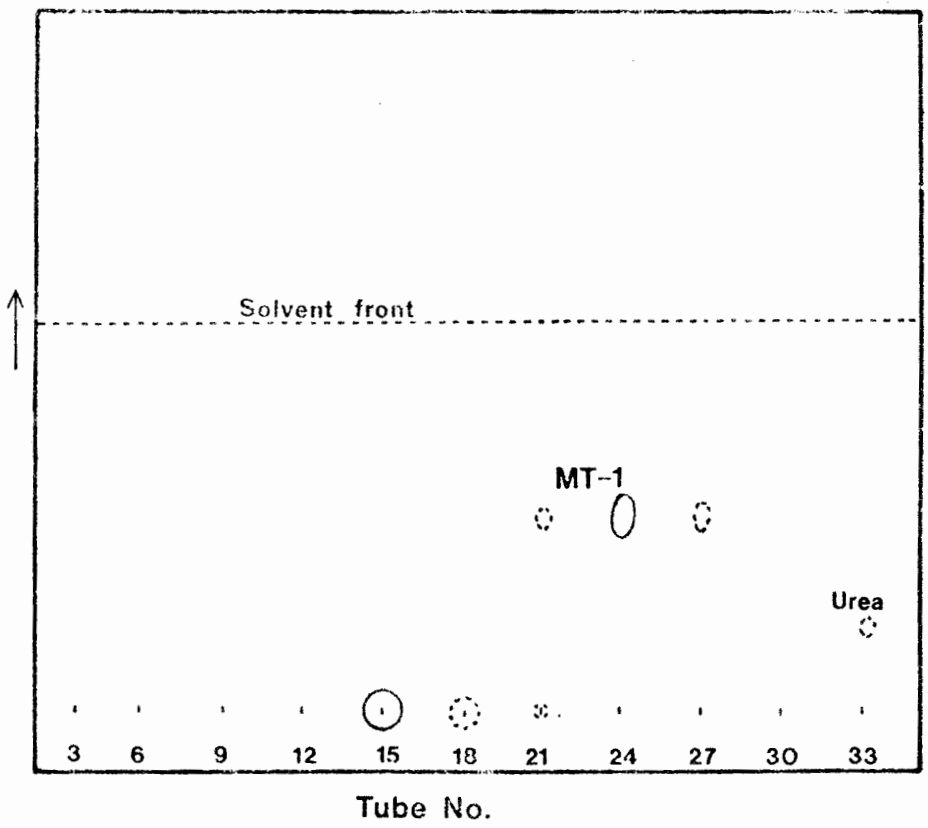


Fig. 2.31.

Gel exclusion chromatography of MT-1 on Sephadex G50 (fine)(1,5x95 cm). Eluant:5% formic acid. Peptides were detected by the ninhydrin reaction (4.4.1) after aliquots of every third fraction had been spotted onto cellulose thin layer plates.

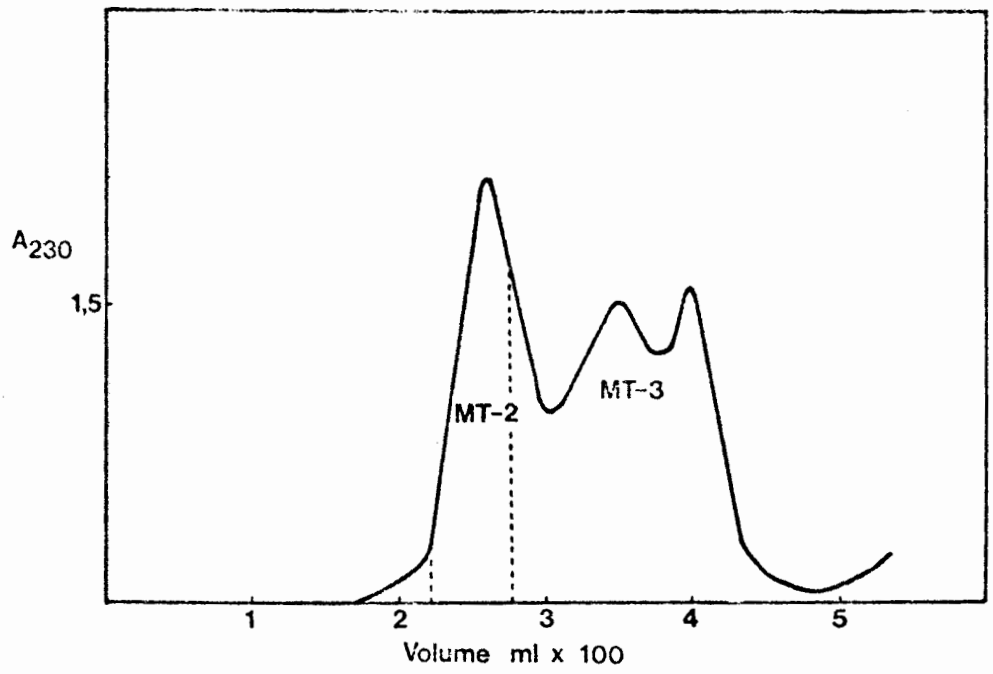


Fig. 2.32.

Gel exclusion chromatography of supernatant from trypsin cleavage of mollusc H2B on Sephadex G50. Conditions described in Fig. 2.30.

Peptide CN-2 was also maleylated prior to digestion with trypsin. The peptide mixture was chromatographed on a Sephadex G50 column (Fig. 2.33).

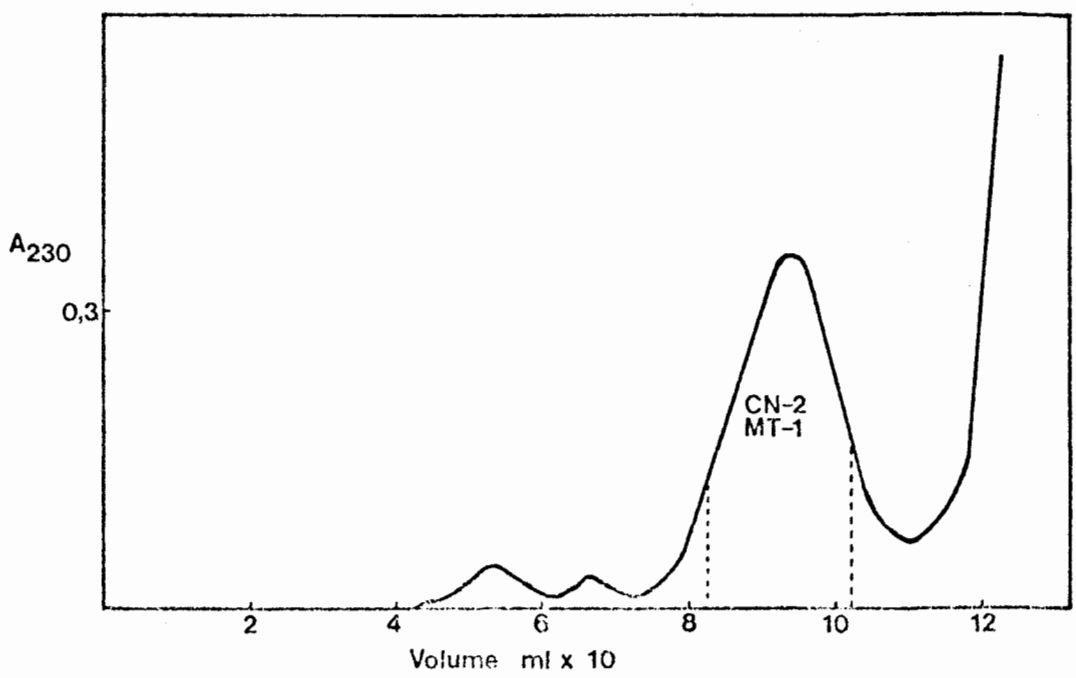


Fig. 2.33.

Gel exclusion chromatography of trypsin cleaved maleylated CN-2 on Sephadex G50 (fine)(2,5x95 cm). Eluant: 0,01 N HCl, flow rate: 100 ml/hr.

The large fraction which eluted near the inner volume of the column consisted of small peptides (6-10 residues) and was discarded.

2.2.1.6 Summary.

A summary of the purification of peptides used in the determination of the primary structure of mollusc H2B is given (Table 2.3.).

Table 2.3.

Purification of peptides used in sequence analysis of histone H2B patella.

Gel chromatography was on Sephadex in 0,01 M HCl or Bio-Gel in 0,02 N HCl, 0,05 M NaCl. Ion-exchange chromatography was on carboxymethyl-cellulose in 50 mM sodium acetate/HCl pH 4,5 with a sodium chloride gradient.

Peptide	Gel chromatography		Ion-exchange chromatography	
	Column size (mm)	gel type	Column size (mm)	
CN-1 and CN-2	25x900	Bio-Gel P30		
CN-3	15x900	Sephadex G15		
CN-1 TH-1	15x950	Sephadex G15		
MT-1	25x950	Sephadex G50(fine)[a]		
	15x950	Sephadex G50(fine)[b]		
CN-1 SA-2-IV	15x900	Sephadex G50(fine)	6x110	0—0,2 M (100 ml)
CN-1 SA-2-IV T-1	15x950	Sephadex G15		
CN-2 NBS-1	16x950	Sephadex G50(fine)		
CN-2 MT-1	25x950	Sephadex G50(fine)		
CN-2 MT-1 NBS-2	15x950	Sephadex G15		
	16x950	Sephadex G50(fine)		

[a] - eluant: NH_4OH pH 8,0.

[b] - eluant: 5% formic acid.

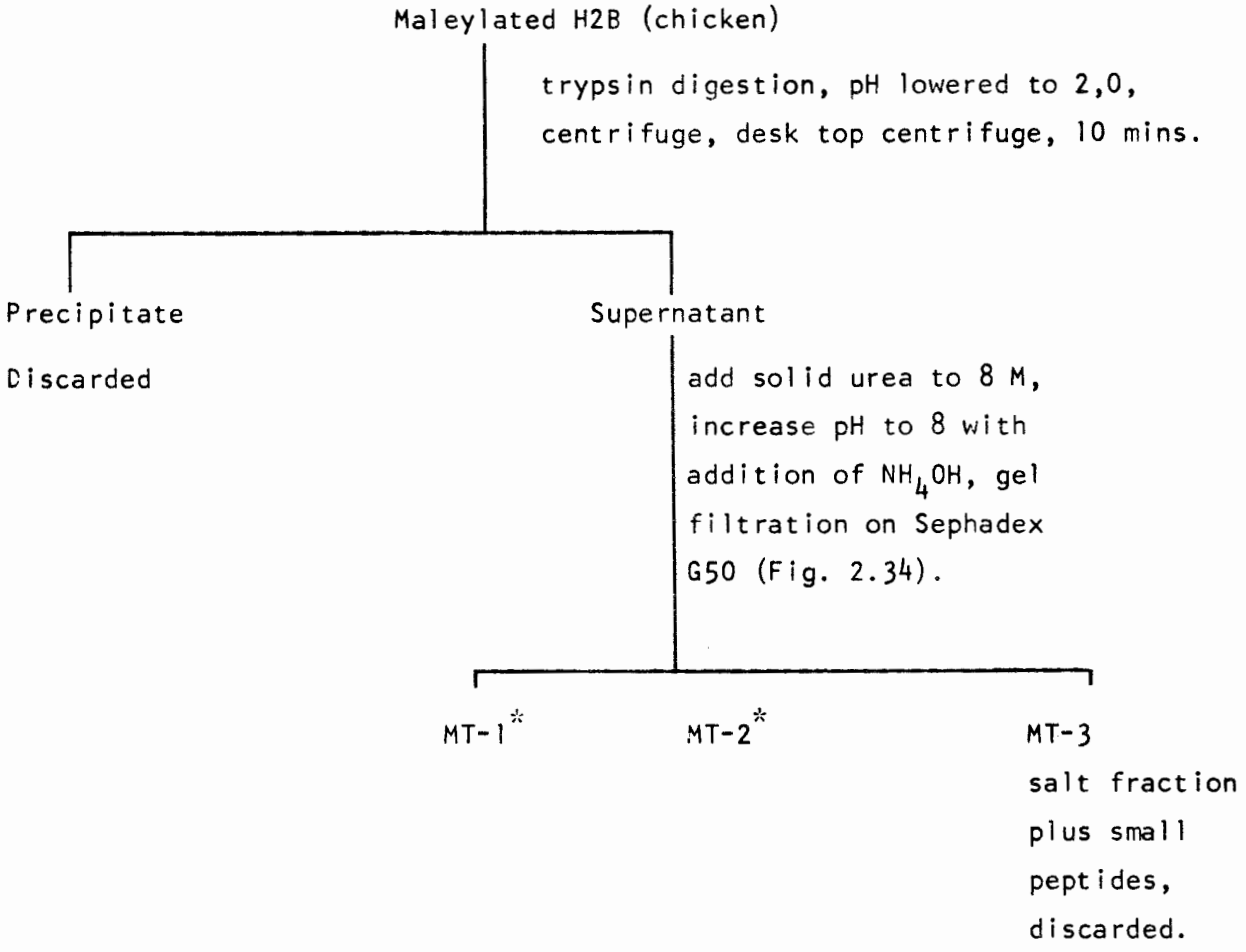
2.2.2 Gallus domesticus (chicken), Crocodylus niloticus (crocodile) and Xenopus laevis.

2.2.2.1 Cyanogen bromide peptides.

Pure H2B from each of the above species was cleaved as described in 4.3.2.1 and the resultant peptide mixture was separated on Bio-Gel P30 as described for Patella (2.2.1.1). The elution profile of the peptides of these three histones on Bio-Gel P30 is the same as that obtained for Patella (2.2.1.1).

2.2.2.2 Tryptic digestion of chicken erythrocyte H2B.

Tryptic cleavage of H2B was done after maleylation as described (4.3.3.1). The reaction was terminated by lowering the pH to 2 by the addition of HCl. The resulting peptide mixture was fractionated as follows;



* Peptides demaleylated (4.3.3.1).

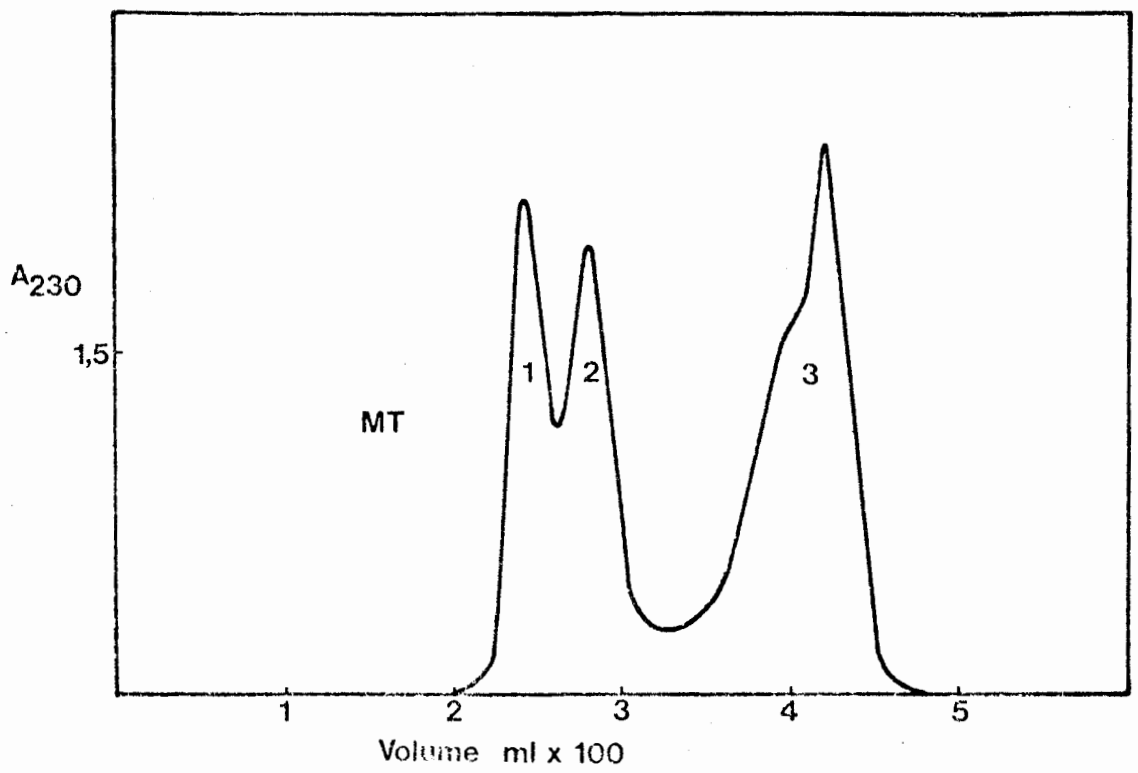


Fig. 2.34.

Gel exclusion chromatography of trypsin peptides from chicken erythrocyte H2B on Sephadex G50 (fine)(2,5x95 cm). Eluant: NH_4OH pH 8,0, flow rate: 100 ml/hr.

2.3 Amino acid analysis of H2B histones and peptides.

Amino acid analysis of histones H2B and constituent peptides was done as described in 4.4.3. The compositions are presented in the following tables.

No corrections for losses during or incompleteness of hydrolysis have been made.

TABLE 2.4. Amino acid composition (mole % and number of residues) of H2B *patella* and peptides used in sequence analysis.

Amino Acid	H2B (1-121)			CN-1 (1-55)			CN-2 (59-121)			CN-3 (56-58)			CN-1;TH-1 (17-34)		
Aspartic Acid	5,21	6,29	(6)	4,18	2,2	(2)	5,96	3,75	(4)				5,23	0,93	(1)
Threonine	5,61	6,78	(7)	1,88	0,99	(1)	9,11	5,74	(6)						
Serine	11,24	13,58	(15)	11,38	5,99	(7)	9,43	5,94	(7)	46,64	0,93	(1)	10,57	1,89	(3)
Glutamic Acid	6,83	8,25	(8)	4,06	2,14	(2)	9,10	5,73	(6)				5,57	0,99	(1)
Proline	3,21	3,88	(4)	5,74	3,02	(3)	1,86	1,17	(1)						
Glycine	5,07	6,12	(6)	6,76	3,56	(4)	3,09	1,95	(2)				5,59	1,0	(1)
Alanine	12,03	14,54	(14)	10,90	5,74	(6)	12,90	8,12	(8)				11,83	2,11	(2)
Valine	6,75	8,15	(8)	7,65	4,03	(4)	7,03	4,42	(4)						
Methionine	1,53	1,85	(2)	*		(1)				*		(1)			
Isoleucine	5,95	7,19	(7)	3,85	2,03	(2)	6,70	4,22	(4)	53,36	1,07	(1)			
Leucine	4,87	5,89	(6)	2,07	1,09	(1)	8,09	5,09	(5)						
Tyrosine	4,02	4,86	(5)	4,69	2,47	(3)	3,48	2,19	(2)				5,92	1,06	(1)
Phenylalanine	1,79	2,16	(2)			(0)	3,39	2,13	(2)						
Lysine	15,13	18,28	(18)	24,70	13,0	(13)	8,10	5,10	(5)				22,41	4,00	(4)
Histidine	2,22	2,68	(3)	2,01	1,06	(1)	3,74	2,35	(2)						
Arginine	8,53	10,31	(10)	10,13	5,33	(5)	8,01	5,05	(5)				32,87	5,86	(5)

No corrections for losses during or incompleteness of hydrolysis have been made.

* Homoserine and homoserine lactone present.

** Spirolactone ultraviolet absorbance present in the peptide. No quantitation after hydrolysis.

*** Residue destroyed on bromination. No quantitation after hydrolysis.

TABLE 2.4. Amino acid composition (mole % and number of residues) of H2B *patella* and peptides used in sequence analysis.

Amino Acid	MT-1 (30-68)			CN-1 SA-2-IV (32-55)			CN-1 SA-2-IV T-1 (43-53)			CN-2 NBS-1 (80-117)			CN-2 MT-1 (96-121)			CN-2 MT-1 NBS-2 (118-121)		
Aspartic Acid	9,85	3,79	(4)	5,42	1,24	(1)	9,57	1,04	(1)	2,72	0,98	(1)						
Threonine	2,57	0,99	(1)	4,51	1,04	(1)	8,69	0,94	(1)	14,38	5,19	(5)	12,08	2,95	(3)	26,38	0,94	(1)
Serine	14,30	5,5	(6)	16,49	3,80	(4)	15,69	1,71	(2)	7,67	2,77	(3)	9,11	2,22	(3)	44,23	1,58	(2)
Glutamic Acid	7,31	2,81	(3)	5,14	1,18	(1)	10,63	1,15	(1)	11,29	4,08	(4)	8,03	1,96	(2)			
Proline	2,81	1,08	(1)	5,26	1,21	(1)	9,69	1,05	(1)	3,02	1,09	(1)	5,12	1,24	(1)			
Glycine	2,66	1,02	(1)	4,11	0,95	(1)	8,73	0,95	(1)	5,21	1,88	(2)	8,35	2,04	(2)			
Alanine	2,95	1,13	(1)	4,57	1,05	(1)				11,35	4,09	(4)	12,47	3,04	(3)			
Valine	10,80	4,15	(4)	13,60	3,13	(3)	18,76	2,04	(2)	8,89	3,21	(3)	5,25	1,28	(2)			
Methionine	5,25	2,02	(2)															
Isoleucine	10,19	3,92	(4)	8,57	1,97	(2)				5,37	1,94	(2)						
Leucine	2,77	1,07	(1)	4,48	1,03	(1)				10,93	3,95	(4)	16,18	3,95	(4)			
Tyrosine	7,91	3,04	(3)	11,41	2,62	(3)				**		(1)						
Phenylalanine	5,91	2,27	(2)															
Lysine	9,57	3,68	(4)	13,02	3,0	(3)	8,42	0,92	(1)	10,42	3,76	(4)	15,36	3,75	(4)	29,40	1,05	(1)
Histidine	2,62	1,01	(1)	3,40	0,78	(1)	9,79	1,06	(1)	***		(1)	5,23	1,27	(1)			
Arginine	2,51	0,97	(1)							8,74	3,15	(3)						

No corrections for losses during or incompleteness of hydrolysis have been made.

* Homoserine and homoserine lactone present.

** Spirolactone ultraviolet absorbance present in the peptide. No quantitation after hydrolysis.

*** Residue destroyed on bromination. No quantitation after hydrolysis.

Table 2.5. Amino acid composition (mole % and number of residues) of H2B *patella* peptides.

Amino Acid	CN-1 SA-1 1-31			CN-1 SA-2 T-2 32-36			CN-2 NBS-2 59-79			MT-2 96-121		
Aspartic acid	3,85	1,19	(1)				16,94	2,88	(3)			
Threonine										8,23	1,96	(3)
Serine	8,92	2,76	(3)	22,66	1,81	(2)	10,44	1,78	(2)	9,38	2,23	(3)
Glutamic acid	3,89	1,20	(1)				10,87	1,85	(2)	8,09	1,93	(2)
Proline	6,41	1,98	(2)							3,93	0,94	(1)
Glycine	9,61	2,98	(3)							8,86	2,11	(2)
Alanine	17,52	5,42	(5)				25,71	4,37	(4)	12,71	3,03	(3)
Valine	3,06	0,95	(1)				4,27	0,73	(1)	7,47	1,78	(2)
Methionine												
Isoleucine				23,91	1,91	(2)	12,22	2,08	(2)			
Leucine							5,83	0,99	(1)	12,96	3,09	(4)
Tyrosine				37,74	3,02	(3)	**		(1)	4,22	1,00	(1)
Phenylalanine												
Lysine	30,26	9,37	(10)	15,69	1,26	(1)				20,43	4,86	(4)
Histidine										4,01	0,96	(1)
Arginine	16,48	5,10	(5)				13,73	2,33	(2)			

Table 2.6. Amino acid composition of Gallus domesticus H2B.

Amino acid	H2B		CN-1		CN-2		MT-1		MT-2	
	Mole %	[a]	Mole %	[a]	Mole %	[a]	Mole %	[a]	Mole %	[a]
Aspartic acid	4,63	5,82	3,83	2,34	6,63	4,02	3,69	1,05		
Threonine	6,74	8,47	5,05	3,09	8,84	5,36	6,64	1,90	11,64	2,90
Serine	7,87	9,90	10,14	6,20	9,06	5,49	6,79	1,94	10,15	2,53
Glutamic acid	8,54	10,74	7,38	4,51	9,80	5,94	7,22	2,06	8,29	2,07
Proline	4,59	5,77	7,92	4,84	1,66	1,01	13,33	3,81	4,96	1,24
Glycine	5,72	7,19	5,24	3,20	4,84	2,93	7,36	2,10	7,72	1,93
Alanine	10,03	12,61	9,08	5,55	11,19	6,78	14,13	4,04	13,09	3,26
Valine	6,38	8,03	6,55	4,01	6,49	3,93	3,31	0,95	8,29	2,07
Methionine	1,58	1,99	*							
Isoleucine	5,36	6,74	3,65	2,23	6,65	4,03				
Leucine	4,74	6,23	1,99	1,22	7,97	4,83			14,39	3,59
Tyrosine	3,59	4,40	4,95	3,03	3,26	1,98			3,95	0,99
Phenylalanine	1,45	1,82			3,11	1,88				
Lysine	19,16	24,1	28,05	17,15	8,86	5,37	34,26	9,79	14,25	3,55
Histidine	2,58	3,25	1,84	1,13	3,15	1,91			3,25	0,81
Arginine	6,55	8,23	4,26	2,61	8,41	5,09	3,23	0,92		

[a] - likely number of residues.

* - identified as homoserine.

Table 2.7. Amino acid composition of Crocodylus niloticus H2B.

Amino acid	H2B		CN-1		CN-2	
	Mole %	a	Mole %	a	Mole %	a
Aspartic acid	4,84	6,05	3,60	2,19	6,65	4,01
Threonine	6,08	7,60	4,94	3,02	10,43	6,29
Serine	9,14	11,43	8,66	5,29	10,52	6,34
Glutamic acid	8,02	10,03	7,06	4,31	10,40	6,27
Proline	4,51	5,64	8,46	5,17	1,89	1,14
Glycine	6,27	7,84	5,39	3,29	5,94	3,58
Alanine	10,16	12,70	8,92	5,44	12,31	7,42
Valine	6,93	8,66	7,06	4,31	6,93	4,17
Methionine	1,51	1,89	*			
Isoleucine	5,91	7,39	3,59	2,19	4,68	2,82
Leucine	4,84	6,05	2,15	1,31	6,77	4,08
Tyrosine	4,14	5,18	4,84	2,96	2,98	1,79
Phenylalanine	1,77	2,21			2,70	1,62
Lysine	16,34	20,43	26,38	16,12	7,53	4,54
Histidine	2,58	3,23	1,63	1,00	3,54	2,14
Arginine	6,78	8,48	5,32	3,25	6,71	4,04

[a] - likely number of residues.

* - identified as homoserine.

Table 2.8. Amino acid composition of Xenopus laevis H2B.

Amino acid	H2B		CN-1		CN-2	
	Mole %	[a]	Mole %	[a]	Mole %	[a]
Aspartic acid	5,12	6,40	4,16	2,29	6,20	3,94
Threonine	6,33	7,91	5,14	3,04	8,40	5,35
Serine	9,21	11,51	9,01	4,95	8,87	5,65
Glutamic acid	8,05	10,06	7,14	4,47	10,20	6,49
Proline	5,16	6,45	8,19	4,87	2,01	1,28
Glycine	5,16	6,45	5,34	3,10	5,04	3,21
Alanine	11,27	14,09	10,01	6,05	12,79	8,15
Valine	7,44	9,30	6,83	3,75	7,64	4,87
Methionine	1,60	2,00	*			
Isoleucine	4,88	6,10	4,44	2,55	5,17	3,29
Leucine	4,98	6,23	2,99	1,92	8,29	5,28
Tyrosine	4,02	5,03	4,39	2,41	2,91	1,85
Phenylalanine	1,76	2,20			2,46	1,58
Lysine	15,89	19,86	24,26	11,06	8,12	5,17
Histidine	2,50	3,13	1,51	0,83	2,94	1,87
Arginine	6,40	8,00	5,21	3,14	8,22	5,24

[a] - likely number of residues.

* - identified as homoserine.

2.4 Amino acid sequencing of histones H2B and peptides.

Introduction.

The possible number of sequential Edman degradations has increased considerably with the introduction of automatic sequence analysis of proteins or peptides. This has resulted in the situation that the sequential degradation of a small number of large peptides may suffice to elucidate the primary structure of the protein. If the protein is split into two or three fragments, it might be possible to align these fragments on the basis of the specificity of the cleavage reaction and by comparing their C and N terminal residues with those of the original protein (Brandt, 1974). This would eliminate the need for a set of overlapping peptides. If the peptide is too large, then it becomes necessary to produce overlapping peptides.

The primary structure of mollusc H2B has been determined from a set of adjoining and overlapping peptides. The H2B histones of chicken, crocodile and toad were only partly sequenced. In each case, intact H2B and CN-2 were sequenced as far as the data could be reliably interpreted.

Methods for sequential degradation were similar to those of Edman and Begg (1967)(see 4.5) and PTH-amino acids were then identified by one of three methods as described (4.5.3).

The solubility of the peptide and the background increase of other amino acids limits the number of Edman degradation cycles in which unambiguous assignments could be made. The background concentration of all amino acids increased as more cycles were completed and this may have been due to the following factors; (a) incomplete coupling, (b) incomplete cleavage, (c) random cleavage (Edman, 1975).

The assignment of an amino acid to a position therefore required quantitation of all amino acids at each cycle. An increase in the amount of one amino acid with no concommitant increase of any other, permitted a positive identification.

An assignment for a particular amino acid was made for an increase followed by a decrease. The absolute increase (in nmoles) and the amplitude of the increase depends on the sequence position and routinely residue yields are tabulated as shown.

Tables 2.10-2.19 show the yield of the amino acid assigned to position R and the background concentration of that particular amino acid in position R-1 and also the carry-over concentration into the next step (R+1).

Preliminary sequencing showed a drop in yield at proline residues, followed by a large carry-over of all amino acids following this residue. This has been shown in previous investigation from this laboratory to be due to a slow cleavage reaction. The proline-rich chicken H2B histone was used to compare cleavage rates of proline.

Proline was cleaved at significantly different rates depending on the nature of the following amino acid (Brandt et al. 1976). To improve the automatic cleavage of proline, the cleavage time of the program had been lengthened, or the sequenator had been operated manually at proline residues, although this would also expose the peptide to increased acidolysis at every step (Edman and Begg, 1967).

In this study the sequenator was set to stop the automatic cycle after cleavage of a proline residue and then the program was manually started at the acid cleavage. This doubled the time of acid cleavage only at proline residues and considerably reduced carry-over.

In previous investigations in this laboratory it had been observed that the yield of PTH-amino acids recovered after a glutamate residue often decreased (Strickland, 1977) and that this could be overcome by coupling of glycine-methyl ester to the carboxyl groups of glutamic and aspartic acids (Gibson, 1972). Hydrolysis and amino acid analysis of these modified residues results in an equimolar increase of glycine, thus clearly differentiating between the amino acids and their amides. The results in table 2.9 confirm that the substitution of the carboxyl groups is nearly complete

in that out of 10 expected glycine residues for the entire H2B, 9 were recovered and from a single Glu residue 89% of the expected glycine was recovered.

Table 2.9.

Modification of acidic amino acids of calf thymus H2B.*

Total acid hydrolysis.

Residue	Mole % Before modification	Mole % After modification	Difference
Aspartic acid	4,8	4,8	-
Glutamic acid	7,9	7,9	-
Glycine	5,5	12,1	+6,6 = 9 residues

	No. of Residues	No. of free carboxyl groups	No. of extra Glycine Residues
Aspartic acid	3		
Asparagine	3	10	9
Glutamic acid	7		
Glutamine	3		

	Results from sequential degradation Recovery after back-hydrolysis of PTH-amino-acid	n.mol.
Residue No. 2	Glutamic acid	8,4
	Glycine	7,5

* Calculated from sequence of calf thymus H2B (Iwai et al. 1972).

Short peptides and hydrophobic peptides were frequently washed out of the sequenator cup. This problem was eliminated by two procedures; (1) reaction of amino groups of lysine to 4-sulfophenylisothiocyanate (Braunitzer, 1973)(S-PITC) thus increasing the hydrophilic nature of the peptide. The reagent was particularly useful when the C-terminal residue was lysine (e.g. CN-1 SA-2 T-1). The S-PITC labelled lysine residues were not recovered (Braunitzer et al. 1973). If S-PITC and PITC were added together, then the yield of PTH-lysine was low, but the residue became identifiable. The first amino acid of S-PITC peptides was identified by dansylation of a separate aliquot or by overlapping peptides.
(2) The other procedure was to add carrier protein to the peptide in the sequenator cup. Poly amino acids have been made (Niall et al. 1974), but during this work, H4 from mollusc, which has a blocked N-terminus, was used. The carrier methodology is also compatible with the use of S-PITC. No double sequence or spurious amino acids appeared during the use of carrier.

These two procedures used together held the peptides firmly in the sequenator cup, e.g. with CN-2 MT-1, a peptide of 26 residues, with lysine as C-terminus, it was possible to detect serine at position 25.

2.4.1 The primary structure of Patella granatina (mollusc) H2B.

Since CN-1 was the only cyanogen bromide peptide with proline as a N-terminal amino acid and since H2B also has N-terminal proline, the peptide CN-1 has to be positioned at the N-terminus of H2B. The absence of homoserine and homoserine lactone in the composition of CN-2 placed it at the C-terminus of H2B and the presence of homoserine in the small peptide put the latter into the centre of the molecule;

CN-1	CN-3	CN-2
55 residues	3 residues	63 residues
Pro ----------	Met-Ser-Ile-Met-As(x)	---------- Lys
1	55	121

3,80 mg of unmodified mollusc H2B were subjected to 33 degradation cycles and the resultant PTH-amino acids quantitated by liquid chromatography (Table 2.10). Proline was identified in the first step. In the second step, proline was present but in a low yield. However, in the second step no other PTH-amino acid was recovered and therefore proline was tentatively assigned to position two. This assignment was proved by a subsequent experiment. The poor cleavage yield of the first two residues resulted in a very high carry-over (Table 2.10) and meaningful assignments were possible up to residue 24 although residues 30-33 were still recognisable. However, a gap was left between residues 25-30.

To achieve assignment in this region a thermolysin peptide was subjected to degradation (CN-1 TH-1, Table 2.12). This peptide overlaps with intact H2B from residue 17 onwards. The peptide was unambiguously sequenced for 14 steps corresponding to position 30 in the intact protein.

The identification of position two in the sequence is difficult. The drop in proline could be interpreted as the result of incomplete cleavage in the first step. However, no new N-terminal residue appears in the quantitation of PTH-amino acids at cycle two. This led to the initial assumption that position two is also occupied by proline.

To prove this 1,3 mg of H2B was added to the sequenator cup and reacted with S-PITC in an equimolar amount to the free amino groups. The terminal proline residue was coupled 100% to S-PITC since no PTH-proline was recovered at step 1 (Table 2.11). However, a total of 61 nmoles PTH-proline was recovered in three successive cleavages of position 2. This PTH-proline could not have been due to carry over and therefore proline was positively identified at position 2. The recovery of 61 nmoles of PTH-proline at position 2 also agrees well with the recovery of 60 nmoles of PTH-valine at position 4. In this experiment as well as the initial degradation from residue two onwards a second minor sequence became apparent, frameshifted by one and complementary to the main sequence. This is shown by the high initial values of R-1 (Table 2.10).

This frameshifted sequence may be the result of the following; the coupled PITC cyclizes to the second peptide bond, thus yielding phenylthiocarbamyl-prolyl-proline (PTC-pro-pro) and that this derivative was not recovered. A model of PITC-pro-pro (Fig. 2.35) illustrates that the reactive group of PITC can come sufficiently close to the second peptide bond to react with it.

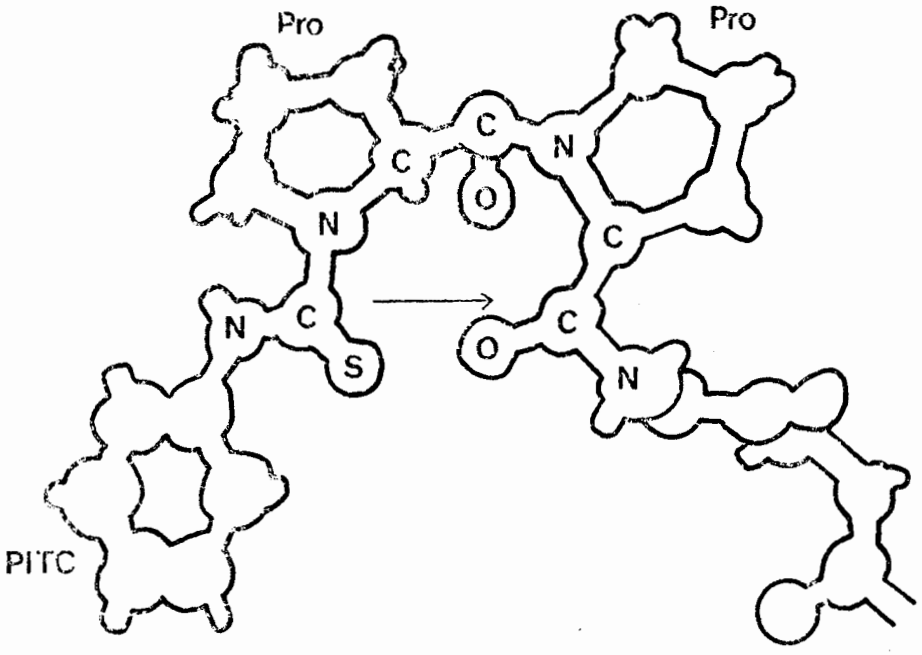


Fig. 2.35.

Molecular model of PITC-pro-pro protein.

Positions 30-33 had been assigned by sequencing intact H2B. To overlap these positions MT-1 was sequenced for eight steps (Table 2.12). Peptide MT-1 formed a bridge between the sequenced portion of H2B histone and the major peptide CN-1 SA-2-IV (Table 2.13).

Cleavage of CN-1 by Staphylococcus aureus protease gave two peptides (2.2.1.4) and they were aligned by N-terminal analysis. Mollusc CN-1 was 55 residues in length and sequencing from the N-terminus,

the first glutamic acid was placed at position 31 (Table 2.10) leaving 24 residues. The alignment is schematically given;

	CN-1 SA-1		CN-1 SA-2-IV	
	31 residues		24 residues	
Pro	----------	Glu-Ser	----------	Met
1		31		55

CN-1 SA-2-IV was sequenced for fifteen steps (Table 2.13). The sequence of this peptide overlapped the previous peptide (MT-1). The first residue was not quantitated, but cleaved and identified by the combined Dansyl-Edman technique described in 4.4.4.3. It was not possible to sequence CN-1 SA-2-IV in its entirety, thus new peptides had to be generated. The amino acid composition of CN-1 SA-2-IV indicated the presence of 3 lysine residues. Two lysine residues were placed at positions 39 and 42 during the sequencing of CN-1 SA-2-IV. The position of the third lysine was determined by carboxypeptidase treatment of CN-1 (Table 2.19 and Fig. 2.36). The unsequenced portion of CN-1 SA-2-IV lay between the second and third lysine.

From the known sequence of CN-1 SA-2-IV and its amino acid composition, four peptides (sequentially 8, 3, 11 and 2 residues) were expected after tryptic digestion. Two peptides were purified. N-terminal and amino acid analysis showed that one (CN-1 SA-2-IV T-1) was 11 residues long, while the second (CN-1 SA-2-IV T-2) was 8 residues long (Table 2.5). This information together with the known sequence made alignment possible as shown schematically;

		CN-1 SA-2-IV				
	8 residues		3 residues	11 residues	2 residues	
Ser	----------	Lys-Val-Leu-Lys-Gln		---------	Lys-Ala-Met	
32		39		43	53	55
	T-2		(T-3)	T-1		(T-3)

The tryptic peptide (CN-1 SA-2-IV T-1) included the unsequenced region (Table 2.13).

The results of carboxypeptidase digestion of CN-1 are presented diagrammatically (Fig. 2.36) and in tabular form (Table 2.19).

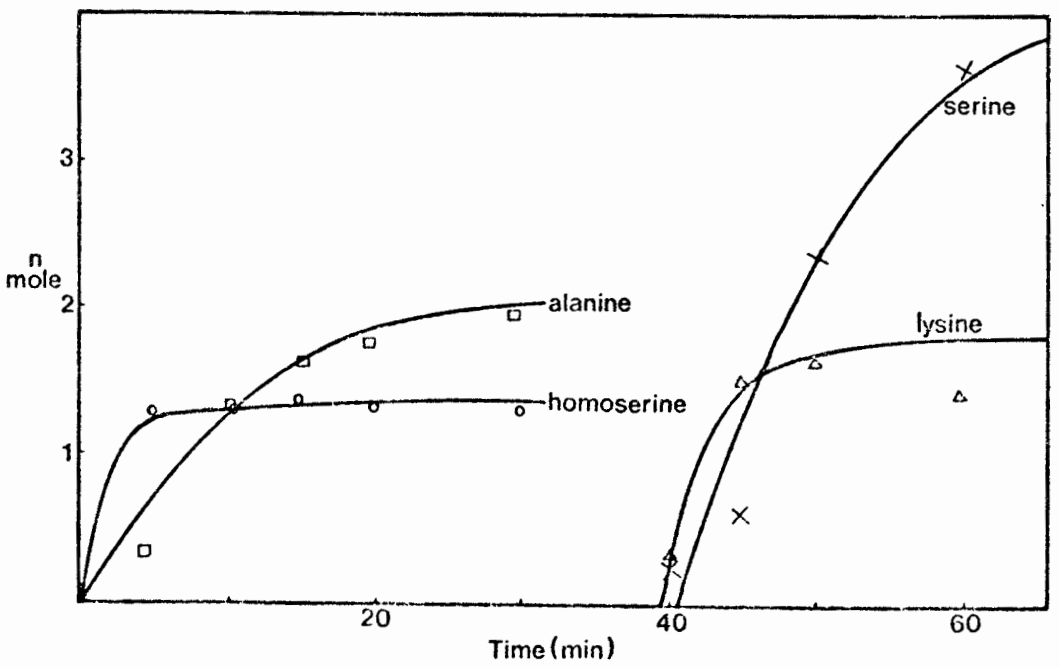


Fig. 2.36.

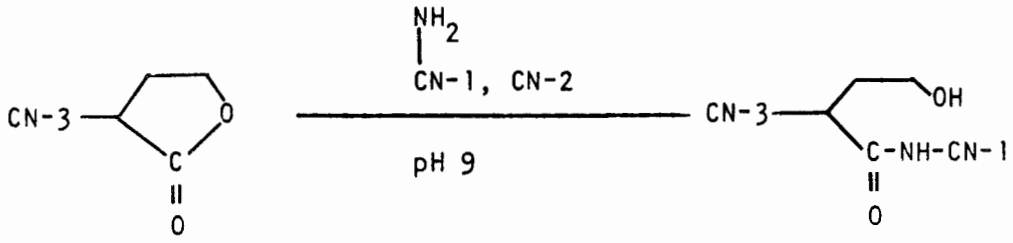
Carboxypeptidase digestion of mollusc CN-1. Carboxypeptidase A added at time 0 and B at 40 min. Quantitation of free amino acids on an analyser as described (4.4.3).

The liberation of amino acids on sequential addition of carboxypeptidases A and B assigns the following sequence to the C-terminus of CN-1; -Ser-Lys-Ala-Met.

From amino acid analysis the composition of CN-3 was found to be Ser_1, Ile_1, Met_1. The specificity of CNBr puts Met at the C-terminus and dansylation Ser at the N-terminus. Therefore the sequence of CNBr is -Ser-Ile-Met.

CN-3 could not be automatically sequenced on its own because it

washed out of the sequenator cup. Therefore intact H2B was cleaved by cyanogen bromide and after the CNBr was removed by freeze-drying, the peptides were treated with trifluoroacetic acid (TFA) to increase the cleavage yield. At acidic pH, the homoserine/homoserine lactone equilibrium moves in the direction of the closed ring lactone structure (Armstrong, 1949). This peptide mixture was placed directly into the sequenator cup and propyne buffer (4.5.2) added to the dried peptides. At the pH of the buffer (approx. 9) it is likely that the lactone attaches itself to amino groups of lysine or arginine;



This reaction of the homeserine lactone has been used to attach peptides to an insoluble resin support (Horn et al. 1973).

During sequencing, CN-1 and CN-2 do not wash out of the cup and thus CN-3 may be sequenced by its covalent attachment to lysine and arginine. The sequence of this mixture is presented in table 2.14. The first degradation cycle produced three residues, i.e. proline, asparagine and serine. Proline is the known end-group for CN-1 and asparagine for CN-2, therefore serine must be the first residue of CN-3. Similarly, three residues became evident for step 2; proline from CN-1, serine from CN-2 and therefore isoleucine from CN-3. These results agree with the sequence given above.

Peptide CN-2 has been assigned to the carboxyl end of the protein. The acidic amino acids were coupled to glycine-methyl ester and then CN-2 was automatically sequenced for 32 steps (Table 2.15). The peptide has a tyrosine residue at position 79 which allowed the production of an NBS peptide to form a long overlap, however, there are two tyrosine

residues in CN-2 so cleavage by NBS should yield three peptides.

Only two peptides were recovered and both had As(x) as N-termini. Peptide CN-2 NBS-1 was sequenced for eighteen steps and was found to overlap the known sequence of CN-2 beginning at position 80 (Table 2.16). Amino acid analysis and Edman degradation for four steps showed that CN-2 NBS-2 corresponded to the amino end region of CN-2.

CN-2

21 residues 38 residues (4 residues)

Asn ---------- Tyr-Asn ---------- Tyr ------ Lys

NBS-2 NBS-1

59 79 117 121

Five arginine residues were found in peptide CN-2 and all five were positioned by partially sequencing either CN-2 or CN-2 NBS-1. The fifth arginine residue occurred at the third last position of the sequenced part in CN-2 NBS-1.

A peptide (MT-2) which overlapped the sequenced portion of CN-2 NBS-1 was recovered from a tryptic digest of maleylated H2B histone (2.2.1.5). It was sequenced for eleven steps; namely from position 96 to 106 (Table 2.17). Following the glutamic acid residue (position 101) the yields of PTH-amino acids were very poor. This peptide (MT-2) was also recovered as the largest peptide (CN-2 MT-1) from tryptic digestion of maleylated CN-2 (2.2.1.5).

This time, however, approximately 200 nmoles of peptide were added to approximately the same amount of carrier protein, i.e. histone H4. This mixture was dried in the sequenator cup and a fourfold excess of S-PITC used for the first step. Twentysix degradation cycles were done (Table 2.18) and the first 23 were unambiguously assigned. Although the peptide was coupled with S-PITC, small amounts of PTH-lysine were observed (positions 104, 112 and 116) indicating that the coupling was incomplete but was sufficient to retain the peptide in the sequenator cup. The quantitation of valine was done by gas chromatography, as after

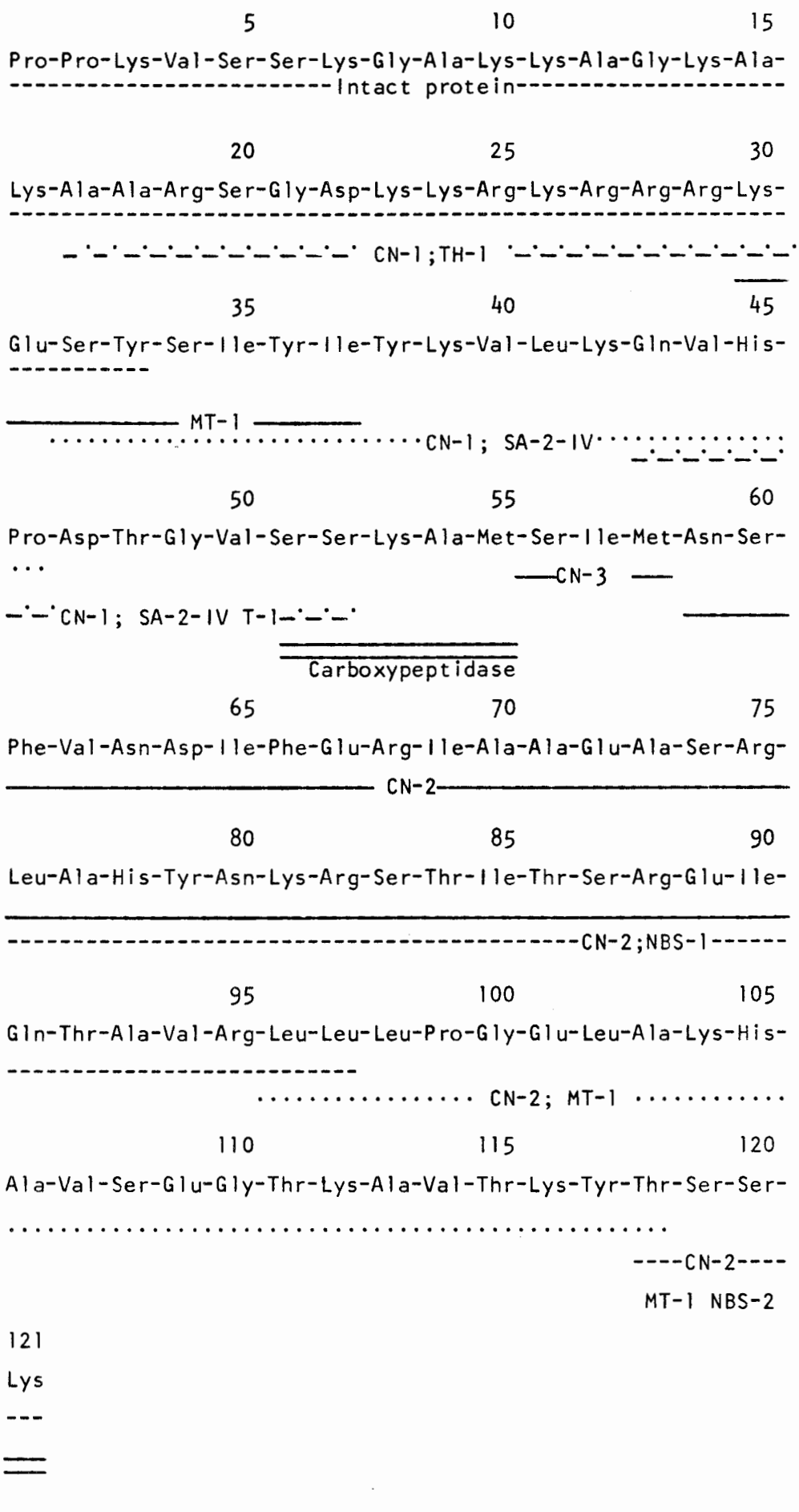
S-PITC labelling, a large contaminant eluted at the same position as valine during liquid chromatography. After glutamic acid (position 109) an increase in carry-over became evident, but the unambiguous assignment of residues was still possible as far as threonine 118. The following two residues showed up as serine breakdown products, but in very low yields.

As there is a tyrosine residue at position 117, the last few residues were identified by producing a peptide with NBS cleavage. The difference between the amino acid composition and the sequenced portion of CN-2 MT-1 indicated that four residues remained after tyrosine 117. This small peptide was recovered as CN-2 MT-1 NBS-2. The peptide was placed in the sequenator cup together with carrier and then a mixture of S-PITC and PITC was added. The results obtained (Table 2.19) support those from the previous peptide and agree with amino acid analysis i.e. thr, ser_2, lys. The last residue, i.e. lys, gave only a small increase, because of S-PITC labelling, but this residue was confirmed by carboxypeptidase digestion.

This was done with carboxypeptidase B for 15 minutes (Table 2.19) and lysine identified as the C-terminal residue of mollusc H2B.

The complete sequence of mollusc H2B was thus established and is given in Fig. 2.37.

Fig. 2.37.


5 10 15
Pro-Pro-Lys-Val-Ser-Ser-Lys-Gly-Ala-Lys-Lys-Ala-Gly-Lys-Ala-
Intact protein
20 25 30
Lys-Ala-Ala-Arg-Ser-Gly-Asp-Lys-Lys-Arg-Lys-Arg-Arg-Arg-Lys-
CN-1;TH-1
35 40 45
Glu-Ser-Tyr-Ser-Ile-Tyr-Ile-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-
MT-1
CN-1; SA-2-IV
50 55 60
Pro-Asp-Thr-Gly-Val-Ser-Ser-Lys-Ala-Met-Ser-Ile-Met-Asn-Ser-
CN-3
CN-1; SA-2-IV T-1
Carboxypeptidase
65 70 75
Phe-Val-Asn-Asp-Ile-Phe-Glu-Arg-Ile-Ala-Ala-Glu-Ala-Ser-Arg-
CN-2
80 85 90
Leu-Ala-His-Tyr-Asn-Lys-Arg-Ser-Thr-Ile-Thr-Ser-Arg-Glu-Ile-
CN-2;NBS-1
95 100 105
Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-
CN-2; MT-1
110 115 120
Ala-Val-Ser-Glu-Gly-Thr-Lys-Ala-Val-Thr-Lys-Tyr-Thr-Ser-Ser-
CN-2
MT-1 NBS-2
121
Lys

Automatic sequencing of histone H2B *patella* and its peptides.

Amounts of amino acid derivatives below 1 nmole have been recorded as 0.

(S-PITC) following peptide identification symbol = 4-sulfophenyl isothiocyanate substituted peptide.

(C) = carrier protein (histone H4) added

(Gly) = Aspartyl-and glutamyl glycine derivative of peptide.

Residues were identified and quantitated by high pressure liquid chromatography, gas chromatography (*) or after acidic hydrolysis via amino acid analysis (_).

R = yield (nmoles) of amino acid derivative assigned to that position, R-1 = yield of that derivative in the preceding cycle and R+1 in the following cycle.

( ) = Sequence position.

a = on hydrolysis accompanied by an equimolar rise in glycine.

b = determined as dansyl derivative.

c = overlap with another peptide.

Peptides CN-1, SA-2, T-1 and CN-2, MT-1 were reached with S-PITC only and peptide CN-2, MT-1, NBS-2 with a 1:1 mixture of S-PITC and PITC.

Table 2.10.

Intact protein.

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1	Pro	-	166	118	17[c]	Ala	43	66	76
2	Pro	166	118	22	18[c]	Ala	66	76	51
3	Lys	11	61	43	19[c]	Arg	23	72	42
4	Val	10	82	33	20[c]	Ser	9	40	31
5	Ser	11	97	93	21[c]	Gly	10	41	38
6	Ser	97	93	31	22[c]	Asp	5	51	30
7	Lys	17	78	36	23[c]	Lys	18	34	45
8	Gly	13	84	45	24[c]	Lys	34	45	35
9	Ala	12	70	45	25[c]	Arg	5	73	51
10	Lys	33	65	87	26[c]	Lys	35	35	27
11	Lys	65	87	47	27[c]	Arg	51	59	65
12	Ala	28	76	53	28[c]	Arg	59	65	66
13	Gly	28	76	56	29[c]	Arg	65	66	47
14	Lys	28	60	40	30[c]	Lys	16	23	20
15	Ala	30	67	43	31[c]	Glu	9	18	16
16	Lys	40	51	38	32[c]	Ser	6	12	11
					33[c]	Tyr	7	17	

Table 2.11.

Intact Protein (S-PITC).

Residue No.	Residue	nmoles				
		R-1	Cleavage 1	Cleavage 2	Cleavage 3	R+1
1	Pro [a]	-	0	0	0	43
2	Pro	0	43	12	6	15
3	Lys [a]	2	28	-	-	18
4	Val	10	60	-	-	-

[a] - low yield due to S-PITC coupling.

Table 2.12.

CN-1 TH-1.

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (17)[c]	Ala	-	70	74
2 (18)[c]	Ala	70	74	30
3 (19)[c]	Arg	25	158	65
4 (20)[c]	Ser	16	31	27
5 (21)[c]	Gly	26	75	33
6 (22)[c]	Asp	16	31	18
7 (23)[c]	Lys	42	106	111
8 (24)[c]	Lys	106	111	41
9 (25)[c]	Arg	33	106	41
10 (26)[c]	Lys	41	85	44
11 (27)[c]	Arg	41	99	105
12 (28)[c]	Arg	99	105	100
13 (29)[c]	Arg	105	100	31
14 (30)[c]	Lys	27	51	23
15 (31)[c]	Glu	5	4	

MT-1

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (30)[c]	Lys	-	43	15
2 (31)[c]	Glu	4	28	13
3 (32)[c]	Ser	0	21	13
4 (33)[c]	Tyr	5	26	13
5 (34)	Ser	13	21	13
6 (35)	Ile	5	25	11
7 (36)	Tyr	10	26	8
8 (37)	Ile	11	31	15

Table 2.13.

CN-1 SA-2-IV(C).

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (32)[b,c]	Ser	-	-	-
2 (33)[c]	Tyr	0	42*	4
3 (34)[c]	Ser	0	63*	15
4 (35)[c]	Ile	0	70*	6
5 (36)[c]	Tyr	0	47*	6
6 (37)[c]	Ile	6	71*	20
7 (38)	Tyr	7	61*	8
8 (39)	Lys	2	14	-
9 (40)	Val	0	51*	19
10 (41)	Leu	5	52*	24
11 (42)	Lys	3	14	-
12 (43)[c]	Glx	1	12	-
13 (44)[c]	Val	14	69*	19
14 (34)[c]	His	0	8	1
15 (46)[c]	Pro	2	22*	19

CN-1 SA-2 T-1 (C,S-PITC).

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (43)[c]	Gln	-	33	13
2 (44)[c]	Val	0	95	19
3 (45)[c]	His	0	33	8
4 (46)[c]	Pro	8	78	43
5 (47)	Asp	11	54	25
6 (48)	Thr	6	44	16
7 (49)	Gly	17	54	20
8 (50)	Val	14	35	16
9 (51)	Ser	2	20	21
10 (52)	Ser	20	21	7

Table 2.14.

CNBr Peptides; no separation.

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (1)	Pro	-	28	35
(56)	Ser	-	22	30
(59)	Asn	-	23	13
2 (2)	Pro	28	35	12
(57)	Ile	18	44	17
(60)	Ser	22	30	13
3 (3)	Lys	0	23	
(61)	Phe	0	40	

CN-1; Pro(1)-Pro(2)-Lys(3) (Table 2.10).

CN-2; Asn(59)-Ser(60)-Phe(61) (Table 2.15)

CN-3; Ser(56)-Ile(57)-Met*

* Met present as homoserine/homoserine lactone.

Table 2.15.

CN-2 (Gly).

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1 (59)	Asn	-	75*	7	15 (73)	Ala	31	116*	32
	Asp	-	33	-	16 (74)	Ser	0	87*	2
2 (60)	Ser	0	214*	-	17 (75)	Arg	4	69	-
3 (61)	Phe	0	158*	12	18 (76)	Leu	9	82*	28
4 (62)	Val	0	180*	23	19 (77)	Ala	12	101*	24
5 (63)	Asn	0	98*	12	20 (78)	His	16	77*	-
	Asp	0	25	101	21 (79)	Tyr	1	61*	24
6 (64)	Asp	25	101	-	22 (80)[c]	Asn	0	29*	8
	Asn	98	12*	-		Asp	0	8	-
7 (65)	Ile	9	155*	31	23 (81)[c]	Lys	9	69	-
8 (66)	Phe	0	133*	17	24 (82)[c]	Arg	3	31	-
9 (67)	Glu	0	76	-	25 (83)[c]	Ser	3	20*	4
10 (68)	Arg	2	84	-	26 (84)[c]	Thr	3	11*	3
11 (69)	Ile	19	120*	29	27 (85)[c]	Ile	10	37*	15
12 (70)	Ala	6	123*	141	28 (86)[c]	Thr	3	9*	2
13 (71)	Ala	123	141*	31	29 (87)[c]	Ser	0	17*	-
14 (72)	Glu	9	52	-	30 (88)[c]	Arg	3	14	-
					31 (89)[c]	Glu	7	13*	-
					32 (90)[c]	Ile	2	4	-

Table 2.16.

CN-2, NBS-1.

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1 (80)[c]	Asn	-	27	16	10 (89)[c]	Glu	0	22*	12
	Asp	-	25	17	11 (90)[c]	Ile	6	33*	18
2 (81)[c]	Lys	1	32	18	12 (91)	Gln	0	6*	2
3 (82)[c]	Arg	0	45	-		Glu	12	13*	5
4 (83)[c]	Ser	6	39*	15	13 (92)	Thr	6	13*	5
5 (84)[c]	Thr	5	48*	22	14 (93)	Ala	6	12*	17
6 (85)[c]	Ile	9	49*	24	15 (94)	Val	1	13*	11
7 (86)[c]	Thr	22	32*	8	16 (95)	Arg	6	15	-
8 (87)[c]	Ser	0	19*	4	17 (96)[c]	Leu	5	14*	24
9 (88)[c]	Arg	7	32	-	18 (97)[c]	Leu	14	24*	

Table 2.17.

MT-2.

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (96)	Leu*	-	-	-
2 (97)	Leu	-	46	39
3 (98)	Leu	46	39	2
4 (99)	Pro	2	31	2
5 (100)	Gly	1	24	4
6 (101)	Glu	0	22	-
7 (102)	Leu	1	10	1
8 (103)	Ala	2	9	3
9 (104)	Lys	1	2	1
10 (105)	His	0	2	
11 (106)	Ala	2	2	

* This step was performed by the micro Dansyl-Edman procedure and was not quantitated.

Table 2.18.

CN-2 MT-1 (C, S-PITC).

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1 (96)[c,b]	Leu	-	-	160	13 (108)	Ser	2	23	7
2 (97)[c]	Leu	-	160	168	14 (109)	Glu	10	29	-
3 (98)	Leu	160	168	39	15 (110)	Gly	19	60	40
4 (99)	Pro	6	143	45	16 (111)	Thr	0	38*	-
5 (100)	Gly	25	140	25	17 (112)	Lys	9	27	20
6 (101)	Glu	25	147	39	18 (113)	Ala	14	48	52
7 (102)	Leu	5	119	53	19 (114)	Val	11	44*	34
8 (103)	Ala	7	101	51	20 (115)	Thr	11	35*	-
9 (104)	Lys	10	45	24	21 (116)	Lys	10	18	18
10 (105)	His	5	15	3	22 (117)	Tyr	11	38	35
11 (106)	Ala	23	88	40	23 (118)[c]	Thr	9	22*	-
12 (107)	Val	9	57*	-					

Table 2.19.

CN-2 MT-1 NBS-2 (C,S-PITC).

Residue No.	Residue	nmoles		
		R-1	R	R+1
1 (118)[c]	Thr	-	38	14
2 (119)	Ser	2	17	20
3 (120)	Ser	17	20	11
4 (121)	Lys	7	9	-

CN-1 (Carboxypeptidase digestion).

Residue	Time (min.)										
nmole	0	5	10	15	20	30	40	45	50	60	70
Met*	0	1,29	1,31	1,38	1,33	1,31	1,39	1,49	1,21	1,31	1,49
Ala	0	,358	1,29	1,68	1,77	1,95	1,96	2,09	2,59	2,64	3,09
Lys	0	0	0	0	0	0	0,19	1,53	1,62	1,41	1,79
Ser	0	0	0	0	0	0	0,15	0,59	2,38	3,69	3,98

* Measured as homoserine.

'0' indicates quantitation less than 0,25 nmoles.

Intact protein (Carboxypeptidase B digestion).

Residue	Time (min.)		
nmole	0	10	15
Lysine (121)	1,62	4,26	4,40

2.4.2 The partial primary structure of Gallus domesticus (chicken) H2B.

The acidic amino acids of pure chicken H2B were modified and then approximately 6,5 mg were automatically sequenced for 37 cycles (Table 2.20). The first 29 residues were unambiguously assigned while the remaining residues were tentatively assigned. For these last residues, quantitative increases were small but there was no increase in other background amino acids.

Since homoserine was absent from CN-2, this peptide was assigned to the carboxyl end of the protein. It was sequenced for 27 steps all of which could be conclusively identified. Acidic amino acids were not modified resulting in decreased yields of PTH-amino acids after glutamic acid residues (Table 2.21).

Peptide MT-2 was obtained by tryptic digestion of maleylated H2B. Since no arginine was present in this peptide (Table 2.6.) it could be placed at the carboxyl terminal end. Dansyl-leucine was the N-terminus. The amino groups were coupled to S-PITC and the acidic amino acids were modified followed by 19 degradation cycles (Table 2.22).

To determine the carboxyl terminal residue, intact H2B was incubated with carboxypeptidase B only for 60 minutes, at which time carboxypeptidase A was added. At 60 minutes, no significant quantities (0,25 nmole) of acidic or neutral free amino acid was detected. The rapid increase in serine content is thus due to the addition of Cp-A (Fig. 2.38).

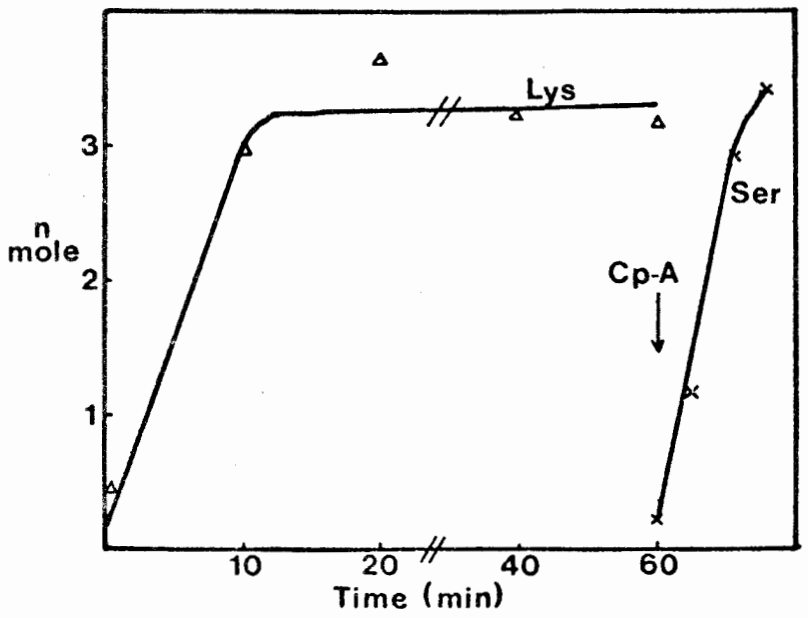


Fig. 2.38.

Carboxypeptidase digestion of chicken H2B histone. Carboxypeptidase B added at time 0 and A at time 60 min.

Therefore the C-terminal sequence is -Ser-Lys-COOH. The partial primary structure obtained from sequencing these peptides and the intact protein is given (Fig. 2.39).

```
                                        5                                  10
Calf      NH2-Pro - Glu - Pro - Ala - Lys - Ser - Ala - Pro - Ala - Pro -
Chicken   NH2-Pro - Glu - Pro - Ala - Lys - Ser - Ala - Pro - Ala - Pro -

                                       15                                  20
Calf           Lys - Lys - Gly - Ser - Lys - Lys - Ala - Val - Thr - Lys -
Chicken        Lys - Lys - Gly - Ser - Lys - Lys - Ala - Val - Thr - Lys -

                                       25                                  30
Calf           Ala - Gln - Lys - Lys - Asp - Gly - Lys - Lys - Arg - Lys -
Chicken        Thr - Gln - Lys - Lys - Gly - Asp - Lys - Lys - Arg -(Arg)-

                                       35
Calf           Arg - Ser - Arg - Lys - Glu - Ser - Tyr -
Chicken       (Lys)-(Ser)-(Arg)-(Lys)-(Glu)-(Ser)-(Tyr)-

                                      ------

                 62                    66                                  71
Calf           Met - Asn - Ser - Phe - Val - Asn - Asp - Ile - Phe - Glu -
Chicken        Met - Asn - Ser - Phe - Val - Asn - Asp - Ile - Phe - Glu -

                                       76                                  81
Calf           Arg - Ile - Ala - Gly - Glu - Ala - Ser - Arg - Leu - Ala -
Chicken        Arg - Ile - Ala - Gly - Glu - Ala - Ser - Arg - Leu - Ala -

                                       86
Calf           His - Tyr - Asn - Lys - Arg - Ser - Thr - Ile -
Chicken        His - Tyr - Asn - Lys - Arg -(Ser)- Thr - Ile -

                                      ------

                 99                   103                                 108
Calf           Arg - Leu - Leu - Leu - Pro - Gly - Glu - Leu - Ala - Lys -
Chicken        Arg - Leu - Leu - Leu - Pro - Gly - Glu - Leu - Ala - Lys -

                                      113                                 118
Calf           His - Ala - Val - Ser - Glu - Gly - Thr - Lys - Ala - Val -
Chicken        His - Ala - Val - Ser - Glu - Gly - Thr -(Lys)- Ala - Val -

                                      123          125
Calf           Thr - Lys - Tyr - Thr - Ser - Ser - Lys - COOH
Chicken       (Thr1, Lys1, Tyr1, Thr1, Ser1,)Ser - Lys - COOH
```

Fig. 2.39.

Comparison of the partial primary structure of H2B histone from erythrocytes of Gallus domesticus to that of calf thymus H2B. Residues are numbered according to alignment positions assuming a comparable total length of the polypeptide chains.

Table 2.20.

Gallus domesticus - Intact Protein (Gly).

Residue No.	Residue	n moles			Residue No.	Residue	n moles		
		R-1	R	R+1			R-1	R	R+1
1	Pro	-	394*	12	20	Lys	21	54*	24
2	Glu[a]	0	133	14	21	Thr	32	50*	27
3	Pro	12	245*	24	22	Gln	3	30*	12
4	Ala	0	335*	50		Glu	10	25*	14
5	Lys	16	201*	60	23	Lys	23	43*	50
6	Ser	17	310*	61	24	Lys	43	50*	28
7	Ala	24	311*	51	25	Gly	10	47*	30
8	Pro	9	147*	26	26	Asp	1	11	6
9	Ala	51	239*	45	27	Lys	17	26*	25
10	Pro	26	128*	18	28	Lys	25	25*	19
11	Lys	9	167*	133	29	Arg	11	24	20
12	Lys	167	133*	97	30	(Arg)	24	20	16
13	Gly	30	286*	124	31	(Lys)	18	23*	18
14	Ser	25	87*	31	32	(Ser)	3	4*	4
15	Lys	37	110*	112	33	(Arg)	13	20	15
16	Lys	110	112*	45	34	(Lys)	11	13*	12
17	Ala	12	74*	35	35	(Glu)[a]	9	13	13
18	Val	7	55*	48	36	(Ser)	10	13*	15
19	Thr	17	57*	32	37	(Tyr)	1	5*	

Table 2.21.

Gallus domesticus - CN-2.

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1	Asn	-	400*	0	14	Glu	9	152*	56
	Asp	-	160*	43	15	Ala	25	140*	73
2	Ser	0	30*	3	16	Ser	10	44*	7
3	Phe	0	492*	70	17	Arg	5	30	10
4	Val	2	248*	36	18	Leu	7	47*	42
5	Asn	0	454*	79	19	Ala	12	62*	30
	Asp	0	61*	198	20	His	14	30	28
6	Asp	61	198*	47	21	Tyr	2	42*	30
	Asn	454	79*	3	22	Asn	0	21*	15
7	Ile	6	118*	28		Asp	16	6*	8
8	Phe	8	428*	98	23	Lys	2	30*	10
9	Glu	7	159*	57	24	Arg	2	27	15
10	Arg	6	91	11	25	(Ser)	5	6*	5
11	Ile	4	166*	66	26	Thr	2	17*	23
12	Ala	5	142*	59	27	Ile	7	26*	
13	Gly	14	268*	133					

Table 2.22.

Gallus domesticus MT-2 (S-PITC).

Residue No.	Residue	n moles			Residue No.	Residue	n moles		
		R-1	R	R+1			R-1	R	R+1
1	Leu	-	22*	89	16	Thr	6	10*	8
2	Leu	22	89*	110	17	Lys	3	5	-
3	Leu	89	110*	18	18	Ala	24	39*	35
4	Pro	0	48*	12	19	Val	6	22*	
5	Gly	0	57*	6					
6	Glu	8	24	-					
7	Leu	10	48*	25					
8	Ala	6	58*	29					
9	Lys	3	16	14					
10	His	1	5	-					
11	Ala	14	48*	41					
12	Val	13	35*	25					
13	Ser	4	22*	11					
14	Glu	0	5	2					
15	Gly	11	41*	30					

2.4.3 The partial primary structure of Crocodylus niloticus (crocodile) H2B.

Approximately 7,5 mg of pure, unmodified crocodile H2B were automatically sequenced for 37 cycles (Table 2.23). The first 29 residues were unambiguously assigned and the remaining residues were only tentatively assigned.

The alignment of CN-2 was made on the same basis as chicken CN-2 (2.4.2). Pure, unmodified CN-2 was subjected to 36 degradation cycles (Table 2.24).

Carboxypeptidase digestion of crocodile H2B was done as for chicken H2B (2.4.2) and the carboxyl-terminus determined was lysine.

The partial primary structure obtained from sequencing CN-2 and the intact protein is given (Fig. 2.40).

```
                                        5                                  10
Calf        NH2-Pro - Glu - Pro - Ala - Lys - Ser - Ala - Pro - Ala - Pro -
Crocodile   NH2-Pro - Glu - Pro - Ala - Lys - Ser - Ala - Pro - Ala - Pro -

                                       15                                  20
Calf            Lys - Lys - Gly - Ser - Lys - Lys - Ala - Val - Thr - Lys -
Crocodile       Lys - Lys - Gly - Ser - Lys - Lys - Ala - Val - Thr - Lys -

                                       25                                  30
Calf            Ala - Gln - Lys - Lys - Asp - Gly - Lys - Lys - Arg - Lys -
Crocodile       Thr - Gln - Lys - Lys - Gly - Asp - Lys - Lys - Arg - Arg -

                                       35
Calf            Arg - Ser - Arg - Lys - Glu - Ser - Tyr -
Crocodile      (Lys)- Ser - Arg -(Lys)-(Glu)-(Ser)- Tyr -

                                    ------

                  62                   66                                  71
Calf            Met - Asn - Ser - Phe - Val - Asn - Asp - Ile - Phe - Glu -
Crocodile       Met - Asn - Ser - Phe - Val - Asn - Asp - Ile - Phe - Glu -

                                       76                                  81
Calf            Arg - Ile - Ala - Gly - Glu - Ala - Ser - Arg - Leu - Ala -
Crocodile       Arg - Ile - Ala - Gly - Glu - Ala - Ser - Arg - Leu - Ala -

                                       86                                  91
Calf            His - Tyr - Asn - Lys - Arg - Ser - Thr - Ile - Thr - Ser -
Crocodile       His - Tyr - Asn - Lys - Arg -(Ser)- Thr - Ile -(Thr)-(Ser)-

                                       96
Calf            Arg - Glu - Ile - Gln - Thr - Ala - Val -
Crocodile      (Arg)- Glu - Ile - Gln -(Thr)- Ala - Val -

                                    ------
```

Table 2.23.

Crocodylus niloticus - Intact Protein.

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1	Pro	-	478*	67	21	Thr	78	106*	60
2	Glu	-	413*	117	22	Gln	0	65*	37
3	Pro	67	388*	53		Glu	31	60*	35
4	Ala	33	533*	96	23	Lys	79	113*	123
5	Lys	15	202*	100	24	Lys	113	123*	74
6	Ser	14	436*	56	25	Gly	24	82*	54
7	Ala	49	499*	87	26	Asp	8	53*	34
8	Pro	22	246*	46		Asn	56	30*	37
9	Ala	87	436*	95	27	Lys	72	99*	106
10	Pro	46	159*	54	28	Lys	99	106*	29
11	Lys	46	192*	144	29	Arg	18	34	33
12	Lys	192	144*	55	30	Arg	34	33	26
13	Gly	81	426*	124	31	(Lys)	29	30*	19
14	Ser	47	146*	45	32	Ser	29	43*	49
15	Lys	54	103*	123	33	Arg	16	27	20
16	Lys	103	123*	49	34	(Lys)	18	22*	18
17	Ala	33	139*	63	35	(Glu)	14	18*	19
18	Val	15	96*	49	36	(Ser)	34	39*	47
19	Thr	41	148*	78	37	Tyr	5	31*	
20	Lys	42	133*	79					

Table 2.24.

Crocodylus niloticus - CN-2.

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1	Asn	-	210*	21	18	Leu	3	50*	20
	Asp	-	12*	0	19	Ala	10	53*	31
2	Ser	0	131*	10	20	His	7	23	8
3	Phe	0	222*	15	21	Tyr	0	72*	60
4	Val	2	130*	16	22	Asn	3	34*	21
5	Asn	15	140*	28		Asp	1	4*	4
	Asp	0	15*	75	23	Lys	1	10*	11
6	Asp	15	75*	12	24	Arg	1	13	6
	Asn	140	28*	0	25	(Ser)	10	13*	12
7	Ile	2	103*	27	26	Thr	2	8*	8
8	Phe	1	134*	24	27	Ile	4	22*	19
9	Glu	6	71*	15	28	(Thr)	8	11*	8
10	Arg	0	45	9	29	(Ser)	13	13*	13
11	Ile	8	79*	25	30	(Arg)	0,5	1,3	-
12	Ala	2	99*	19	31	Glu	3	6*	10
13	Gly	15	135*	30	32	Ile	8	14*	15
14	Glu	5	44*	13	33	Gln	4	15*	16
15	Ala	8	52*	23	34	(Thr)	3	5*	4
16	Ser	10	27*	18	35	Ala	4	7*	9
17	Arg	2	25	5	36	Val	2	6*	7
					37	Ser	6	13*	

	125
Calf	Lys - COOH
Crocodile	Lys - COOH

Fig. 2.40.

Comparison of the partial primary structure of H2B histone from erythrocytes of Crocodylus niloticus to that of calf thymus H2B. Residues are numbered according to alignment positions assuming a comparable total length of the polypeptide chains.

2.4.4 The partial primary structure of Xenopus laevis H2B.

Approximately 11 mg of pure unmodified Xenopus H2B were sequenced for 34 cycles. The first 30 residues were unambiguously assigned but the remaining four assignments were tentative (Table 2.25).

Pure, unmodified CN-2 was aligned on the same basis as chicken CN-2 and subjected to 27 degradation cycles (Table 2.26).

Carboxypeptidase digestion of Xenopus H2B was done as for chicken and the carboxyl terminus determined was lysine.

The partial primary structure obtained from sequencing CN-2 and intact protein is given (Fig. 2.41).

					5					10
Calf	NH_2-Pro - Glu - Pro - Ala -	Lys	-	Ser	-	Ala	-	Pro	-	Ala	- Pro -
Xenopus	NH_2-Pro - Glu - Pro - Ala -	Lys	-	Ser	-	Ala	-	Pro	-	Ala	- Pro -
					15					20	
Calf	Lys - Lys - Gly - Ser -	Lys	-	Lys	-	Ala	-	Val	-	Thr	- Lys -
Xenopus	Lys - Lys - Gly - Ser -	Lys	-	Lys	-	Ala	-	Val	-	Thr	- Lys -
					25					30	
Calf	Ala - Gln - Lys - Lys -	Asp	-	Gly	-	Lys	-	Lys	-	Arg	- Lys -
Xenopus	Thr - Gln - Lys - Lys -	Asp	-	Gly	-	Lys	-	Lys	-	Arg	- Arg -

					34
Calf	Arg - Ser - Arg -	Lys	-		
Xenopus	(Lys)-(Ser)-(Arg)-	(Lys)-			

					66					71
Calf	Met -	Asn -	Ser -	Phe -	Val -	Asn -	Asp -	Ile -	Phe -	Glu -
Xenopus	Met -	Asn -	Ser -	Phe -	Val -	Asn -	Asp -	Val -	Phe -	Glu -
					76					81
Calf	Arg -	Ile -	Ala -	Gly -	Glu -	Ala -	Ser -	Arg -	Leu -	Ala -
Xenopus	Arg -	Ile -	Ala -	Gly -	Glu -	Ala -	Ser -	Arg -	Leu -	Ala -
					86					
Calf	His -	Tyr -	Asn -	Lys -	Arg -	Ser -	Thr -	Ile -		
Xenopus	His -	Tyr -	Asn -	Lys -	(Arg)-	Ser -	Thr -	Ile -		

	125
Calf	Lys - COOH
Xenopus	Lys - COOH

Fig. 2.41.

Comparison of the partial primary structure of H2B histone from erythrocytes of Xenopus laevis to that of calf thymus H2B. Residues are numbered according to alignment positions assuming a comparable total length of the polypeptide chains.

Table 2.25.

Xenopus laevis - Intact Protein.

Residue No.	Residue	nmoles R-1	nmoles R	nmoles R+1	Residue No.	Residue	nmoles R-1	nmoles R	nmoles R+1
1	Pro	-	526*	83	20	Lys	28	128*	25
2	Glu	0	587*	145	21	Thr	33	94*	67
3	Pro	83	659*	55	22	Gln	2	24*	14
4	Ala	0	896*	65	23	Lys	37	192*	247
5	Lys	0	335*	36	24	Lys	192	247*	48
6	Ser	38	538*	69	25	Asp	7	39*	24
7	Ala	77	858*	87	26	Gly	10	45*	25
8	Pro	23	426*	46	27	Lys	23	59*	43
9	Ala	87	682*	239	28	Lys	59	43	30
10	Pro	46	162*	47	29	Arg	17	33	44
11	Lys	20	367*	325	30	Arg	33	44	29
12	Lys	367	325*	77	31	(Lys)	25	22	23
13	Gly	30	364*	105	32	(Ser)	21	28*	25
14	Ser	31	173*	36	33	(Arg)	21	30	29
15	Lys	47	291*	254	34	(Lys)	20	27	24
16	Lys	291	254*	82					
17	Ala	22	135*	51					
18	Val	14	99*	43					
19	Thr	18	147*	33					

Table 2.26.

Xenopus laevis - CN-2.

Residue No.	Residue	nmoles			Residue No.	Residue	nmoles		
		R-1	R	R+1			R-1	R	R+1
1	Asn	-	241*	44	16	Ser	0	92*	22
	Asp	-	49*	10	17	Arg	4	14*	5
2	Ser	0	150*	0	18	Leu	11	132*	38
3	Phe	0	480*	14	19	Ala	23	114*	50
4	Val	0	555*	24	20	His	0	4	5
5	Asn	26	241*	50	21	Tyr	2	13*	5
	Asp	0	39*	281	22	Asn	19	94*	35
6	Asp	39	281*	24	23	Lys	3	19	8
	Asn	241	50*	12	24	(Arg)	9	12*	6
7	Val	0	249*	24	25	Ser	0	8*	6
8	Phe	0	304*	24	26	Thr	13	27*	
9	Glu	21	291*	64	27	Ile	16	47*	
10	Arg	0	26*	5					
11	Ile	18	368*	56					
12	Ala	9	437*	41					
13	Gly	0	85*	12					
14	Glu	19	129*	82					
15	Ala	23	233*	110					

Chapter 3.

Discussion.

The once widely held view that histones are almost invariable has had to be changed. Whereas histones H3 and H4 still remain the most conservative proteins known, (Dayhoff 1972, 1973, 1976) the histones H2A and H2B are variable and H1 histones are highly variable.

3.1 Variability of histones H2B.

Many statements have been made concerning the number and kind of histones on the basis of electrophoretic studies (e.g. Panyim et al. 1971, Wright and Olins 1975, Destree et al. 1972). Generally the method of Panyim and Chalkley (1969) has been used, but modifications have given increased resolution : long gels (25 cm) (McMaster-Kaye 1973), 6,25 M urea (ibid), 4 M urea (Oliver and Chalkley, 1972) and nonionic detergents, e.g. Triton X-100 (Franklin and Zweidler, 1977). Reservations should be placed on conclusions drawn only from electrophoretic analyses because the mobility of a protein is a function of size, shape and charge. Furthermore, the protein may be modified (e.g. acetylation or phosphorylation) after synthesis. Positive identification cannot be made on the basis of electrophoretic data alone.

Amino acid and end group analysis will provide additional information as to the identity of a histone. However, a positive identification can only be made by primary structure determination.

Panyim et al. (1971) studied the variability of histone preparations from various tissues of several vertebrates and concluded that there was little electrophoretic variability. On the contrary, Zweidler (1976) estimated that 15-20 different species of H2B could be found in various mouse tissues. This heterogeneity of H2B could be due to sequence differences as reported in mouse ascites (Franklin and Zweidler, 1977) or due to charge modifications such as acetylation of lysine residues as reported

in Tetrahymena (Johmann and Gorovsky, 1976) and rainbow trout (Candido and Dixon, 1972) or phosphorylation of serine residues as in calf (Kuroda et al. 1976). Variation in the type and extent of these modifications could be responsible for the appearance of multiple H2B species on high resolution gel electrophoresis.

Two H2B histones were identified electrophoretically in histone preparations from rat-, monkey-, rabbit- and mouse-testis. The major fraction corresponded to somatic H2B in mobility whereas the minor fraction migrated slower. This new H2B designated TH2B, appeared at an early stage of spermatocyte differentiation. This TH2B from rat contained cysteine, less lysine, but more of both arginine and acidic amino acids (Shires et al. 1975 and 1976).

The histone H2B from the starfish Asterias rubens has a higher electrophoretic mobility than that from calf thymus (Vanhoutte-Durand et al. 1977) and has a higher content of glycine and arginine, similar to the sea-urchin sperm histones H2B (Strickland et al. 1977).

Histone H2B from late spermatids and spermatozoa of the cricket (Pallota and Tessier, 1975) also have a higher electrophoretic mobility than calf H2B probably due to its increased glycine and arginine but decreased lysine content.

Silk worm H2B (Yokotsuka et al. 1971) has a blocked N-terminus.

Tetrahymena contains two H2B species with distinct amino acid composition and electrophoretic mobilities (Johmann and Gorovsky, 1976).

Though no sequence data are available yet, it appears from electrophoretic studies (Spiker 1975 and 1976, Nadeau et al. 1977) amino acid composition (ibid.) and tryptic fingerprints (Hayashi et al. 1977) that the histone H2B protein(s) in plants are distinctly different.

A histone "H2B-like" protein migrating on electrophoresis between H1 and H3 has been found in Neurospora crassa (Goff, 1976). The classification of "H2B-like" was based on amino acid analysis even though

there was a lower arginine content. Similar to silk-worm histone H2B, the amino end is blocked.

An H2B isolated from rat chloroleukaemia cells (Martinage et al. 1976) differs slightly in primary structure and amino acid composition from calf H2B with Thr-Ala, Ser-Thr substitutions at positions 122 and 123. Similarly mouse Ascites cells (Franklin and Zweidler, 1977) contain two H2B histones, one of which has a Ser-Gly substitution at residue 75. The occurrence of a variant in calf thymus and other mammalian tissue with a Gln-Glu substitution at residue 76 has also been reported (ibid.).

The primary structure of histone H2B from trout testis (Kootstra and Bailey, 1976) shows many differences from calf thymus H2B (Fig. 3.2).

The fruit flies Drosophila (Oliver and Chalkley, 1972) and Ceratitis (Franco et al. 1974) have histone H2B which has a greater electrophoretic mobility than calf thymus H2B. The mobility of Drosophila H2B is identical in larval and adult tissues. However, McMaster-Kaye et al. (1973) showed that there is a major H2B (90%) and a minor H2B (10%) in Drosophila. The amino acid composition of Drosophila histone H2B (Oliver and Chalkley, 1972) is very different from calf thymus H2B in that the lysine content is lower with a concommitant increase in acidic amino acids. A partial sequence has been reported (Annual Review of Biochemistry 1975, p 729) which is distinctly different from calf H2B but similar to mollusc H2B reported in this investigation.

Although histones are found in sperm of sea-urchins (Strickland et al. 1977), in arthropods (McMaster-Kaye and Kaye, 1973 and Vaughn and Thomson, 1972), teleosts (Sung and Dixon, 1970 and Bols and Kasinsky, 1976), higher vertebrates and mammals, the histones in mature spermatozoa have been replaced by protamines (Coelingh et al. 1972 and Marushige et al. 1974). In Mollusca, the mature spermatozoa also contain proteins intermediate in composition between histones and protamines, but larger in size (Subirana et al. 1973). I also found a protein of this type in the mollusc Patella granatina (Table 2.1).

In this study, partial structures of chicken, crocodile and Xenopus erythrocyte H2B histones are reported as well as the complete sequence of the histone H2B from the gonads of the mollusc Patella granatina. The electrophoretic mobility of the H2B histones from the three vertebrate species is identical to that of calf, whereas the mobility of the mollusc H2B is higher (Fig. 2.15).

Amino acid composition of H2B histones from chicken, crocodile and Xenopus (Table 3.1) are very similar to that of calf thymus H2B. Approximately one half the amino acid residues have been assigned with only four variable positions (van Helden et al. 1978b and Fig. 3.1). All these mutations involve single base changes. This comparison of erythrocyte histones H2B is the first sequence study done which compares histones H2B from the same tissue of different species.

Although the amino acid composition of the mollusc H2B (Table 3.1) is not strikingly different from that of the vertebrate H2B, the sequence analysis of the first 33 residues of the intact protein demonstrated substantial differences in the primary structure. Mollusc H2B consists of 121 residues, whereas calf thymus H2B has 125 residues. If the carboxyl terminal 50 amino acids of the two H2B histones are compared, there are no differences (Fig. 3.1). Proceeding towards the amino terminal region, there are only six mainly conservative changes in the next 49 residues (Fig. 3.1). The amino terminal ends of the two H2B histones are almost entirely different. This pattern of variability is found in the structure of every other H2B histone that has been studied (sea-urchin - Strickland, 1977, trout - Kootstra 1976, Drosophila - Annual Review of Biochemistry, 1975). There is a close similarity in the N-terminal region between Drosophila and mollusc histone H2B.

In Fig. 3.2 the known histones H2B have been aligned for maximal homology. Sea-urchin $H2B_1$ and $H2B_2$ are unique in having a repeating pentapeptide structure. The pentapeptide begins with proline and ends

	1								10										20
Mollusc	Pro	-	Pro	-	Lys	Val	Ser	-	-	Ser	Lys	Gly	Ala	Lys	Lys	Ala	Gly	Lys	Ala	Lys
Calf	Pro	Glu	Pro	Ala	Lys	Ser	Ala	Pro	Ala	Pro	Lys	Lys	Gly	Ser	Lys	Lys	Ala	Val	Thr	Lys
Chicken	Pro	Glu	Pro	Ala	Lys	Ser	Ala	Pro	Ala	Pro	Lys	Lys	Gly	Ser	Lys	Lys	Ala	Val	Thr	Lys
Crocodile	Pro	Glu	Pro	Ala	Lys	Ser	Ala	Pro	Ala	Pro	Lys	Lys	Gly	Ser	Lys	Lys	Ala	Val	Thr	Lys
Xenopus	Pro	Glu	Pro	Ala	Lys	Ser	Ala	Pro	Ala	Pro	Lys	Lys	Gly	Ser	Lys	Lys	Ala	Val	Thr	Lys

	21								30										40
Mollusc	Ala	Ala	Arg	Ser	Gly	Asp	Lys	Lys	Arg	Lys	Arg	Arg	Arg	Lys	Glu	Ser	Tyr	Ser	Ile	Tyr
Calf	Ala	Gln	Lys	Lys	Asp	Gly	Lys	Lys	Arg	Lys	Arg	Ser	Arg	Lys	Glu	Ser	Tyr	Ser	Val	Tyr
Chicken	Thr	Gln	Lys	Lys	Gly	Asp	Lys	Lys	Arg(Arg	Lys	Ser	Arg	Lys	Glu	Ser	Tyr)			
Crocodile	Thr	Gln	Lys	Lys	Gly	Asp	Lys	Lys	Arg	Arg	Lys	Ser	Arg(Lys	Glu	Ser	Tyr)			
Xenopus	Thr	Gln	Lys	Lys	Asp	Gly	Lys	Lys	Arg	Arg(Lys	Ser	Arg	Lys)						

	41								50										60
Mollusc	Ile	Tyr	Lys	Val	Leu	Lys	Gln	Val	His	Pro	Asp	Thr	Gly	Val	Ser	Ser	Lys	Ala	Met	Ser
Calf	Val	Tyr	Lys	Val	Leu	Lys	Gln	Val	His	Pro	Asp	Thr	Gly	Ile	Ser	Ser	Lys	Ala	Met	Gly

	61								70										80
Mollusc	Ile	Met	Asn	Ser	Phe	Val	Asn	Asp	Ile	Phe	Glu	Arg	Ile	Ala	Ala	Glu	Ala	Ser	Arg	Leu
Calf	Ile	Met	Asn	Ser	Phe	Val	Asn	Asp	Ile	Phe	Glu	Arg	Ile	Ala	Gly	Glu	Ala	Ser	Arg	Leu
Chicken		Met	Asn	Ser	Phe	Val	Asn	Asp	Ile	Phe	Glu	Arg	Ile	Ala	Gly	Glu	Ala	Ser	Arg	Leu
Crocodile		Met	Asn	Ser	Phe	Val	Asn	Asp	Ile	Phe	Glu	Arg	Ile	Ala	Gly	Glu	Ala	Ser	Arg	Leu
Xenopus		Met	Asn	Ser	Phe	Val	Asn	Asp	Val	Phe	Glu	Arg	Ile	Ala	Gly	Glu	Ala	Ser	Arg	Leu

	81								90										100
Mollusc	Ala	His	Tyr	Asn	Lys	Arg	Ser	Thr	Ile	Thr	Ser	Arg	Glu	Ile	Gln	Thr	Ala	Val	Arg	Leu
Calf	Ala	His	Tyr	Asn	Lys	Arg	Ser	Thr	Ile	Thr	Ser	Arg	Glu	Ile	Gln	Thr	Ala	Val	Arg	Leu
Chicken	Ala	His	Tyr	Asn	Lys	Arg(Ser)Thr	Ile												Arg	Leu
Crocodile	Ala	His	Tyr	Asn	Lys	Arg(Ser)Thr	Ile(Thr	Ser	Arg)Glu	Ile	Gln(Thr)Ala	Val								
Xenopus	Ala	His	Tyr	Asn	Lys(Arg)Ser	Thr	Ile													

	101								110										120
Mollusc	Leu	Leu	Pro	Gly	Glu	Leu	Ala	Lys	His	Ala	Val	Ser	Glu	Gly	Thr	Lys	Ala	Val	Thr	Lys
Calf	Leu	Leu	Pro	Gly	Glu	Leu	Ala	Lys	His	Ala	Val	Ser	Glu	Gly	Thr	Lys	Ala	Val	Thr	Lys
Chicken	Leu	Leu	Pro	Gly	Glu	Leu	Ala	Lys	His	Ala	Val	Ser	Glu	Gly	Thr(Lys)Ala	Val				

	121				125
Mollusc	Tyr	Thr	Ser	Ser	Lys
Calf	Tyr	Thr	Ser	Ser	Lys
Chicken				Ser	Lys
Crocodile					Lys
Xenopus					Lys

Fig. 3.1. Comparison of histones H2B from; gonads of the mollusc (Patella granatina), erythrocytes of chicken (Gallus domesticus), crocodile (Crocodylus niloticus) and amphibian (Xenopus laevis), and calf thymus (Iwai et al. 1972).

Alignment position:																				
					5					10					15					20
Calf	Pro		Glu			Pro	Ala	Lys	Ser	Ala	Pro	Ala				Pro	Lys	Lys	Gly	Ser
$H2B_1$*	Pro	Ser	Gln	Lys	Ser	Pro	Thr	Lys	Arg	Ser	Pro	Thr	Lys	Arg	Ser	Pro	Thr	Lys	Arg	Ser
$H2B_2$*	Pro			Arg	Ser	Pro	Ala	Lys	Thr	Ser	Pro	Arg	Lys	Gly	Ser	Pro	Arg	Lys	Gly	Ser
Trout	Pro		Glu			Pro	Ala	Lys	Ser	Ala	Pro						Lys	Lys	Gly	Ser
Drosophila	Pro					Pro		Lys	Thr	Ala										
Mollusc	Pro					Pro		Lys	Val	Ser										
					25					30					35					40
Calf																		Lys	Lys	Ala
$H2B_1$*	Pro	Gln	Lys	Gly	Gly				Lys	Gly	Gly	Lys	Gly	Ala	Lys	Arg	Gly	Gly	Lys	Ala
$H2B_2$*	Pro	Ser	Arg	Lys	Ala	Ser	Pro	Lys	Arg	Gly	Gly	Lys	Gly	Ala	Lys	Arg	Ala	Gly	Lys	Gly
Trout																		Lys	Lys	Ala
Drosophila							Gly	Lys		Ala	Ala	Lys			Lys		Ala	Gly	Lys	Ala
Mollusc							Ser	Lys		Gly	Ala	Lys			Lys		Ala	Gly	Lys	Ala
					45					50					55					60
Calf					Val	Thr	Lys	Ala	Gln	Lys	Lys	Asp	Gly	Lys	Lys	Arg	Lys	Arg	Ser	Arg	Lys
$H2B_1$*					Gly	Lys	Arg	Arg	Arg	Gly	Val	Gln	Val	Lys	Arg	Arg	Arg	Arg	Arg	Arg	
$H2B_2$*					Gly	Arg	Arg	Arg	Arg		Val		Val	Lys	Arg	Arg	Arg	Arg	Arg	Arg	
Trout					Val	Thr	Lys	Thr	Ala	Gly	Lys	Gly	Gly	Lys	Lys	Arg	Lys	Arg	Ser	Arg	Lys
Drosophila	Glx	Lys	Asx	Ile	Thr	Lys	Asx	Lys	Lys												
Mollusc						Lys	Ala	Ala	Arg	Ser	Gly	Asp	Lys	Lys	Arg	Lys	Arg	Arg	Arg	Arg	Lys

Fig. 3.2. Comparison of the N-termini of histones H2B. Histones are aligned for maximum homology.

* Refers to the two H2B's from sea-urchin (Parechinus angulosus).

generally with serine. The other three positions are two basic residues and one helix breaker. No clear homology can be seen between mollusc histone H2B and the other H2B's for the first 23 alignment positions. From position 24 up to alignment position 52 the common origin of the various H2B histones becomes apparent. This common origin is reinforced by a block of basic amino acids (alignment positions 53-60) common to all the sequences. From alignment position 61 the sequences of all the reported H2B histones is very strongly conservative.

TABLE 3.1.

AMINO ACID COMPOSITION OF H2B.

Amino Acid	Mole %				
	Calf*	Chicken	Crocodile	Xenopus	Mollusc
Lys	16,0	19,16	16,34	15,89	15,44
His	2,42	2,58	2,58	2,50	2,23
Arg	6,79	6,55	6,78	6,40	8,88
Asp	4,80	4,63	4,84	5,12	5,23
Thr	6,39	6,74	6,08	6,33	5,90
Ser	11,15	7,87	9,14	9,21	10,07
Glu	7,98	8,54	8,02	8,05	7,26
Pro	4,80	4,59	4,51	5,16	3,21
Gly	5,60	5,72	6,27	5,16	5,16
Ala	10,36	10,03	10,16	11,27	11,68
Val	7,18	6,38	6,93	7,44	6,84
Met	1,63	1,58	1,51	1,60	1,73
Ile	4,80	5,36	5,91	4,88	5,59
Leu	4,80	4,74	4,84	4,98	4,65
Tyr	4,01	3,59	4,14	4,02	4,19
Phe	1,63	1,45	1,77	1,76	1,85

* Calculated from known sequence, Iwai et al. 1972.

No corrections for losses during hydrolysis or for incomplete cleavage.

3.2 Prediction of H2B conformation.

The prediction of secondary structure from amino acid sequence provides a useful tool for protein studies. Several methods to predict the secondary structure of proteins from sequence data have been developed. Argos et al. (1976) and Matthews (1975) have compared these methods and concluded that the Chou-Fasman (1974) method is relatively reliable. The latter is based on statistical probabilities of amino acids present in certain conformations in proteins investigated by X-ray crystallography. Argos et al. (1976) also showed that helix predictions were substantially more accurate than β-sheet predictions.

3.2.1 Prediction of H2B conformation by the Chou-Fasman method.

Chou and Fasman, on the basis of their statistical method, have made proposals concerning the conformation of histones (Fasman et al. 1976, 1977). Their proposal for histone H2B is given in fig. 3.9 which is based on probability profiles shown in Figs. 3.3 and 3.4 and values of table 3.2.

Table 3.3.

Conformational prediction of H2B for α-helix, β-sheet and β-turn regions: (Pα), (Pβ) and (Pt) values. Calf thymus H2B, (125) residues.

	Region	$(P\alpha)^a$	$(P\beta)^a$
α-helix	15-24 (10)	1,16	0,94
	69-82 (14)	1,15	0,96
	93-102 (10)	1,16	1,16 [b]
	105-113 (9)	1,23	0,86
β-sheet	39-48 (13)	1,03	1,36
	61-66 (6)	1,03	1,23
	88-90 (3)	0,91	1,33
	117-122 (6)	1,00	1,19

Table 3.2.

Conformational Parameters for α-Helical, β-Sheet and β-Turn Residues in 29 Proteins.

P_α		P_β		P_t	f_i	f_{i+1}	f_{i+2}	f_{i+3}
Glu 1.51		Val 1.70		Asn 1.56	Asn 0.161	Pro 0.301	Asn 0.191	Trp 0.167
Met 1.45	H_α	Ile 1.60	H_β	Gly 1.56	Cys 0.149	Ser 0.139	Gly 0.190	Gly 0.152
Ala 1.42		Tyr 1.47		Pro 1.52	Asp 0.147	Lys 0.115	Asp 0.179	Cys 0.128
Leu 1.21		Phe 1.38		Asp 1.46	His 0.140	Asp 0.110	Ser 0.125	Tyr 0.125
Lys 1.16		Trp 1.37		Ser 1.43	Ser 0.120	Thr 0.108	Cys 0.117	Ser 0.106
Phe 1.13		Leu 1.30		Cys 1.19	Pro 0.102	Arg 0.106	Tyr 0.114	Gln 0.098
Gln 1.11		Cys 1.19		Tyr 1.14	Gly 0.102	Gln 0.098	Arg 0.099	Lys 0.095
Trp 1.08	h_α	Thr 1.19	h_β	Lys 1.01	Thr 0.086	Gly 0.085	His 0.093	Asn 0.091
Ile 1.08		Gln 1.10		Gln 0.98	Tyr 0.082	Asn 0.083	Glu 0.077	Arg 0.085
Val 1.06		Met 1.05		Thr 0.96	Trp 0.077	Met 0.082	Lys 0.072	Asp 0.081
Asp 1.01		Arg 0.93		Trp 0.96	Gln 0.074	Ala 0.076	Thr 0.065	Thr 0.079
His 1.00	I_α	Asn 0.89		Arg 0.95	Arg 0.070	Tyr 0.065	Phe 0.065	Leu 0.070
Arg 0.98		His 0.87	i_β	His 0.95	Met 0.068	Glu 0.060	Trp 0.064	Pro 0.068
Thr 0.83		Ala 0.83		Glu 0.74	Val 0.062	Cys 0.053	Gln 0.037	Phe 0.065
Ser 0.77	i_α	Ser 0.75		Ala 0.66	Leu 0.061	Val 0.048	Leu 0.036	Glu 0.064
Cys 0.70		Gly 0.75		Met 0.60	Ala 0.060	His 0.047	Ala 0.035	Ala 0.058
Tyr 0.69		Lys 0.74	b_β	Phe 0.60	Phe 0.059	Phe 0.041	Pro 0.034	Ile 0.056
Asn 0.67	b_α	Pro 0.55		Leu 0.59	Glu 0.056	Ile 0.034	Val 0.028	Met 0.055
Pro 0.57		Asp 0.54	B_β	Val 0.50	Lys 0.055	Leu 0.025	Met 0.014	His 0.054
Gly 0.57	B_α	Glu 0.37		Ile 0.47	Ile 0.043	Trp 0.013	Ile 0.013	Val 0.053

P_α, P_β, P_t are conformational parameters of helical, β-sheet and β-turns. f_i, f_{i+1}, f_{i+2}, f_{i+3} are band frequencies in the four positions of the β-turn. H_α, H_β, etc., as defined previously (Chou and Fasman, 1974). Based on 408 turns.

β-bends - when $p_t > 0.75 \times 10^{-4}$
as well as $(P_t) > 1.00$
and $(P_\alpha) < (P_t) > (P_\beta)$

$$p_t = f_i \times f_{i+1} \times f_{i+2} \times f_{i+3}$$

Chou and Fasman, to be published.

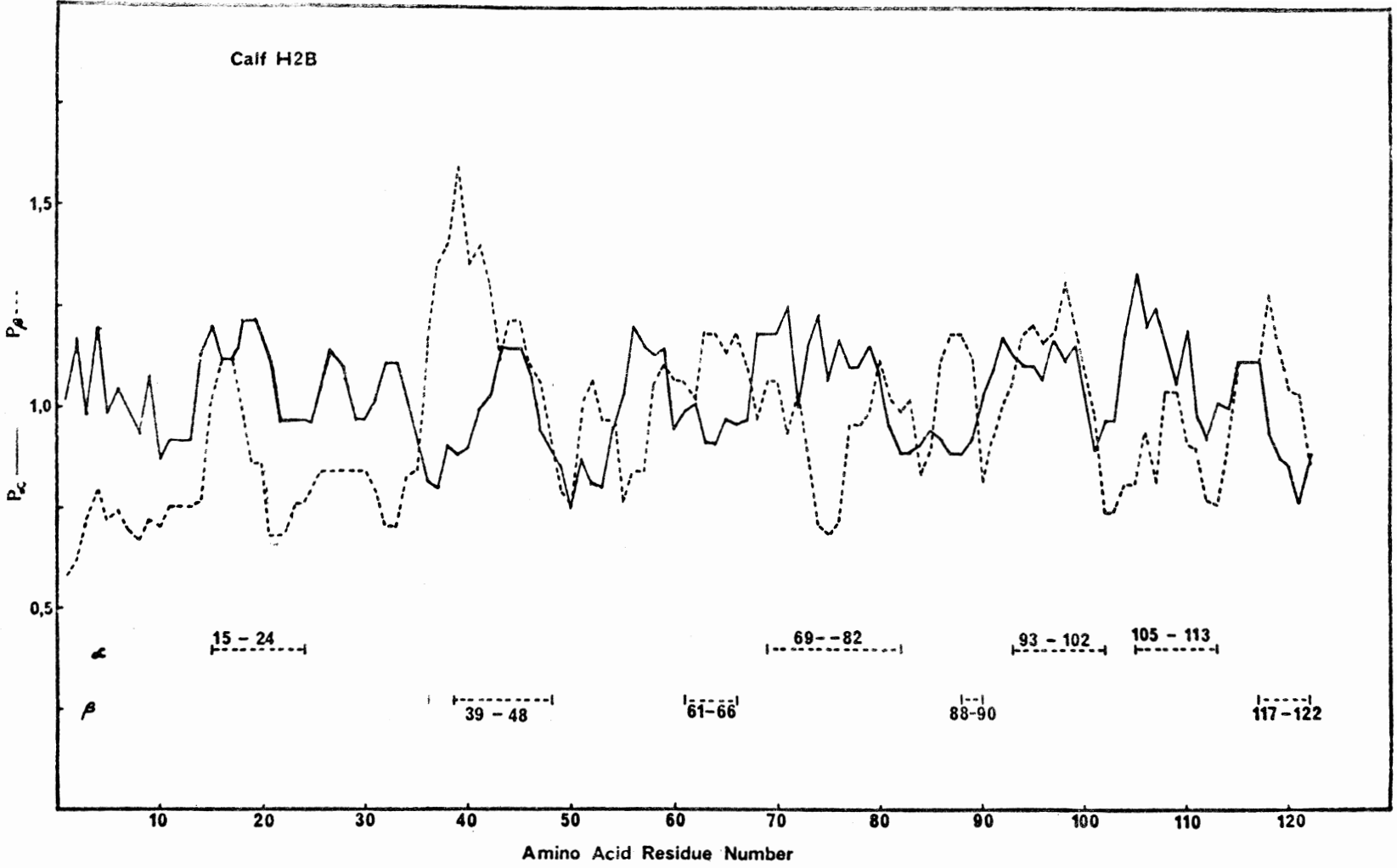


Fig. 3.3. The conformational profile of calf thymus H2B. The predicted regions are indicated according to a cut-off value of 1,0 (Fasman et al. 1977).

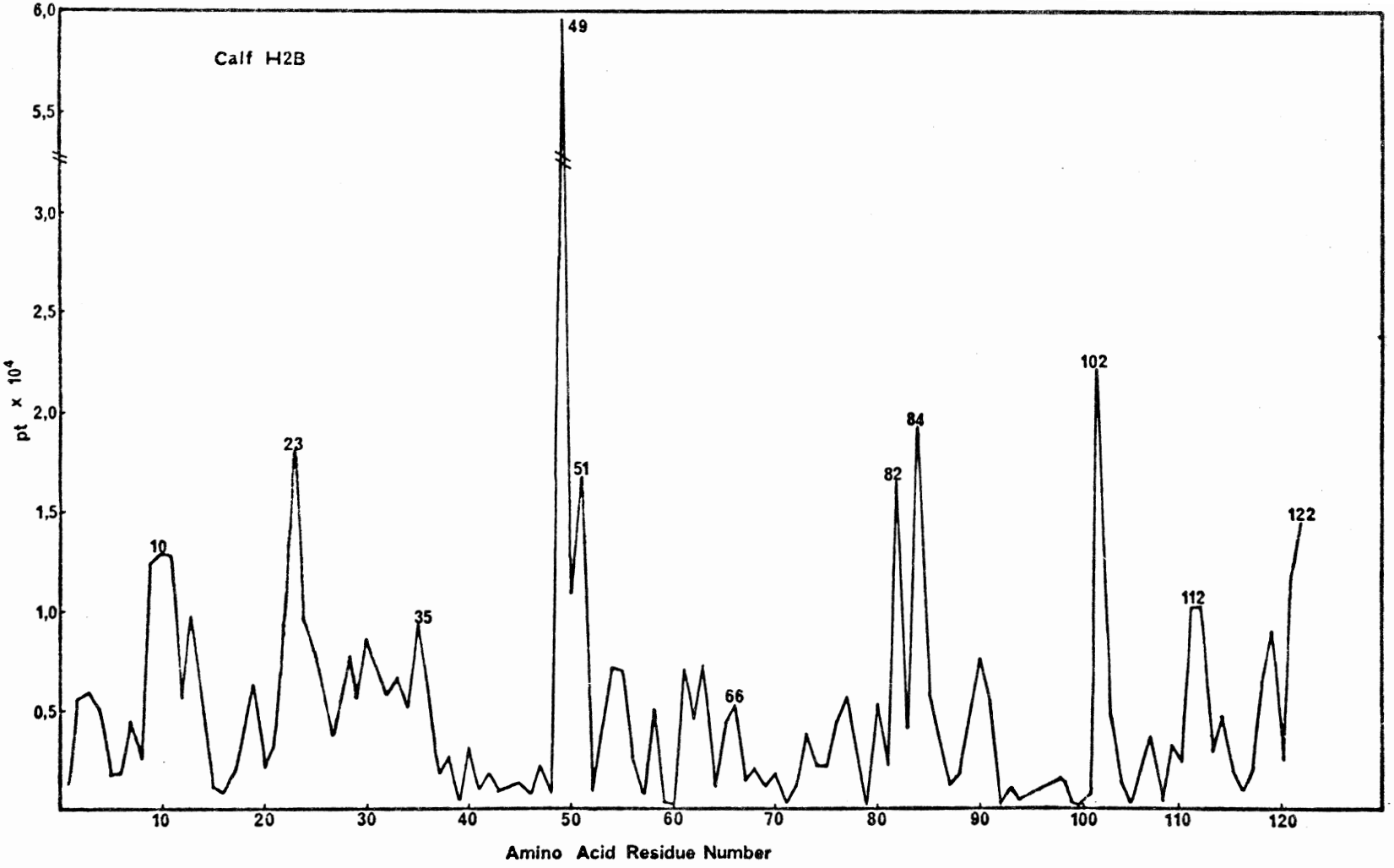


Fig. 3.4. Probability of tetrapeptide β-turns in H2B according to the method of Lewis et al. 1971.

β-turn	Tetrapeptide	$(P\alpha)^a$	$(P\beta)^a$	$(Pt)^{a,c}$
10-13	Pro-Lys-Lys-Gly	0,87	0,70	1,27
24-27	Lys-Asp-Gly-Lys	0,98	0,69	1,26
35-38	Glu-Ser-Tyr-Ser	0,94	0,84	1,18
49-52	His-Pro-Asp-Thr	0,85	0,79	1,22
66-69[d]	Val-Asn-Asp-Ile	0,96	1,18	1,00
84-87	Asn-Lys-Arg-Ser	0,90	0,83	1,24
102-105	Leu-Pro-Gly-Glu	0,97	0,74	1,10
122-125	Thr-Ser-Ser-Lys	0,88	0,86	1,21

[a] - From Fasman et al. 1977.

[b] - Predicted as α-helical, as analysis shows 5 strong helix formers but only 2 strong β-formers.

[c] - Probability profile, not shown, calculation as for 'a'.

[d] - Region 51-82 has no tetrapeptide with a high β-turn probability. The region 60-66 is a predicted β-sheet which changes to a predicted α-helix (69-82) and is separated by Asn-Asp. These two residues have a high predicted β-turn probability. Therefore the tetrapeptide 66-69 is assigned to a turn.

The profiles for helix and sheet areas of the sequenced regions of chicken H2B are very similar to those for calf thymus H2B (Fasman et al. 1977). Two differences are evident. The first difference is that in the α-helix 15-24 where (Pα) has dropped from 1,16 in calf to 1,11 in chicken H2B. This is due to the fact that residue 21 in calf is alanine (Hα) whereas threonine in chicken is iα. The second difference is a minor peak in the β-turn probability profile at residue 26 which is due to the Asp-Gly inversion from calf to chicken. (Figs. 3.5 and 3.6.).

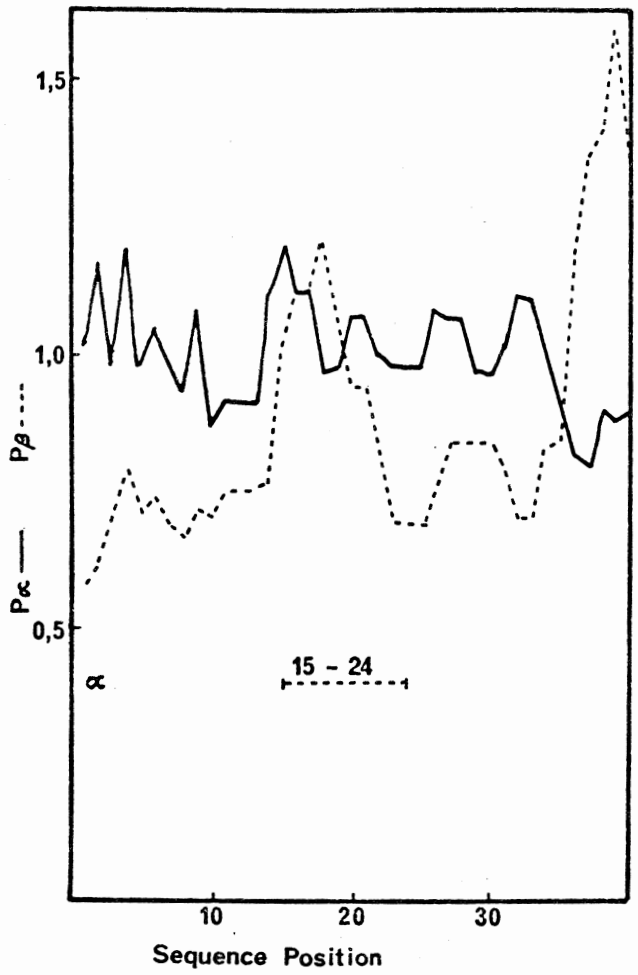


Fig. 3.5. The partial conformational profile of chicken erythrocyte H2B.

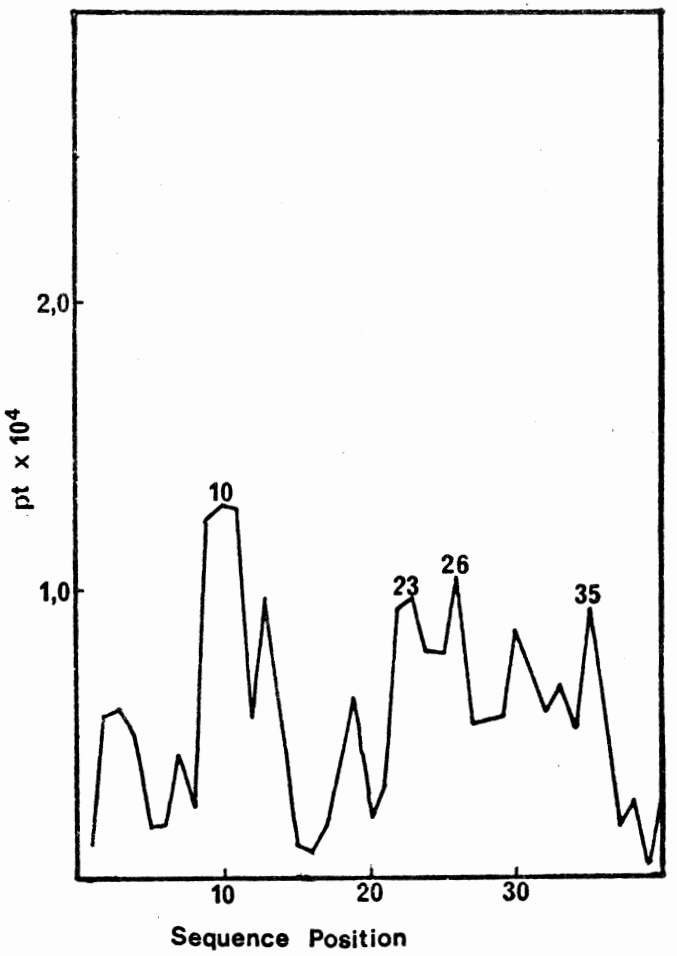


Fig. 3.6. Probability of tetrapeptide β-turns in chicken erythrocyte H2B.

The effect on conformation of the mutations is very small. The predicted conformations for chicken, crocodile and Xenopus are the same as that for calf thymus H2B except that the helix probability in region 15-24 is stronger in calf than the other H2B's.

Similar profiles to those for calf thymus H2B were drawn for mollusc H2B and the predicted secondary conformation was deduced and drawn in Fig. 3.9.

As expected the helical and β-sheet regions for mollusc and calf thymus are very similar. With respect to helix prediction, there are three areas of difference viz. region 1-20, 50-55 and 65-78 and with respect to β-turns there is one area of difference viz. region 20-23.

Mollusc H2B has a helix in the region 9-18 which is closer to the N-terminus than the first helix in chicken- and calf histone (15-24). Although the primary structure is entirely different, the length of these α-helices is constant. In addition, this helix in mollusc has a higher (Pα) than chicken; mollusc 1,23, chicken 1,11, calf 1,16.

In calf thymus H2B, residues 54-59 have not been assigned to an α-helical conformation by Fasman et al. 1977; however, it is feasible that this region may exist as a short helix in view of the (Pα) value, i.e. 1,11 and because 4 of the 6 residues are strong helix formers whereas 2 are indifferent and these two are surrounded by the strong formers. In addition, only 3,6 consecutive residues are required to form the hydrogen bonds of a single α-helical turn (Pauling et al. 1951). Consequently the equivalent area in mollusc H2B has been assigned to a helical conformation with the values as shown.

Table 3.4: Mollusc H2B.

Conformational prediction of H2B for α-helix, β-sheet and β-turn regions: (Pα), (Pβ) and (Pt) values. Mollusc H2B, 121 residues.

	Region	(Pα)[a]	(Pβ)[a]
α-helix	9-18 (10)	1,23	0,79
	50-55 (6)	1,11	0,97
	65-78 (14)	1,21	0,96
	89-98 (10)	1,16	1,16 [b]
	101-109 (9)	1,23	0,86
β-sheet	35-44 (13)	0,96	1,26
	57-62 (6)	1,03	1,23
	84-86 (3)	0,91	1,33
	113-118 (6)	1,00	1,19

β-turn	Tetrapeptide	(Pα)[a]	(Pβ)[a]	(Pt)[c]
5-8	Ser-Ser-Lys-Gly	0,82	0,75	1,36
20-23	Ser-Gly-Asp-Lys	0,88	0,69	1,36
31-34	Glu-Ser-Tyr-Ser	0,94	0,84	1,18
45-48	His-Pro-Asp-Thr	0,85	0,79	1,22
62-65[d]	Val-Asn-Asp-Ile	0,96	1,18	1,00
80-83	Asn-Lys-Arg-Ser	0,90	0,83	1,24
98-101	Leu-Pro-Gly-Glu	0,97	0,74	1,10
118-121	Thr-Ser-Ser-Lys	0,88	0,86	1,21

[a]- From figures in table 3.2.

[b,c,d]- See "Calf H2B".

The (Pα) for 65-78 in mollusc is higher than the (Pα) for the corresponding area in chicken and calf, as is the (Pt) for 20-23 in mollusc. The table below assigns to each region a particular conformation;

Table 3.5.

Region	Conformation
1-4	Coil
5-8	Turn
9-18	Helix
20-23	Turn
24-30	Coil
31-34 [a]	Turn
32-44 [a]	Sheet
45-48	Turn
50-55	Helix
57-62	Sheet
62-65	Turn
65-78	Helix
80-83	Turn
84-86	Sheet
89-98	Helix
98-101	Turn
101-109	Helix
113-118	Sheet
118-121	Turn

The above table leads to the diagrammatic representation of the two dimensional conformation of mollusc H2B (Fig. 3.9).

Certain areas may undergo transitions from α-helices to β-sheets, e.g. region 93-102 in calf H2B (Pα = Pβ = 1,16) has initially been assigned to a helix due to the greater number of helix formers. In addition β-sheet destabilization may occur due to charge repulsion. In region 1-35 (calf and mollusc), there is a concentration of positive charges and in region 69-76 there are two negative charges (acidics). The stable helices in

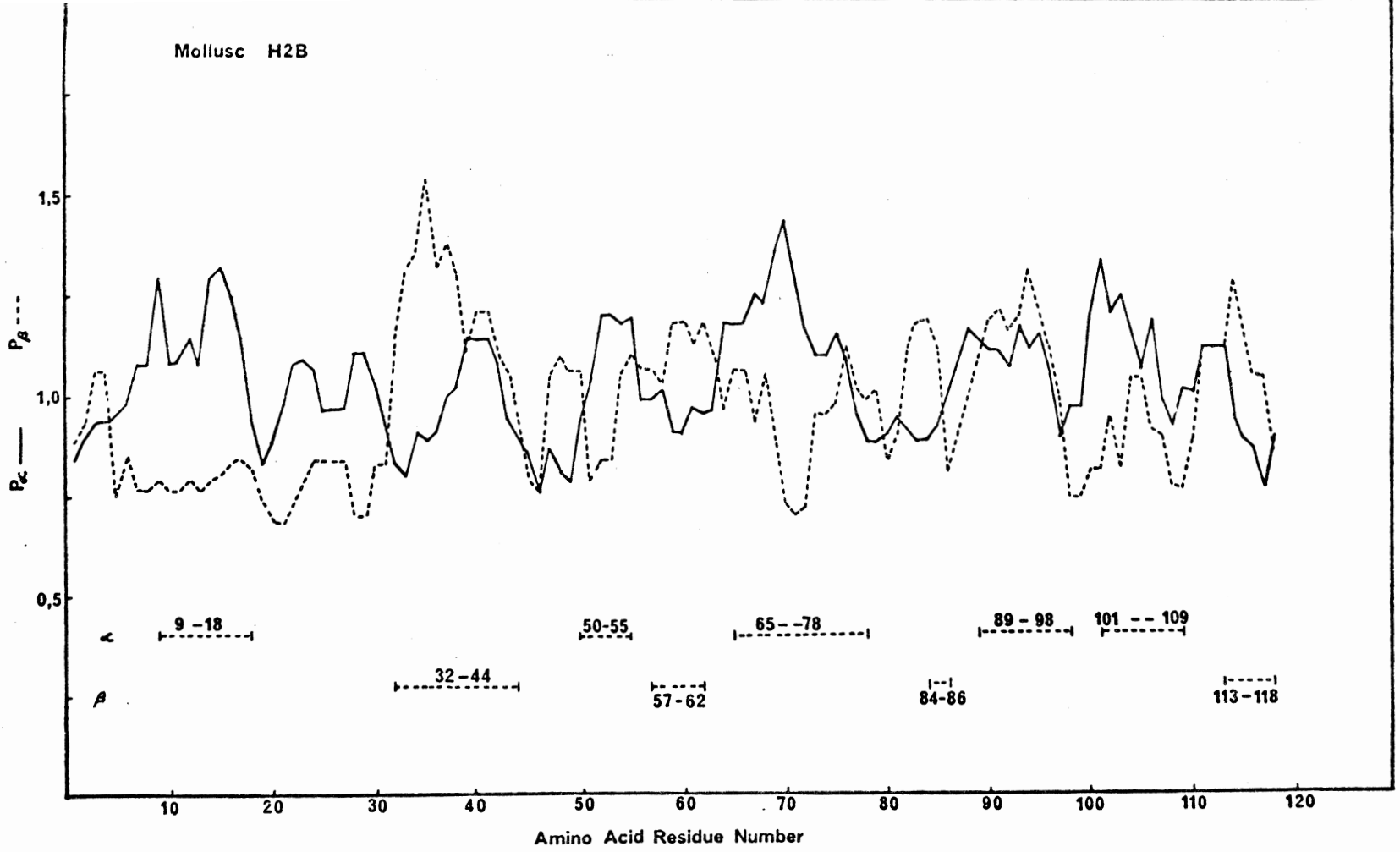


Fig. 3.7. The conformational profile of H2B patella. The predicted regions are indicated according to a cut-off value of 1,0.

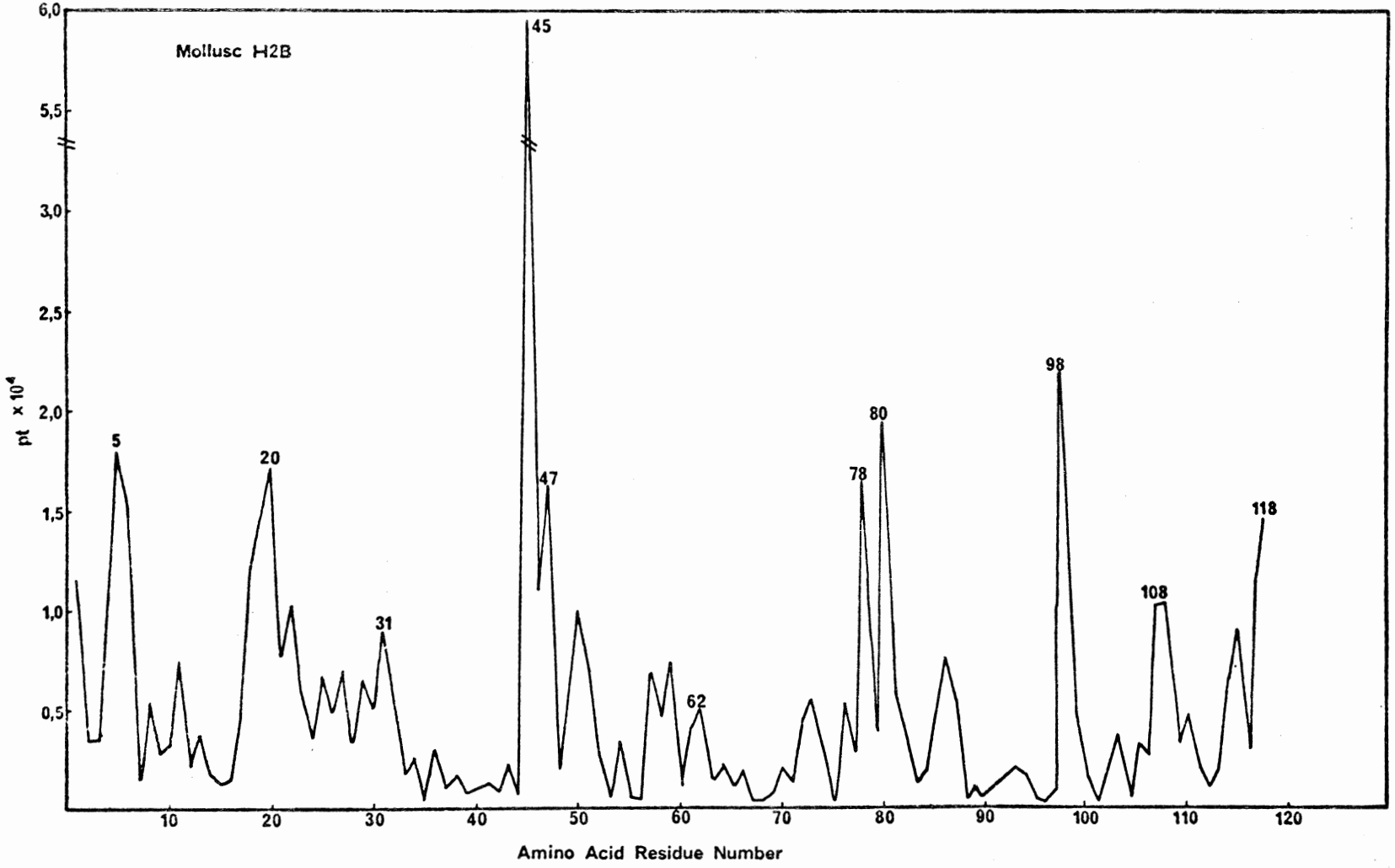


Fig. 3.8. Probability of tetrapeptide β-turns in H2B patella.

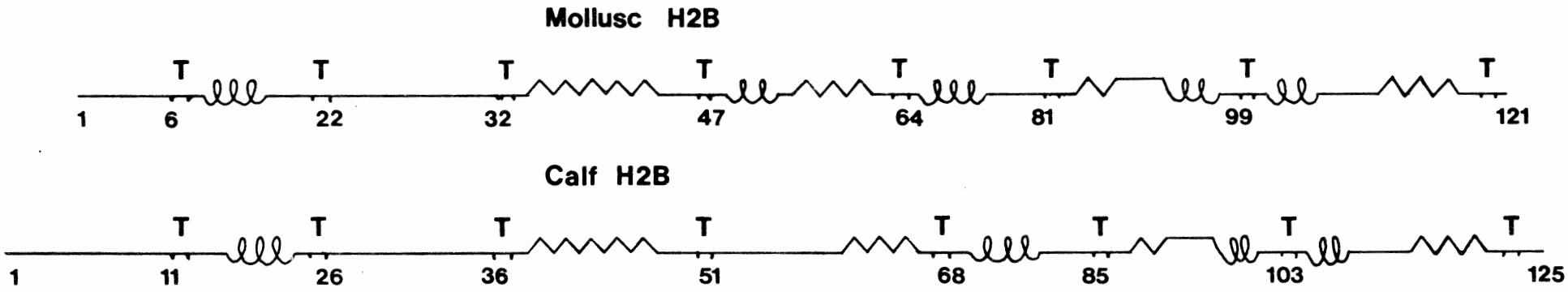


Fig. 3.9. The predicted secondary structure of H2B, indicating α-helices, β-sheets and random coil areas (——). Positions of β-turns are indicated by 'T'.

calf would then be 76-82, 93-102, 105-113, thus reducing the helical content from 39,2% to 20,8% (Lewis 1974, Fasman, 1977).

The maximum and minimum helix and sheet content for H2B based on such considerations is given in table 3.6.

Table 3.6.

Histone	% Helix		% Sheet	
	max. [a]	min. [b]	max. [c]	min. [a]
Calf H2B	39,2	20,8	28,0	20,0
Mollusc H2B	40,5	26,4	34,7	26,4

[a] - Calculated from predicted regions of histones.

[b] - Stable helices, not affected by charge repulsion.

[c] - Predicted sheet, plus helix-sheet transition.

The predicted secondary structure is that of the isolated protein and therefore these values may not be true estimates of the conformation *in situ*; furthermore the method only takes into account local interaction and not long-range interactions. The *in situ* conformation of the histone will be affected by protein-protein and protein-DNA interactions, which will be discussed briefly.

It is well established that clustering of certain types of amino acids occurs in histones. A survey of mollusc H2B illustrates this (Table 3.7.).

Table 3.7.

Amino Acid Class *	Residues 1-30	31-53	54-73	74-89	90-107	108-121
Basic	15	4	1	5	3	3
Acidic	1	2	3	1	1	1
Uncharged Polar	6	10	4	7	3	8
Hydrophobic	8	7	12	3	11	2

* Amino acids are classed according to Lehninger (1972).

The N-termini of all known H2B histones are very basic and show evidence of repetitive structures. These repetitive structures are most clearly defined in the two sperm H2B histones from sea-urchin (Strickland et al. 1977) where the N-terminal regions exhibit a repeating pentapeptide structure. The pentapeptide is variable and usually consists of proline at position 1 followed by 2 basic residues and 2 other residues which are normally polar.

At least three types of repetitive sequences may be found in protein (McLachlan, 1977):

(1) Sequence copy repeats, where a succession of amino acids is found repeated in a similar manner (such as the sea-urchin pentapeptides).

(2) Regular type-repeats, where amino acids of one class occur at regular intervals which are coherent throughout the length of the molecule.

(3) Irregular type-repeats, where amino acids of one class occur at roughly regular intervals and are usually caused by adaptation to a regular structural pattern.

A prominent feature of H2B histones from calf, chicken, crocodile, Xenopus and trout is the proline content in the first ten residues. A repeat is evident in the form of $\overset{1}{\text{Pro}}\text{-X-}\overset{2}{\text{Pro}}\overset{3}{\text{-X}}$ and $\overset{8}{\text{Pro}}\text{-X-}\overset{9}{\text{Pro}}\overset{10}{\text{-X}}$ except that -X-Pro is missing in trout H2B.

Circular dichroism of $(\text{Lys-Ala-Pro})_n$ under conditions of varying ionic strength and pH (Cernosek et al. 1974) showed that this peptide has no ionic strength dependence of higher order structure and does not appear to form any helix. This is probably due to the regular presence of Pro, which is helix destabilizing.

This leads to the postulate that the proline-containing N-termini of H2B may contain a number of 130° turns which may impart a peculiar conformation to the H2B's.

The N-termini of the histones H2B have two other features which need to be considered.

Within the first 20 residues mollusc H2B is rich in lysine, alanine and glycine:

```
          9
Ala-Lys-Lys-Ala-Gly-Lys-Ala-Lys-Ala-Ala
```

There is also a small palindrome, i.e.

```
                  10  11
Lys-Gly-Ala-Lys-Lys-Ala-Gly-Lys
```

The presence of a palindrome is a unique feature of mollusc H2B.

Calf thymus histone H2B contains repetitive segment of four residues. Four out of the five repeats contain two adjacent lysine residues:

```
          11
Lys-Lys-Gly-Ser
          15
Lys-Lys-Ala-Val
          19
Thr-Lys-Ala-Gln
          23
Lys-Lys-Asp-Gly
          27
Lys-Lys-Arg-Lys
```

Number of lysine	4	5	0	1

The N-terminus of mollusc H2B similarly exhibits a repeat involving lysine, but not quite as markedly as calf H2B histone:

```
          3
Lys-Val-Ser-Ser
          7
Lys-Gly-Ala-Lys
          11
Lys-Ala-Gly-Lys
          15
Ala-Lys-Ala-Ala
```

Number of lysine	3	1	0	2

```
                    11  12   14  15      16  17
```

One other type of repeat occurring is Lys-Ala, Lys-Ala and Lys-Ala.

Polypeptides containing these amino acids have been synthesized and used in conformation studies (Snell, 1972 and Yaron, 1972). Two types have been used, viz. (1) random sequence polypeptides, (2) fixed sequence polypeptides usually made up of a repeating unit of a di- or tripeptide. These polypeptides can form secondary structures in solution but these investigations do not give hints as to the possible interaction of these peptides with DNA. The DNA however, certainly has a conformation inducing effect on polypeptide chains, e.g. a $(\text{lysine}^{68}, \text{leucine}^{32})_n$ copolymer may form an α-helix on associating with DNA (Ong, Snell and Fasman, 1976) due to the neutralization of charges on lysine residues by the phosphates of DNA. Such helices may then be capable of hydrophobic interactions causing condensation of DNA. In addition, random sequence polypeptides containing lysine, alanine and glycine in the presence of salt may assume an α-helical conformation and cause condensation of DNA to an ordered and asymmetric structure whereas repetitive $(\text{Lys-Ala-Gly})_n$ polypeptides form random coils (Schwarz and Fasman, 1977).

Residues 11-30 in calf thymus H2B and 3-18 in mollusc H2B are rich in lysine, glycine and alanine residues. These regions on interacting with DNA may well assume α-helical forms contrary to the previous predictions.

Residues 23-30 of mollusc H2B are all basic. A homologous basic block is found in all other H2B histones. Such a block of basic residues in a protein would be predicted to be in a random coil conformation due to electrostatic repulsions of like charges. This prediction is not always correct, however. The poly-arginine random coil structure of protamine is changed to a structure containing helical segments in contact with tRNA. These helical portions are sandwiched between two different tRNA molecules (Warrant, 1978). In view of this, it may well be that the basic region (27-34 calf and 23-30 mollusc) in H2B may exist in a helical conformation when associated with DNA, which it does (Kato and Iwai, 1977).

This was also confirmed by thermal denaturation studies with DNA and the two large CNBr cleavage products (Li and Bonner, 1971).

Reversible phosphorylation (Stevely and Stocken, 1966) of serine residues in the amino terminal region occurs at the following positions; Ser (6) (Dixon et al. 1975), Ser (14) (Williams, 1976), Ser (32) and Ser (36) (Yeaman, 1977). These phosphorylated residues are usually surrounded by basic amino acids or separated from basic amino acids by only one residue;

$$\overset{6}{\text{Pro-Ala-Lys-Ser(P)-Ala-Pro-Lys}}$$

$$\overset{14}{\text{Lys-Lys-Gly-Ser(P)-Lys-Lys}}$$

$$\overset{32}{\text{Arg-Ser(P)-Arg-Lys}}$$

$$\overset{36}{\text{Lys-Glu-Ser(P)-Tyr}}$$

Ser (36) is an exception and has an acidic amino acid as a neighbour. These phosphorylated sites are located in regions of low hydrophobicity.

Acetylation of lysine involves residues 5, 10, 13 and 18 (Dixon et al. 1975).

All these modifications occur in the amino terminal 36 residues, therefore this region must be exposed to enzyme action and not buried inside a globular core. These modifications are reversible and would therefore alter histone/DNA interactions.

The sequence of mollusc H2B differs only slightly from that of calf thymus H2B and the predicted secondary structure is very similar (Fig. 3.9.), hence the structure and orientation of this H2B with regard to DNA and the other histones is likely to be identical from residues 8-121 (mollusc) onwards with the only difference in the amino terminal first 5 residues and this similarity exists in all the other H2B histones.

It is possible to envisage interactions between the various segments of the predicted secondary structure of H2B. It is possible that a β-turn could allow interaction between the β-sheet from residue

34-44 and the α-helix plus β-sheet from 49-62, thereby forming a 'face' which could interact with H2A (Martinson et al. 1976, Weintraub, 1975). Similarly a 'face' for H4 interaction (ibid.) could form if the sheet plus helix of region 83-98 folded back on region 101-118.

3.3 Conclusion and functional significance of H2B.

H2B is a relatively stable histone with a very conservative C-terminal half. The majority of the mutations after the ubiquitous basic block are conservative, but the N-terminus is highly variable. In a tissue with little transcriptional activity i.e. nucleated erythrocytes, the changes are negligible, but at lower stages of evolution there is greater variability. H2B is not the histone primarily responsible for the formation of the nucleosome, but helps to stabilize this higher order structure.

The different parts of the histone allow for functional flexibility.

The variable N-terminus of the H2B could possibly play a part in determining the exact number of base-pairs of DNA per nucleosome core. The sequence conservation of residues 35-125 (approximately - ref. calf thymus H2B) is probably necessary to ensure a constant core formation with the other histones, particularly in view of the conservatism of the primary core-formers H3 and H4.

Hence the variability of the H2B structure in the region where the molecule interacts with DNA allows differentiation while the constancy of the core and C-terminal region allows the ubiquitous chromatin structure.

Chapter 4.

Materials and Methods.

4.1 Reagents.

Peptide and protein solutions were usually dissolved in urea to allow disaggregation prior to gel filtration. The urea used was reagent grade from Merck Chemicals.

Cyanate can convert the $-NH_2$ group of proteins to hydantoin. (Stark et al., 1960). Acidification of urea solutions before use will decompose any cyanate present to NH_3 and CO_2.

Enzymes; Staphylococcus aureus protease was obtained from Miles (36-900).

Thermolysin from Serva (36015).

Trypsin from Sigma (grade XI).

Carboxypeptidase A from Sigma (Bovine Pancreas).

Carboxypeptidase B from Boehringer-Mannheim (Porcine Pancreas).

In addition, the following were used;

Digitonin from BDH (reagent grade).

Dansyl chloride from Serva.

Maleic anhydride (Merck) was freshly recrystallized from chloroform prior to use.

Hydrochloric acid was redistilled twice to produce constant boiling HCl for amino acid analysis.

4.2 Extraction and Purification of Histones.

4.2.1 Calf Thymus.

Calf thymus histones have been extensively studied (Hnilica 1972) and the primary structure of calf H2B was elucidated by Iwai et al. in 1970, therefore calf H2B was isolated as a reference standard (Panyim et al. 1968, Bohm et al. 1973 and von Holt, 1977).

4.2.2 Gonads.

Gonads were obtained from male Patella granatina by removing the animal from its shell, slicing open the mantle skin and gently pushing the gonad off the other tissues. Gonads were collected during the spawning season and also during the resting season. In order to obtain spermatozoa, ripe gonads were placed in a beaker of sea-water and stirred once very gently with a glass rod. After allowing to stand at room temperature (22°C) for about 20 minutes, the supernatant sea-water, now cloudy with spermatozoa, was decanted and the spermatozoa centrifuged. Acid extractions were done on the spermatozoan pellet as well as on the intact gonads.

4.2.2.1 Preparation of nucleoprotein.

The gonads were homogenized in 3 volumes of saline citrate (SSC - 0,15 M NaCl - 0,015 M trisodium citrate) in a Dounce glass homogenizer with a loose pestle. The homogenate in all cases was spun down at 10000xg for 10 minutes. The supernatant (cloudy orange) was poured off and the pellet homogenized and centrifuged twice more after which the supernatant was a clear solution. The pellet was then homogenized in SSC with 0,2% Triton X100 and repelleted. The supernatant was orange coloured and the slimy, orange top layer of the pellet was carefully scraped off and discarded. This was repeated, followed by two more washes in 3 volumes SSC and then twice washed in 1,5 volumes of 95% (v/v) ice-cold ethanol. This gave a cream coloured leathery pellet of crude nucleoprotein which was used as starting material for acid extraction of histones. Sperm cells were processed by the same method.

4.2.2.2 Extraction of acid-soluble proteins.

Nucleoprotein was extracted twice with 5 volumes of 0,25 M HCl at 4°C. The nucleoprotein was homogenized with a Waring Blendor at full speed for 30 seconds. The suspension was allowed to stand for 15 minutes

and then centrifuged at 12000xg for 20 minutes. The supernatant (approx. 150 ml) was filtered through a number 4 sintered glass filter and then the extracted proteins were dialysed for two days against 3 changes of 20 l twice glass distilled water. The dialysed proteins were subsequently freeze dried.

4.2.2.3 Purification of H2B.

The acid extracted proteins of gonads contain a slow-moving protein with a mobility different from the known histones (Fig. 2.1). This protein was subsequently found to be a large basic protein of approximately 298 residues of which 51 mole % were arginine (Table 2.1).

Pilot exclusion chromatography showed that this basic protein apparently complexed to the other histones under the usual conditions of histone exclusion chromatography. It therefore became necessary to separate the basic protein prior to fractionation of histones. Arginine-rich proteins are retained on the Bio-Gel matrix (Bonilla, 1969 and Dixon, 1975) as a result of the ion-exchange property of polyacrylamide gels which may have undergone partial hydrolysis of their amide groups. It was therefore attempted to utilize this property of Bio-Gel in the purification of the mollusc histones.

In a batch process approximately 5,5 g of Biogel P100 was used to form a slurry of 100 ml at pH5. To this slurry, 100 mg of acid extracted crude chromosomal proteins were added and allowed to stand with occasional stirring, for 2 hours at 22°C. The solution was filtered and the gel washed with twice glass distilled water adjusted to pH5 with dilute HCl. The pH of the slurry was then reduced to 1,6 and the gel was filtered a second time and washed. After dialysis and freeze-drying the initial filtrate yielded a protein mixture of mainly histones from

which the basic protein contaminant had been removed and the second filtrate recovered after reducing the pH yielded mostly basic protein and only small amounts of histone. The total recovery of proteins was approximately 50%.

To improve this process, a column of Biogel P60 (2,5x30 cm) was set up. The gel was washed extensively with twice distilled water (pH 6,5) and the column loaded with 64 mg of total protein extract dissolved in 6 M urea pH 6,5. The proteins were eluted with a pH gradient (distilled water plus HCl). Two fractions were recovered (Fig. 2.2). The recovery (86%) was improved compared to the batch process. The Biogel was regenerated with dilute NaOH at pH8 and then adjusted with HCl to pH 6,5 before applying the next sample. Up to 200 mg of acid extracted proteins could be fractionated per column run. With continued usage of the Biogel, however, the efficiency of separation decreased and the histones became increasingly contaminated with the basic protein. Different batches of Biogel gave different separation profiles. Attempts to standardize the ion-exchange capacity of Bio-Gel were unsuccessful. The Biogel particles also undergo a considerable volume change, increasing in size as the pH increased from 1,6 to 8. This volume change necessitated column repacking after each run. The inconsistency and inconvenience of this method led to a search for a more reliable, convenient ion-exchange separation technique.

The suitability of the cation-exchanger carboxymethylcellulose was investigated. The sample of acid extracted proteins was applied to a column of CM-52 in 0,05 M sodium acetate/HCl pH 4,5 and a gradient from 0,2 to 0,8 M NaCl was applied in the above buffer. After the salt gradient, any remaining proteins were eluted by 1 M NaCl in 0,05 M sodium acetate/HCl, 6 M urea at pH 4,5.

All the histones were eluted with the salt gradient. Therefore in subsequent runs the procedure was simplified to a stepwise elution

procedure. The histones were eluted with 0,8 M NaCl and the basic protein was then eluted with 1 M NaCl 6 M urea in the same sodium acetate/HCl buffer.

The histone fraction was dialysed and freeze-dried, dissolved in 8 M cyanate free urea, then applied to a Bio-Gel P30 column and eluted in 0,05 M NaCl, 0,02 N HCl.

After the first gel filtration column run, the H2B fraction was still impure. After recycling the H2B fraction twice more through the same Bio-Gel column pure histone H2B was obtained. The H2B was judged to be pure on the criteria of end-group analysis, amino acid analysis and electrophoresis.

4.2.3 Erythrocytes.

4.2.3.1 Collection of blood.

In all cases fresh blood was collected in 10% (w/v) ice-cold trisodium citrate (100 ml/l of blood).

Blood from white Leghorn chickens was collected at the local abbatoirs. Yield was approximately 50 ml blood/chicken.

Crocodile blood was collected at a crocodile farm and was obtained by shooting the animal in the back of its head, rapidly turning it on its back and allowing the blood to drip from the wound. The yield was approximately 250 ml for a crocodile of one metre in length.

Blood from Xenopus was obtained by anaesthetising the animal, pinning it on its back to a board and opening the thorax with a pair of scissors leaving the heart undamaged. The board was then inverted over a beaker and a large incision was made through the heart with a pair of scissors. The blood was collected in a beaker. The yield was approximately 3 ml blood/animal.

4.2.3.2 Preparation of nucleoprotein and extraction of histones.

Blood was collected as described and kept on ice. All subsequent operations were executed at 4°C. The blood was strained through three thicknesses of muslin and the erythrocytes sedimented at 600xg for 15 minutes in a refrigerated centrifuge. The plasma was discarded. The cells were washed three times by suspending them in 5-8 volumes of 0,14 M NaCl - 0,01 M trisodium citrate (wash medium) followed by centrifugation as above.

The erythrocytes were pooled and suspended in two volumes of medium 1 (0,14 M NaCl - 0,01 M trisodium citrate - 0,01 M EDTA - 5 mM sodium bisulfite - Panyim et al. 1968). Digitonin (0,5 g/20 ml) was dissolved in medium 1 and the erythrocytes lysed by slowly adding digitonin solution, with stirring, to a concentration of 20 ml digitonin solution per litre of blood initially used. After one hour the solution was centrifuged at 10000xg for 10 minutes (Purkayastha and Neelin, 1966). The red supernatant was discarded. The sediment was again suspended in medium 1 and treated with digitonin as before. In the case of chicken and crocodile erythrocytes, a homogeneous yellowish sediment was obtained, which was washed 5-8 times in 5-8 volumes of medium 1 followed by centrifugation until the supernatant was colourless. The final sediment consists of nuclei uncontaminated by ghosts (phase contrast microscopy). In the case of Xenopus, lysis of erythrocytes occurs prior to digitonin treatment. Digitonin treatment completes lysis of erythrocytes but also results in the lysis of many nuclei. This results in a viscous solution of chromatin which is contaminated by haemoglobin. This mixture can be pelleted.

The nuclei (or chromatin) were homogenized in 1-2 volumes of medium 1 in a Waring Blendor at low speed for 2 minutes. This homogenate

was suspended in 5 volumes of wash medium followed by centrifugation at 10000xg for 10 minutes. This wash was repeated 4-10 times to remove nucleoplasmic proteins. The pellet was twice homogenized in 2 volumes of 95% (v/v) analytical grade ethanol (Johns 1964). The yellowish supernatant was discarded and the cream coloured leathery pellet of crude nucleoprotein was used for histone extraction.

The osmotically fragile Xenopus erythrocytes, resulting in haemoglobin-contaminated chromatin, required extra washes to remove haemoglobin.

Extraction of the chromatin was done in two ways:-

The first method was that of Johns (1964) with modifications. (Fig. 4.1). Chromatin was homogenized with the loose pestle of a Dounce glass homogenizer in 1 volume of distilled water and allowed to stand at $4^{o}C$ for one hour, after which an equal volume of 10% perchloric acid was added slowly with stirring. A flocculant precipitate formed as the solution was stirred for 20 minutes. The suspension was centrifuged at 14000xg for 15 minutes. The supernatant was filtered, dialysed and freeze-dried. The pellet was suspended with homogenization (loose Dounce pestle) in 5-8 volumes of ethanol-HCl (4 volumes ethanol : 1 volume 1,25 N HCl) together with 0,4 cm glass beads and shaken in the cold for 18 hours. The suspension was centrifuged at 12000xg for 20 minutes. The sediment was re-extracted as above for 20 minutes. The supernatants were separately filtered and the histones recovered after dialysis (4.2.2.1) by freeze-drying. The pellet was extracted twice in 5 volumes of 0,25 N HCl, the supernatant filtered, dialysed (4.2.2.1) and freeze-dried.

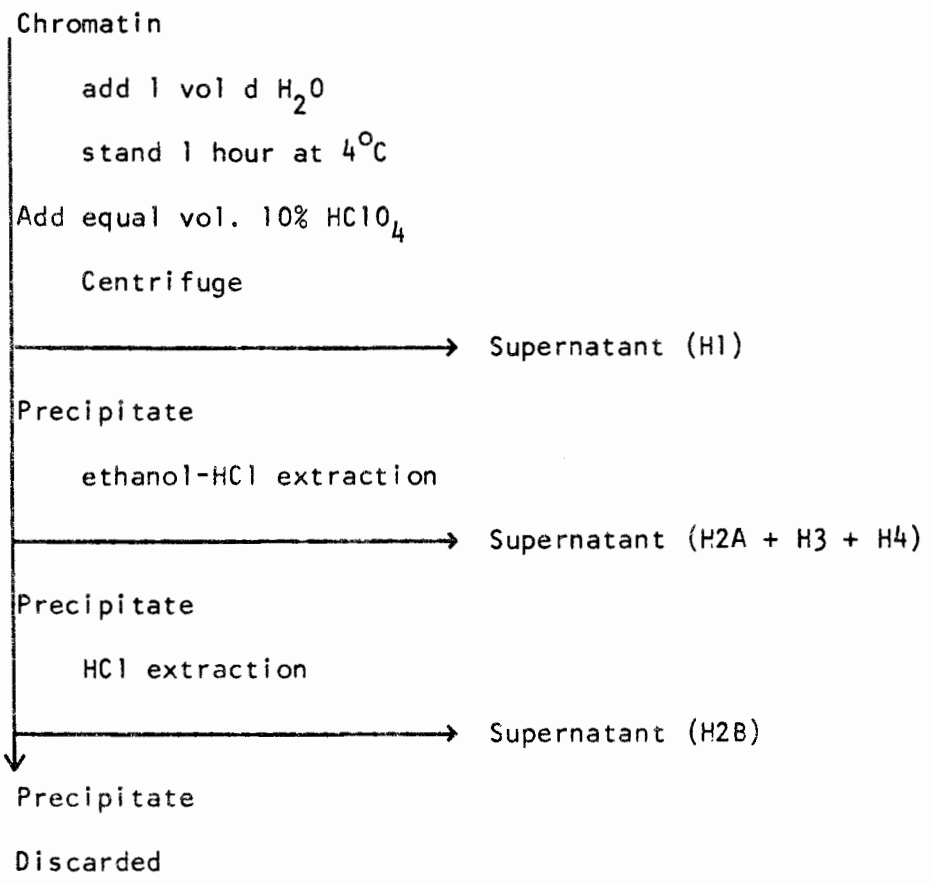


Fig. 4.1

The second method involved total histone extraction in 0,25 N HCl. These histones were partially purified by a modified method of Oliver et al. (1972), which is based on the procedures developed by Johns (1964). In a typical experiment (Fig. 4.2) 500 mg total histone were dissolved in 100 ml of 1 mM HCl at 4^oC. To this was added dropwise with stirring 7,08 ml 70% $HClO_4$ to make the final solution 0,88 M in $HClO_4$. Oliver et al. (1972) used a 0,5 M $HClO_4$ solution but it was found that this concentration was too low for optimal removal of histone H1. After 15 minutes the suspension was centrifuged at 20000xg for 20 minutes. The above was repeated using half the volumes. The supernatant contained H1. The pellet was homogenised for two periods of 30 seconds in an Ultraturrax homogeniser in 60 ml ethanol-HCl (4 volumes ethanol : 1 volume 1,25 N HCl) followed by centrifugation. The homogenization was repeated twice, the

volumes decreasing to 40 ml and 20 ml. These supernatants contained H2A, H3 and H4. The final pellet was dissolved in 1 mM HCl followed by dialysis and freeze-drying.

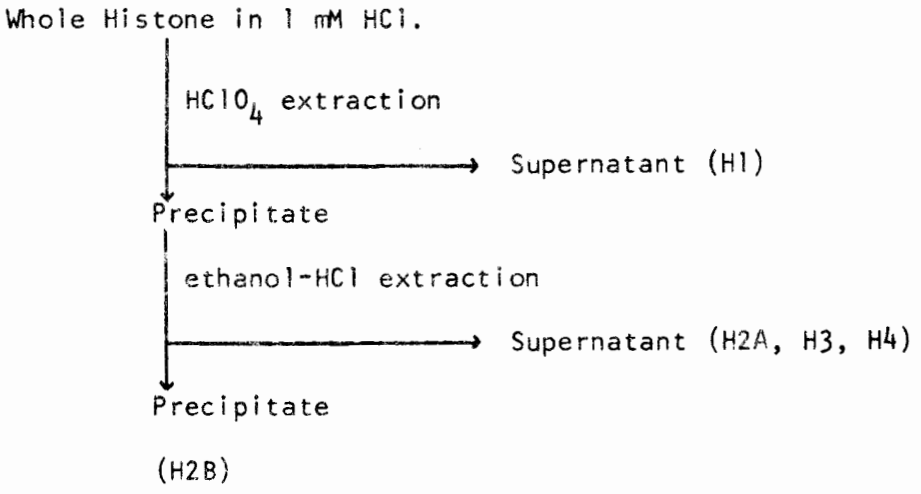


Fig. 4.2

Neither of these methods yielded pure H2B.

4.2.3.3 Purification of H2B's.

The impure H2B prepared by the previous methods was finally purified by gel filtration. Since histones have a tendency to aggregate all samples were dissolved in 8 M urea (4.1) prior to application to Bio-Gel or Sephadex columns. These samples were chromatographed on Bio-Gel P30 or P60 (van der Westhuyzen et al. 1974). The elution from two column types was similar. Crude chicken histone H2B contained an additional fraction corresponding to H5 and some Xenopus histone samples also had a small peak in that area. No evidence of such a peak was found in crocodile material.

The H2B obtained from gel filtration was still contaminated with small amounts of histone H3. Chicken histone H3 and those which have been studied from organisms at a lower stage of evolution contain one cysteine (Brandt, 1971 and 1974, Hooper et al. 1973). Chicken histone H3 readily forms disulfide bridges under oxidative conditions resulting in

the formation of a dimer. The molecular dimension of the H3 may thus be artificially increased enabling separation of H2B and H3 on gel chromatography (Brandt et al. 1971). Since the dimerization does not go to completion, the H2B fraction after gel chromatography has to be recycled through this process twice to result finally in a 96-98% pure protein.

Typically, the experiment was performed using an H2B fraction approximately 30% (w/w) contaminated by H3. This fraction (100 mg) was dissolved in 4 ml 6 M urea previously acidified to pH 2,5 (4.1). All operations were performed at 4°C. The protein solution was allowed to stand for one hour, after which 2 ml 0,5 M tris was added and the solution adjusted to pH 7 with 0,1 N NaOH. O-iodosobenzoic acid was weighed out in an equimolar amount to that of the estimated H3, dissolved in an equivalent amount of 0,1 N NaOH and 1 ml water. This solution was slowly added to the protein with stirring. The oxidation was allowed to proceed for one hour and then the solution was dialysed extensively for 18 hours against distilled water and finally freeze-dried. The resulting protein solution was applied to a Sephadex G100 column.

4.3 Production of Peptides.

Peptides were produced by chemical and enzymatic means.

4.3.1 Nomenclature of Peptides.

The nomenclature of peptides follows the scheme set out below. The letters indicate the type of cleavage, e.g. CN denotes cyanogen bromide cleavage, NBS denotes cleavage by N-bromosuccinimide, SA - by Staphylococcus aureus protease, TH - by thermolysin and MT by maleylation followed by tryptic cleavage. This is followed by a number which indicates the relative elution order of the peptide on chromatography. A Roman numeral indicates ion-exchange chromatography and an

Arabic numeral indicates exclusion chromatography. Subsequent cleavages and purifications were added to the name of the parent fragment, e.g. CN-2 NBS-1; this peptide was first produced by cyanogen bromide cleavage and then the pure peptide was eluted as the second peptide on gel chromatography. This peptide was cleaved by N-bromosuccinimide and purified as the first peptide from gel chromatography.

4.3.2 Chemical cleavages.

4.3.2.1 Cyanogen bromide.

Cyanogen bromide (CNBr) cleaves specifically at the C-terminal side of methionine and the derivatized methionine residues may be identified as homoserine or homoserine lactone. The generated carboxyl-terminal peptide is easily identified by its lack of these derivatives. The method followed is that of Gross (1967).

Typically, 100 mg H2B were dissolved in 15 ml 70% (v/v) formic acid to which 100 mg of CNBr were added with stirring. The solution was kept at room temperature (21^{o}C) under N_2 and in the dark. Further aliquots of 50 mg of CNBr were added at 4 hours and 24 hours. After 30 hours, the solution was either dried down at reduced pressure at 35^{o}C or diluted by 10 volumes of distilled water, frozen in liquid nitrogen and freeze dried. The dried material was dissolved in 8 M urea at pH 1,9 and applied to a Biogel P30 column.

4.3.2.2 N-Bromosuccinimide.

This reagent has been used to cleave peptides containing tryptophan, tyrosine and histidine residues (Ramachandran and Witkop, 1967).

Between pH 0-5 1 mole tyrosine reacts with 3 moles of N-bromosuccinimide (NBS) to result in the cleavage of the carboxyl-peptide bond of tyrosine. Tyrosine is converted to a dibromodienone spirolactone and a

peptide with a new N-terminal group is released.

The reaction may be monitored spectrophotometrically at 260 nm at which the dienone spirolactone shows characteristic absorption (ε = 10000-11000).

In the case of the histones, tryptophan is absent, so the tyrosyl-peptide bonds will be selectively cleaved. Histidyl bonds require high temperatures for cleavage, although the NBS reacts with the histidine residues concurrently with tyrosine.

Peptide CN-2 (3 mg) was dissolved in 3,3 ml 50% (v/v) acetic acid. Freshly recrystallised NBS (from 90% v/v acetic acid) was dissolved in 50% acetic acid (22 mg/ml) and added in 10 µl aliquots to the peptide solution with stirring at 22°C. The absorbance at 260 nm was recorded at 5 minute intervals using as a blank 50% acetic acid to which the same amount of NBS was added as in the peptide solution. Near the end point a faint, transient yellow colour, which persisted after the end point was passed, became visible. Addition of NBS was stopped after no further increase in 260 nm absorption was evident. The solution was diluted with 5 volumes of distilled water and freeze-dried.

4.3.3 Enzymatic cleavages.

4.3.3.1 Trypsin.

Trypsin exhibits a high degree of specificity, catalyzing the hydrolysis of the peptide bond between the carboxyl group of arginine and lysine and the amino group of the adjacent amino acid (Hirs et al. 1956). Due to the high mole % of basic amino acids in H2B, trypsin would produce a large number of peptides. However, the cleavage sites may be reduced by reversible modification of lysine. Maleic anhydride is one such reagent (Butler et al. 1969).

Histone H2B (30 mg) was dissolved in 1,5 ml 6 M urea and 2 ml

0,2 M $NaHCO_3$ pH 8,5 were added. Based on an estimate of 41 µmol lysine, a twentyfold excess of maleic anhydride was used to modify the ε-NH_2 groups of lysine. The maleic anhydride was added in small batches with vigorous stirring at 22°C over a period of two hours, during which time the pH was maintained above 8,0 by addition of 2 M NaOH (Butler et al. 1969). The solution was dialysed against distilled water and freeze-dried.

The maleylated H2B was dissolved in 3 ml H_2O and the pH adjusted to 8,4 with 10 mM NaOH. 0,3 mg trypsin was added and the pH maintained at 8,4 with 10 mM NaOH for 2 hours in a pH-stat at room temperature (22°C). A further aliquot of 0,1 mg of trypsin was added and the digestion allowed to proceed for a further 2 hours, after which the pH was lowered to 2,0 (by the addition of HCl) to terminate the digestion. The peptides produced were fractionated by column chromatography and the lysine residues were finally demaleylated.

Demaleylation of peptides was carried out in 4 ml 30% (v/v) acetic acid at 40°C for 48 hours, after which the solution was diluted to 5% acetic acid and freeze-dried.

4.3.3.2 Staphylococcus aureus protease.

Staphylococcus aureus protease specifically cleaves peptide bonds on the carboxyl side of aspartate or glutamate residues (Drapeau et al. 1972). The enzyme exhibited maximum proteolytic activity between pH 4,0-7,8 in sodium or potassium phosphate buffer. Later work showed that in ammonium bicarbonate or ammonium acetate only glutamoyl and Asp-Gly bonds were cleaved (Houmard et al. 1972). This suggested that the enzyme specificity was affected by the buffer used.

Typically, 7 mg CN-1 were dissolved in 0,8 ml 0,05 M ammonium acetate buffer pH 7,0. 0,24 mg protease was added and digestion was

allowed to proceed for 16 hours at 37°C. The pH was lowered to 2,0 by the addition of HCl and the peptides produced were fractionated by column chromatography.

4.3.3.3 Thermolysin.

Thermolysin is produced by Bacillus thermoproteolyticus. The enzyme exhibits a substrate specificity for peptide bonds involving the amino groups of hydrophobic amino acid residues (Matsubara et al. 1965 and 1966). Most frequently the bonds X-Leu, X-Ile, X-Phe and X-Val type are cleaved, but X-Ala cleavages also have been noted (Bradshaw, 1969).

The enzyme has a calcium dependence (Drucker, 1971) and a pH optimum near pH 8 (Stauffer, 1971).

The cleavage of H2B or CN-1 by thermolysin was done as follows; H2B (6 mg) was dissolved in 1 ml of distilled water to which 0,1 ml of 25 mM $CaCl_2$ was added. The pH was adjusted to 7,5 with 0,1 M NaOH. Thermolysin was dissolved in 25 mM $CaCl_2$ (2 mg enzyme/ml) and used at an enzyme substrate ratio of 1:50 (w/w). The pH was maintained in a pH-stat at room temperature for 1,5 hours after which the reaction was terminated by lowering the pH to 2,0 with HCl. The peptide mixture resulting was separated by column chromatography on Sephadex G15.

4.4 Peptide Characterization.

4.4.1 Thin-layer chromatography.

To test the purity of peptides with identical end groups or to test for the presence of peptides when the eluant used for column chromatography has a high U.V. absorbance (e.g. 30% acetic acid or 5% formic acid), the following technique proved useful.

From individual column fractions an aliquot (e.g. 5-10 µl) was spotted onto a cellulose thin-layer plate (Polygram cell 300 or 400, Macherey and Nagel, Duren)(Chen et al. 1975). This plate was developed

in the following buffer for two hours; 60 ml glacial acetic acid, 200 ml pyridine, 300 ml butanol, 240 ml twice distilled water. The plate was dried and sprayed with a ninhydrin solution; 3 g ninhydrin, 30 ml collidine, 100 ml glacial acetic acid, made up to 1000 ml with ethanol and kept cold in the dark. The plate was then placed in an oven or dried with hot air to develop spots.

4.4.2 Polyacrylamide gel electrophoresis.

This was done in 15% acrylamide gels containing 2,5 M urea, according to the method of Panyim and Chalkley (1969) using 10 cm gels. The following solutions were prepared:

A: 60% (w/v) acrylamide, 0,42 (w/v) N,N-methylenebisacrylamide.

B: 4% (v/v) N,N,N,N-tetramethylethylenediamine (TEMED) in 43,2% (v/v) glacial acetic acid (A.R. grade).

C: 0,2% (w/v) ammonium persulfate, 24% (w/v) urea.

All solutions were stored at $4^{o}C$, solution C being made up fresh before use. Gels were prepared by mixing A, B and C in the volume ratio 2:1:5. This solution was degassed under vacuum in a side-arm flask for 1 minute and quickly pipetted to within approximately 1 cm of the top into glass tubes (10x0,5 cm), which were sealed at the bottom by Parafilm (American Can Co.). Water was carefully layered over the gel and polymerisation was allowed for 1,5 hr. at room temperature. The parafilm was removed and the 10 cm gels pre-electrophoresed in a Shandon disc gel electrophoresis apparatus using 0,9 M acetic acid as tray buffers at 2 mA per tube until the voltage remained constant at approximately 150 V. This took about 4 hours.

The sample was dissolved at a protein concentration of 1 mg/ml in freshly prepared 8 M urea and allowed to stand for 1 hour prior to application.

After pre-electrophoresis, the tray buffer was changed, sample applied (5-10 µg per expected band) and electrophoresed for 3½ hours at a constant voltage of 150 V.

Gels were removed from tubes by forcing water between the gel and glass with a syringe and stained with 0,25% amido black (Merck) - 10% acetic acid for 1 hour and destained in a transverse disc destainer with 7% acetic acid. Gels were stored in 25% ethanol-7% acetic acid in corked tubes.

4.4.3 Amino acid analysis.

Approximately 20-100 nmoles of sample were hydrolysed in twice distilled constant boiling HCl with 0,025 % phenol. The samples were flushed twice with nitrogen and sealed under vacuum (0,02 mm Hg) prior to heating at 110°C for 24 hours. Analysis was done on Beckman amino acid analysers (model 116 or 119) in conjunction with a Hewlett-Packard Lab. Data System 3352B using the internal standards tryptophan, AGP (ε-amino-B-guanidino propionic acid) or norleucine. No corrections for hydrolytic losses were made.

4.4.4 N-terminal Group.

4.4.4.1 Dansylation.

The method used for labelling the N-terminal amino acids of proteins and peptides is essentially that of Gray (1972).

Protein method; Protein (5-25 nmoles) was dissolved in 50 µl of 1% (w/v) sodium-dodecyl-sulfate (SDS) in a 3 mmx60 mm Pyrex test tube. The solution was heated in a boiling water bath for 2-5 minutes to dissolve the protein and then cooled. N-ethylmorpholine (50 µl) was added and then 50 ul dansyl chloride (1-dimethylaminonaphthalene-5-sulfonyl chloride) freshly prepared in dimethylformamide or water-free acetone (50 mg/ml). The mixture was incubated at 45°C for 30 minutes

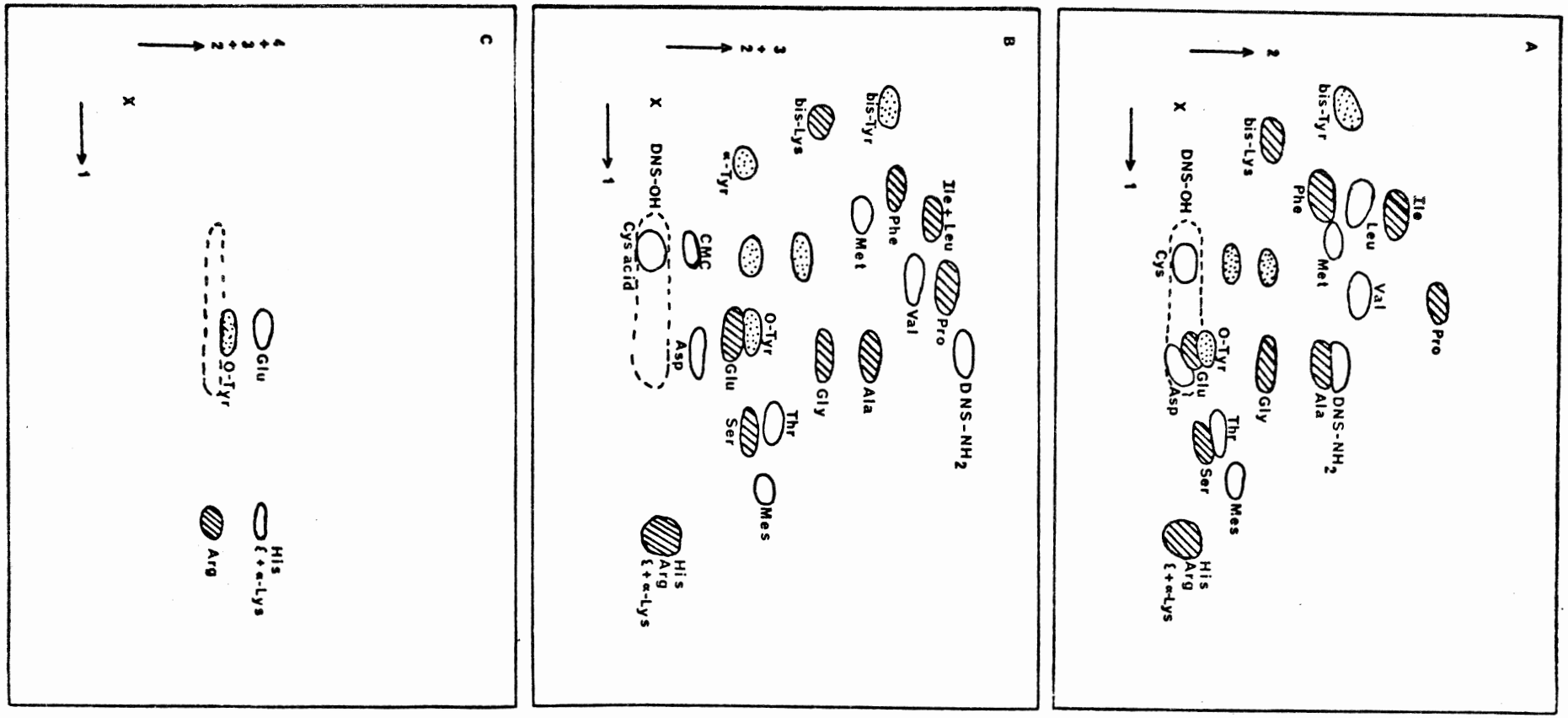


Fig. 4.3. Separation of Dansyl-amino acids on polyacrylamide thin layer plates. Dimension 1 (Water : formic acid, 100 : 1). Dimension 2 (Benzene : Acetic acid, 9 : 1). Dimension 3 (Ethyl Acetate : Methanol : Acetic acid, 20 : 1 : 1). Dimension 4 (0,05 M Na_3PO_4 : Ethanol, 3 : 1).

after which 0,5 ml acetone was added to precipitate protein. Low speed centrifugation pelleted the protein which was washed with 0,5 ml 80% acetone and dried.

Peptide method; Peptides (0,5-5 nmoles) were dissolved in 5 µl 0,2 M $NaHCO_3$ pH 8,5. To this was added 5 µl dansyl chloride in acetone (2,5 mg/ml). This was incubated for 30 minutes at 45°C and then dried.

Hydrolysis of labelled protein or peptide. To the solid residue 20 µl of 5,7 N (constant boiling) HCl was added, the tube sealed and incubated overnight at 105°C, after which the tube was opened and the HCl dried in an evacuated dessicator with NaOH pellets.

Identification of dansyl amino acids; This followed the procedures of Neuhoff (1973) and Gray (1972). The identification was done by thin layer chromatography on Schleicher and Schull F1700 micro-polyamide plastic supported plates, coated on both sides. The original 15x15 cm sheets were cut into 9 separate 5x5 cm pieces. The solvent systems used were;

1. Water : Formic acid 100 : 1,5
2. Benzene : Acetic acid 9 : 1
3. Ethyl Acetate : Methanol : Acetic Acid 20 : 1 : 1
4. 0,05 M Na_3PO_4 : Ethanol 3 : 1

The sample was applied to one of these 5x5 cm plates with a capillary tube as a spot no larger than 1 mm, 1 cm equidistant from the sides of the plate in the right bottom corner. At precisely the same position on the other side of the plate (i.e. left bottom corner, with the plate turned over) the mixture of standards was applied. The standard mixture consisted of 10 µl of each standard (2 mg/ml in 95% ethanol individually) mixed and made up to 400 µl. One spot on the plate is sufficient to identify all the standards. The plate was developed in solvent system 1, dried, turned on its side and run in

system 2. It was dried, examined under U.V. light and developed further in solvent systems 3 and 4 whenever necessary. By inverting the plate, the unknown may be easily identified according to the standards on the other side (see fig. 4.3).

4.4.4.2 Manual Edman Degradation.

The method is that of Peterson et al. 1972.

A 0,4 M solution of dimethylallylamine in 1-propanol-water (3:2 v/v) was titrated to pH 9,5 with trifluoroacetic acid. Approximately 50-200 nanomoles of peptide are dissolved in 100 µl of the above coupling buffer in a glass stoppered 12 ml centrifuge tube. This was flushed with nitrogen and phenylisothiocyanate (10 µl) was added under a stream of nitrogen. The contents were thoroughly mixed and incubated (stoppered under nitrogen) at 50°C for 20 minutes. The mixture was extracted once with benzene and the extract discarded. The coupled peptide was dried at 50°C under a nitrogen stream and then by evacuating at 50°C with a vacuum pump for 30 minutes.

In order to effect cleavage, 100 µl of trifluoroacetic acid were added under nitrogen and the mixture incubated at 50°C for 7 minutes. The trifluoroacetic acid (TFA) was then evaporated with a stream of nitrogen, rotating the tube on its side to spread the material in an even film over a small area at the bottom of the tube. The thiazolinone derivative was extracted from this film with 1 ml of peroxide-free ethyl ether. The peptide may be recycled. The cleavage product was identified by gas chromatography (4.5.3.2).

4.4.4.3 Combined Dansyl-Edman degradation.

This is a method which is useful for identifying peptides which may have the same N-terminus, but differ in subsequent residues. It has also been used successfully to sequence peptides up to 30 residues

(Chen 1976).

The N-terminus of an aliquot of the peptide is determined by the dansylation procedure, while the majority of the sample is subjected to manual Edman degradation. The new N-terminus exposed is identified by dansylation of an aliquot of this peptide. The procedure may be repeated as far as the residues are identifiable. The method has the advantage of requiring very little material, but the disadvantage of quantitation difficulties.

Method: 20-100 nmol of peptide were introduced to a glass-stoppered tube and evaporated to dryness. 50 µl of water were added and 5 µl withdrawn for dansylation. Next, 50 µl 5% PITC in pyridine were added, the tube flushed with nitrogen, stoppered and incubated at 45°C for 1 hour. The solution was evaporated to dryness in a preheated dessicator (60°C) over P_2O_5 and NaOH with an oil pump. Redistilled trifluoroacetic acid (100 µl) was added, the tube sealed and incubated at 45°C for 30 minutes then evaporated to dryness over NaOH. The residue was dissolved in 50 µl water and extracted three times with 200 µl n-butyl acetate, the non-aqueous phase being discarded. The sequence may be repeated for as many steps as are required. The dansylation procedure was the same as that described in 4.4.4.1.

4.4.5 C-Terminal Group.

4.4.5.1 Carboxypeptidase.

Amino acid residues are hydrolysed by carboxypeptidase A and B at different rates (Ambler, 1967). A combination of these may be used to advantage. Carboxypeptidase A hydrolyses homoserine and alanine at a greater rate than carboxypeptidase B, while carboxypeptidase B hydrolyses lysine faster than carboxypeptidase A. One of the enzymes was added singly, the mixture incubated for a set time, and then the second

carboxypeptidase was added. Aliquots were removed from the mixture at various time intervals and subjected to amino-acid analysis to check for free amino acids.

In a typical experiment, approximately 200 nm mollusc CN-1 was dissolved in 500 µl 0,05 M NH_4HCO_3. An aliquot of 20 µl of carboxypeptidase A (25 mg/ml) was diluted to 1 ml with distilled water, the protein centrifuged down and the pellet resuspended in 0,2 ml 1 M NH_4HCO_3. This was diluted to 0,5 ml with water and an aliquot of 50 µl removed and added to the peptide solution. Aliquots of 50 µl were withdrawn during the 37°C incubation at the times indicated (Table 2.19) and added to 50 µl 50% acetic acid. After 40 minutes an aliquot of 5 µl of carboxypeptidase B (5 mg/ml) was added and aliquots withdrawn as above. The time point aliquots were all freeze-dried and applied directly to an amino acid analyser to quantitate free amino acids.

4.5 Sequence Analysis.

4.5.1 Preparative Steps - Peptide modifications.

4.5.1.1 Carboxyl group modification.

During the automatic sequencing of proteins, the yield of PTH-amino acid recovered would frequently drop drastically after a glutamic acid residue. This decreased yield could be due to cyclization of free N-terminal glutamic acid residues to pyroglutamic acid which would render these molecules unavailable for further sequencing.

It had been reported from our laboratory that side-chain carboxyl group modification introduced by Gibson (1972) to differentiate between amide and free carboxyl groups prevented such a drop in yields (Strickland et al. 1977).

The conditions used for the modification were the following:

9 mg of chicken CN-1 were dissolved in 1,5 ml 6 M urea and 150 mg glycine methyl ester were dissolved in a similar aliquot of urea. The two aliquots were mixed and the pH was adjusted to 4,75 with 0,1 M NaOH. 15 mg carbodiimide dissolved in 0,2 ml 6 M urea were added and the mixture reacted at room temperature for 4 hours. The pH was automatically maintained at 4,75 with 0,1 N HCl in a pH-stat. The solution was finally desalted over Sephadex G25 (medium) and freeze-dried.

4.5.1.2 S-PITC Treatment.

During automatic Edman degradation certain hydrophobic peptides are washed out of the sequenator cup by the solvents used. To prevent this loss, the hydrophobicity of the peptide was decreased by the addition of a hydrophilic group to the ε-amino of lysine residues.

This was accomplished by the use of 4-sulfophenylisothiocyanate (S-PITC)(Braunitzer et al. 1973). All amino groups become labelled by this irreversibly attached hydrophilic group. The method has the disadvantage that the amino acids with the S-PITC coupled are not recovered and therefore the sequence shows a gap for the first residue and any subsequent lysine residues. A mixture of S-PITC and PITC used initially should have allowed identification in yields according to the ratio of S-PITC to PITC.

Typically, the peptide was placed in solution into the sequenator cup and a three-fold excess of S-PITC (15 mg/ml in 50% water, 50% pyridine) was added. This mixture was incubated for 30-60 minutes at 55°C, then dried before the first cycle of the Edman degradation.

4.5.1.3 Use of Carrier Protein.

When using the automatic sequenator, short peptides, especially in low amounts, were often extracted during washing steps. Even

after S-PITC treatment, they were not always retained in the cup. The addition of a carrier protein (histone H4 which has a blocked N-terminus) to the peptide solution before introduction into the sequenator cup helped to reduce mechanical losses. Presumably, addition of sufficient carrier produces an even film of protein on the inner surface of the cup and this protein 'net' holds the peptide in place, preventing mechanical losses.

4.5.2 The Automatic Degradation Cycle.

The discovery of an efficient, repetitive degradation cycle for peptides by Edman (1950) led to the manufacture of a commercially available sequenator (Edman and Begg, 1967). The sequenator performs the coupling and cleavage reactions and also collects each derivatized amino acid in a fraction collector, from whence they are collected and identified (4.5.3). All degradations were performed in a Beckman 890 protein sequenator using programs developed previously in our laboratory (Brandt, 1971). For intact H2B, the protein program was used and for peptides the peptide program was used. The quadrol buffer system in these programs has been replaced by DMAP (3-dimethylamino-propyne) buffer (Braunitzer and Schrank, 1970).

The buffer was made up as follows; 1 M 3-dimethylamino-propyne was adjusted to pH 9,0 with trifluoroacetic acid and diluted 4:5 with aldehyde-free propanol.

The nomenclature used for the program statements is as follows; R1 - PITC, R3 - heptafluorobutyric acid, R4 - buffer, R5 - nitrogen, S1 - benzene, S2 - ethyl acetate, S3 - butyl chloride. The protein program has two cleavages each followed by an extraction step, whereas the peptide program has one cleavage and one extraction, which minimize losses from the cup at the expense of reducing cleavage efficiency.

PROTEIN PROGRAM

Step	Program Statement	Step Time (sec)	Cup Speed	Step	Program Statement	Step Time (sec)	Cup Speed
1	Stop slew	2	L	20	Reaction	300	L
2	Delay	4	L	21	R5 deliver	120	L
3	Blank	2	L	22	Blank	2	L
4	R4 vent.	14	L	23	Vac. restricted	60	L
5	R4 press.	14	L	24	Delay	4	L
6	R1 vent.	14	H	25	N_2 dry	300	L
7	R1 pressurize	14	H	26	Vac. restricted	100	H
8	R1 deliver	6	H	27	Vac. rough	200	H
9	Blank	2	H	28	Vac. fine	100	H
10	Vac. restricted	30	H	29	Blank	6	H
11	Delay	6	H	30	S1 vent.	30	H
12	N_2 dry	60	H	31	S1 press.	30	H
13	R4 deliver	28	H	32	S1 deliver	200	H
14	Reaction	300	H	33	N_2 dry	200	H
15	R4 deliver	4	H	34	Vac. restricted	30	H
16	Reaction	300	L	35	Vac. rough	140	H
17	R4 deliver	4	L	36	Delay	3	H
18	Reaction	300	L	37	S2 vent.	30	H
19	R5 deliver	10	L	38	S2 press.	30	H

PROTEIN PROGRAM

Step	Program Statement	Step Time (sec)	Cup Speed	Step	Program Statement	Step Time (sec)	Cup Speed
39	S2 deliver	440	H	59	Vac. fine	60	H
40	Delay	82	H	60	Delay	4	H
41	Vac. restricted	60	H	61	R3 vent.	14	L
42	Vac. rough	40	H	62	R3 press.	14	L
43	Vac. fine + F/C step	300	H	63	R3 deliver	24	H
44	Delay	6	L	64	Reaction	90	H
45	R3 vent. + F/C vent	20	L	65	Vac. restricted	40	L
46	R3 press.	14	L	66	Vac. rough	50	L
47	R3 deliver	26	H	67	Vac. fine	20	L
48	Reaction	190	H	68	Delay	4	L
49	Vac. restricted	40	L	69	S3 vent.	0	L
50	Vac. rough	50	L	70	S3 press.	0	L
51	Vac. fine	20	L	71	S3 deliver + waste	200	H
52	Delay	4	L	72	Delay	40	H
53	S3 vent.	0	L	73	Vac. restricted	60	H
54	S3 press.	0	L	74	Vac. rough	40	H
55	S3 deliver + collect	200	H	75	Vac. fine	200	H
56	Delay	40	H	76	Vac. fine	0	L
57	Vac. restricted	60	H	77	Vac. fine	2	L
58	Vac. rough	60	H	78	Vac. fine	0	L

PEPTIDE PROGRAM

Step	Program Statement	Step Time (sec)	Cup Speed	Step	Program Statement	Step Time (sec)	Cup Speed
1	Stop slew	2	L	20	R5 deliver	10	L
2	Delay	6	L	21	Reaction	300	L
3	Blank	2	L	22	R5 deliver	120	L
4	R4 vent.	14	L	23	Blank	2	L
5	R4 press.	2	L	24	Vac. restricted	60	L
6	Delay	14	L	25	Delay	4	L
7	R1 vent.	14	H	26	N_2 dry	400	L
8	R1 press.	14	H	27	Vac. restricted	100	H
9	R1 deliver	6	H	28	Vac. rough	200	H
10	Blank	2	H	29	Vac. fine	400	H
11	Vac. restricted	30	H	30	Blank	2	H
12	Delay	6	H	31	Delay + F/C step	6	H
13	N_2 dry	60	H	32	S1 vent	30	H
14	R4 deliver	28	H	33	S1 press. + F/C vent.	30	H
15	Reaction	300	H	34	S1 deliver	20	H
16	R4 deliver	0	L	35	Blank	30	H
17	Reaction	300	L	36	Blank	2	H
18	R5 deliver	4	L	37	Blank	2	H
19	Reaction	300	L	38	Blank	2	H

PEPTIDE PROGRAM

Step	Program Statement	Step Time (sec)	Cup Speed	Step	Program Statement	Step Time (sec)	Cup Speed
39	Delay	6	H	59	Vac. rough	30	L
40	S1 deliver + collect	150	H	60	Delay	6	L
41	Delay + collect	30	H	61	S3 deliver	25	L
42	Blank	2	H	62	N_2 dry	230	L
43	Vac. restricted	30	H	63	Vac. restricted	62	L
44	Delay	6	H	64	Vac. rough	62	L
45	N_2 dry	200	H	65	Vac. fine	62	L
46	Vac. restricted	100	H	66	Blank	2	L
47	Vac. rough	300	H	67	Delay	6	L
48	Vac. fine	300	H	68	S3 deliver + collect	150	H
49	Blank	2	H	69	Delay	30	H
50	Delay	6	H	70	Blank	2	H
51	S3 vent.	30	H	71	Vac. restricted	60	H
52	S3 press. + F/C vent. + F/C step	30	H	72	Delay	6	H
53	R3 vent.	14	H	73	N_2 dry	200	H
54	R3 press.	14	H	74	Vac. restricted	60	H
55	R3 deliver	24	H	75	Vac. rough	300	H
56	Reaction	80	H	76	Vac. fine	800	H
57	N_2 dry	40	L	77	Vac. fine	0	L
58	Vac. restricted	20	L	78	Vac. fine	800	L

The cup speeds L and H correspond to 1000 and 1500 r.p.m. respectively. F/C ≡ fraction collector.

In addition, the peptide program has no ethyl acetate wash, which can lead to losses of hydrophobic peptides.

The reagents and solvents were purified according to the methods of Edman and Henschen (1975) except for DMAP, which was purified according to Braunitzer and Schrank (1970).

4.5.3 Quantitation of phenylthiohydantoin (PTH) amino acids.

The thiazolinones were first converted to identifiable PTH-amino acids (4.5.3.1) and then quantitated by one of three methods (4.5.3.2). To each tube in the sequenator fraction collector 0,1 mg of dithioerythritol was added to improve the recovery of serine and threonine and also 100 nanomoles of norleucine as an internal standard.

4.5.3.1 Conversion of amino acid thiazolinones.

Thiazolinones were converted as described by Edman and Begg (1967). To the dried material 0,25 ml 1 N HCl containing 1% (v/v) ethanethiol was added, after which the tube was flushed with nitrogen and stoppered with a silicone stopper. The tube was heated at 80°C for 10 minutes in a sand-filled heating block. After the solution was cooled, it was twice extracted with 1 ml peroxide free ethyl acetate (sequencer grade, to which solid ascorbic acid was added).

The organic phase was evaporated in a separate tube under a stream of nitrogen at 50°C. The residue was dissolved in 50 µl peroxide free ethyl acetate and aliquots (2 to 8%) of this solution used for gas chromatographic identification of the PTH-amino acids. The aqueous phase was dried under a stream of nitrogen at 50°C and the residue transferred with methanol to a hydrolysis tube. The residual ethyl acetate solution was also transferred and the combined solutions evaporated to dryness.

In later experiments, the thiazolinones were converted by the method of Wittman-Liebold et al. (1975). To the dry sample from the fraction collector 0,2 ml of 20% TFA was added and then the tube was flushed with nitrogen. The contents were mixed well and heated at 80°C for 10 minutes. The residue was dried with nitrogen at 40°C and re-dissolved in ethyl acetate for gas chromatographic identification or methanol for liquid chromatographic identification.

4.5.3.2 Identification of PTH-amino acids by gas chromatography.

The converted thiazolinones (4.5.3.1) were examined before and after silylation on a Beckman GC-45 gas chromatograph (Pisano, 1969) using a SP-400 (Chromosorb W (100-120 mesh) column. Those amino acids which were not identifiable by gas chromatography, i.e. arginine, lysine, histidine and derivatized glutamic acid, were identified by amino acid analyses (4.5.3.3).

4.5.3.3 Hydrolysis and amino-acid analysis of PTH-amino acids.

Samples were hydrolysed in 0,2 ml constant boiling HCl (twice distilled) with 1% thioglycollic acid for 24 hours at 130°C (Van Orden et al. 1964) after which identification of amino acids was done on a Beckman 116 amino-acid analyser with modifications to increase the sensitivity.

4.5.3.4 Identification of PTH-amino acids by High-Pressure liquid chromatography.

A Hewlett-Packard 1084A liquid chromatograph with a 250x4,6 mm Lichrosorb RP-8 column was used. The buffer was 1,5 mM sodium acetate pH 4,75 and the runs were done at 35°C. A linear methanol gradient from 5 to 45% over twentyfive minutes separated 18 PTH-amino acids (Strickland et al. 1978). The unresolved PTH-amino acids (phe, ile) were checked by gas chromatography. PTH-serine and PTH-threonine were not stable, but were identified by characteristic derivative peaks.

These derivatives were identified from sequencing the serine and threonine residues of a protein with a known primary structure. Quantitation using an internal standard (i.e. norleucine) was done by a dedicated computer.

References.

Ambler, R.P. (1967). Carboxypeptidases A and B. Methods in Enzymology, 11, 436-444.

Argos, P., Schwarz, J. and Schwarz, J. (1976). An Assessment of Protein Secondary Structure Prediction Methods Based on Amino Acid Sequence. Biochim. Biophys. Acta 439, 261-273.

Armstrong, M.D. (1949). The Relationship between Homoserine and its Lactone. Am. Chem. Soc. J. 71, 3399-3402.

Bailey, G.S. and Dixon, G.H. (1973). Histone IIb1 from Rainbow Trout. J. Biol. Chem. 248, 5463-5472.

Bohm, E.L., Strickland, W.N., Strickland, M., Thwaits, B.H., van der Westhuyzen, D.R. and von Holt, C. (1973). Purification of the five main calf thymus histone fractions by gel exclusion chromatography. FEBS. Lett. 34, 217-221.

Bols, N.C. and Kasinsky, H.E. (1976). On the Diversity of Sperm Histones in the Vertebrates. J. Exp. Zool. 198, 109-114.

Bonilla, C.A. (1969). Rapid isolation of basic proteins and polypeptides from salivary gland secretions by adsorption chromatography on polyacrylamide gel. Anal. Biochem. 32, 522-529.

Bradshaw, R.A. (1969). The Amino Acid Sequence of Bovine Carboxypeptidase A. III. Specificity of Peptide-Bond Cleavage by Thermolysin and the Complete Sequence of the Cyanogen Bromide Fragment F III. Biochemistry 8, 3871-3877.

Brandt, W.F., Edman, P., Henschen, A. and von Holt, C. (1976). Abnormal Behaviour of Proline in the Isothiocyanate Degradation. Hoppe-Seyler's Z. Physiol. Chem. 357, 1505-1508.

Brandt, W.F., Strickland, W.N. and von Holt, C. (1974). The Primary Structure of Histone F3 from Shark Erythrocytes. FEBS. Lett. 40, 349-352.

Brandt, W.F. and von Holt, C. (1972). The Complete Amino Acid Sequence of Histone F3 from Chicken Erythrocytes. FEBS. Lett. 23, 357-360.

Brandt, W.F. and von Holt, C. (1971). Purification and Characterization of the Cysteine-containing F3 Histone from Chicken Erythrocyte. FEBS. Lett. 14, 338-345.

Brandt, W.F. and von Holt, C. (1974). The Determination of the Primary Structure of Histone F3 from Chicken Erythrocytes by Automatic Edman Degradation. Eur. J. Biochem. 46, 407-429.

Braunitzer, G. and Schrank, B. (1970). 3-Dimethylamino-1-propin als Puffersubstanz. Hoppe-Seyler's Z. Physiol. Chem. 351, 417.

Braunitzer, G., Schrank, B., Petersen, S. and Petersen, U. (1973). Automatische Sequenzanalyse des Insulins. Hoppe-Seyler's Z. Physiol. Chem. 354, 1563-1566.

Butler, P.J.G., Harris, J.I., Hartley, B.S. and Leberman, R. (1969). The Use of Maleic Anhydride for the Reversible Blocking of Amino Groups in Polypeptide Chains. Biochem. J. 112, 679-689.

Candido, E.P.M. and Dixon, G.H. (1972). Amino-Terminal Sequences and Sites of In Vivo Acetylation of Trout-Testis Histones III and IIb_2 Proc. Nat. Acad. Sci. USA 69, 2015-2019.

Cernosek, S.F., Wells, M. and Fasman, G.D. (1974). Synthesis and Conformational Studies of Model Histones: II. Sequential and Random Polypeptides with the composition L-Lysyl : L-Alanyl : L-Proline. Israel J. Chem. 12, 47-66.

Chen, R. (1976). The Sequence Determination of a Protein in a Micro Scale: The Sequence Analysis of Ribosomal Protein L34 of Escherichia Coli. Hoppe-Seyler's Z. Physiol. Chem. 357, 873-886.

Chen, R., Mende, L. and Arfsten, U. (1975). The Primary Structure of Protein L27 from the Peptidyl-tRNA Binding site of Escherichia Coli Ribosomes. FEBS. Lett. 59, 96-99.

Chou, P.Y. and Fasman, G.D. (1974). Conformational Parameters for Amino Acids in Helical, β-Sheet and Random Coil Regions Calculated from Proteins. Biochemistry 13, 211-222.

Chou, P.Y. and Fasman, G.D. (1974). Prediction of Protein Conformation. Biochemistry 13, 222-245.

Coelingh, J.P., Monfoort, C.H., Rozijn, T.H., Leuven, J.A.G., Schiphof, R., Steyn-Parve, E.P., Braunitzer, G., Schrank, B. and Ruhfus, A. (1972). The Complete Amino Acid Sequence of The Basic Nuclear Protein of Bull Spermatozoa. Biochim. Biophys. Acta 285, 1-14.

Dayhoff, M.O. (1972), (1973), (1976). Atlas of Protein Sequence and Structure.

Delange, R.J., Hooper, J.A. and Smith, E.L. (1973). Histone III. III. Sequence Studies on the Cyanogen Bromide Peptides; Complete Amino Acid Sequence of Calf Thymus Histone III. J. Biol. Chem. 248, 3261-3274.

Destree, O.H.J., d'Adelhart-Toorop, H.A. and Charles, R. (1972). Analysis of Histones from Different Tissues and Embryos of Xenopus laevis. Acta Morphol. Neerl. - Scand. 10, 233-248.

Dixon, G.H. (1975). In discussion. Ciba Found. Symp. Struct. Func. Chromatin 28, 267.

Dixon, G.H., Candido, E.P.M., Honda, B.M., Louie, A.J., Macleod, A.R. and Sung, M.T. (1975). The biological roles of post-synthetic modifications of basic nuclear proteins. in The Structure and Function of Chromatin - Ciba Foundation Symposium, North-Holland.

Doyle, B.B., Traub, W., Lorenzi, G.P. and Blout, E.R. (1971). Conformational Investigations on the Polypeptide and Oligopeptides with the Repeating Sequence L-Alanyl-L-prolylglycine. Biochemistry 10, 3052-3060.

Drapeau, G.R., Boily, Y. and Houmard, J. (1972). Purification and Properties of an Extracellular Protease of Staphylococcus aureus. J. Biol. Chem. 247, 6720-6726.

Drucker, H. and Borchers, S.L. (1971). The Role of Calcium in Thermolysin: Effect on Kinetic Properties and Autodigestion. Arch. Biochem. Biophys. 147, 242-248.

Edman, P. (1950). Method for Determination of Amino Acid Sequence in Peptides. Acta Chem. Scand. 4, 283.

Edman, P. and Begg, G. (1967). A Protein Sequenator. Eur. J. Biochem. 1, 80-91.

Edman, P. and Henschen, A. (1975). Sequence Determination in Molecular Biology, Biochemistry and Biophysics, Vol. 8. Ed. S.B. Needleman, Springer-Verlag. Chapter 8, pp 222-279.

Fambrough, D.M. and Bonner, J. (1966). On the similarity of plant and animal histones. Biochemistry 5, 2563-2569.

Fasman, G.D., Chou, P.Y. and Adler, A.J. (1976). Prediction of the Conformation of the Histones. Biophys. J. 16, 1201-1238.

Fasman, G.D., Chou, P.Y. and Adler, A.J. (1977). Histone conformation predictions and experimental studies. in The Molecular Biology of the Mammalian Genetic Apparatus 1. Ed. P.O.P. Ts'o, North-Holland.

Franco, L., Montero, F., Navlet, J.M., Perera, J. and Rojo, M.C. (1974). Histones from the Fruit Fly Ceratitis capitata. Eur. J. Biochem. 48, 53-61.

Franklin, S.G. and Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. Nature 266, 273-275.

Gibson, D. and Anderson, P.J. (1972). Carboxyl Group Modification and Amide Assignments in Automated Sequencing of Proteins. Biochem. Biophys. Res. Comm. 49, 453-460.

Goff, C.G. (1976). Histones of Neurospora crassa. J. Biol. Chem. 251, 4131-4138.

Gray, W.R. (1972). Sequence analysis with Dansyl Chloride. Methods in Enzymology 25, 333-344.

Gross, E. (1967). The Cyanogen Bromide Reaction. Methods in Enzymology 11, 238-255.

Hayashi, H., Iwai, K., Johnson, J.D. and Bonner, J. (1977). Pea Histones H2A and H2B. J. Biochem. 82, 503-510.

Hewish, D.R. and Burgoyne, L.A. (1973). Chromatin Sub-Structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem. Biophys. Res. Comm. 52, 504-510.

Hirs, C.H.W., Moore, S. and Stein, W.H. (1956). Peptides obtained by Tryptic Hydrolysis of Performic Acid-Oxidized Ribonuclease. J. Biol. Chem. 219, 623-642.

Hnilica, L.S. (1972). The Structure and Biological Functions of Histones. The Chemical Rubber Co.

Hooper, J.A., Smith, E.L., Sommer, K.R. and Chalkley, R. (1973). Histone III. IV. Amino acid sequence of histone III of the testes of the carp (Letiobus bubalus). J. Biol. Chem. 248, 3275-3279.

Horn, M.J. and Laursen, R.A. (1973). Solid-Phase Edman Degradation: Attachment of Carboxyl-Terminal Homoserine Peptides To An Insoluble Resin. FEBS. Lett. 36, 285-288.

Houmard, J. and Drapeau, G.R. (1972). Staphylococcal Protease: A Proteolytic Enzyme Specific for Glutamoyl Bonds. Proc. Nat. Acad. Sci. USA 69, 3506-3509.

Iwai, K., Hayashi, H. and Ishikawa, K. (1972). Calf Thymus Lysine and Serine-Rich Histone. J. Biochem. 72, 357-367.

Iwai, K., Ishikawa, K. and Hayashi, H. (1970). Amino-acid Sequence of Slightly Lysine-rich Histone. Nature 226, 1056-1058.

Johmann, C.A. and Gorovsky, M.A. (1976). Purification and Characterization of the Histones Associated with the Macronucleus of Tetrahymena. Biochemistry 15, 1249-1256.

Johns, E.W. (1964). Studies on Histones 7. Preparative Methods for Histone Fractions from Calf Thymus. Biochem. J. 92, 55-60.

Kato, Y. and Iwai, K. (1977). DNA-binding Segments of Four Histone Sequences Identified in Trypsin-treated H1-depleted Chromatin. J. Biochem. 81, 621-630.

Kootstra, A. and Bailey, G.S. (1976). The Primary Structure of Histone H2B from brown trout (Salmo trutta) testes. FEBS. Lett. 68, 76-78.

Kornberg, R.D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. Science 184, 868-871.

Kuroda, Y., Hashimoto, E. and Nishizuka, Y. (1976). Phosphorylated Sites of Calf Thymus H2B Histone by Adenosine 3':5' - Monophosphate-Dependent Protein Kinase from Bovine Cerebellum. Biochem. Biophys. Res. Comm. 71, 629-635.

Lehninger, A.L. (1972). Biochemistry. Worth Publishers, Inc. New York.

Lewis, P.N. and Bradbury, E.M. (1974). Effect of Electrostatic Interactions on the Prediction of Helices in Proteins: The Histones. Biochim. Biophys. Acta, 336, 153-164.

Lewis, P.N., Momany, F.A. and Scheraga, H.A. (1971). Folding of Polypeptide Chains in Proteins: A proposed mechanism for folding. Proc. Nat. Acad. Sci. USA 68, 2293-2296.

Li, H-J. and Bonner, J. (1971). Interaction of Histone Half-Molecules with Deoxyribonucleic Acid. Biochemistry 10, 1461-1470.

Martinage, A., Sautiere, P., Kerckaert, J-P. and Biserte, G. (1976). Purification by Preparative Electrophoresis and Characterization of Histone H2B from Rat Chloroleukaemia. Biochim. Biophys. Acta 420, 37-41.

Martinson, H.G. and McCarthy, B.J. (1976). Histone-Histone Interactions within Chromatin. Preliminary Characterization of Presumptive H2B-H2A and H2B-H4 Binding Sites. Biochemistry 15, 4126-4131.

Marushige, Y. and Marushige, K. (1974). Properties of Chromatin Isolated from Bull Spermatozoa. Biochim. Biophys. Acta 340, 498-508.

Matsubara, H., Sasaki, R., Singer, A. and Jukes, T.H. (1966). Specific Nature of Hydrolysis of Insulin and Tobacco Mosaic Virus Protein by Thermolysin. Arch. Biochem. Biophys. 115, 324-331.

Matsubara, H., Singer, A., Sasaki, R. and Jukes, T.H. (1965). Observations on the specificity of a thermostable bacterial protease "thermolysin". Biochem. Biophys. Res. Commun. 21, 242-247.

Matthews, B.W. (1975). Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. Biochim. Biophys. Acta 405, 442-451.

McLachlan, A.D. (1977). Analysis of Periodic Patterns in Amino Acid Sequences: Collagen. Biopolymers 16, 1271-1297.

McMaster-Kaye, R. and Kaye, J.S. (1973). An Electrophoretic Analysis of the Histones of the House Cricket. Arch. Biochem. Biophys. 156, 426-436.

Miki, B.L.A. and Neelin, J.M. (1977). Comparison of the histones from fish erythrocytes. Canad. J. Biochem. 55, 1220-1227.

Nadeau, P., Pallotta, D. and Lafontaine, J-G. (1977). Comparative study of rye and thymus histones: Amino acid analysis and tryptic fingerprinting. Can. J. Biochem. 55, 721-727.

Neuhoff, V. (1973). Micromethods in Molecular Biology, pp 85-145, Neuhoff, V. ed. Springer-Verlag, Berlin-Heidelberg - New York.

Niall, H.D., Jacobs, J.W., van Rietschoten, J. and Tregear, G.W. (1974). Protected Edman degradation: A new approach to microsequence analysis of proteins. FEBS. Lett. 41, 62-64.

Noll, M. (1974). Internal structure of the chromatin subunit. Nucl. Acids Res. 1, 1573-1578.

Oliver, D. and Chalkley, R. (1972). An Electrophoretic Analysis of Drosophila Histones. I and II. Exp. Cell Res. 73, 295-310.

Oliver, D., Sommer, K.R., Panyim, S., Spiker, S. and Chalkley, R. (1972). A Modified Procedure for Fractionating Histones. Biochem. J. 129, 349-353.

Ong, E.C., Snell, C. and Fasman, G.D. (1976). Chromatin Models. The Ionic Strength Dependence of Model Histone-DNA Interactions. Circular Dichroism Studies of Lysine-Leucine Polypeptide-DNA Complexes. Biochemistry 15, 468-476.

Pallotta, D. and Tessier, A. (1975). Amino acid composition of sperm histones in the house cricket Acheta domesticus. Can. J. Biochem. 54, 56-61.

Panyim, S., Bilek, D. and Chalkley, R. (1971). An Electrophoretic Comparison of Vertebrate Histones. J. Biol. Chem. 246, 4206-4215.

Panyim, S. and Chalkley, R. (1969). High Resolution Acrylamide Gel Electrophoresis of Histones. Arch. Biochem. Biophys. 130, 337-346.

Panyim, S., Jensen, R.H. and Chalkley, R. (1968). Proteolytic contamination of calf thymus nucleohistone and its inhibition. Biochim. Biophys. Acta 160, 252-255.

Patthy, L. and Smith, E.L. (1973). Histone III. V. The Amino Acid Sequence of Pea Embryo Histone III. J. Biol. Chem. 248, 6834-6840.

Pauling, L., Corey, R.B. and Branson, H.R. (1951). The Structure of Proteins: Two Hydrogen-Bonded Helical configurations of the Polypeptide chain. Proc. Nat. Acad. Sci. USA 37, 205-212.

Peterson, J.D., Nehrlich, S., Oyer, P.E. and Steiner, D.F. (1972). Determination of the Amino Acid Sequence of the Monkey, Sheep and Dog Proinsulin C-Peptides by a Semi-micro Edman Degradation Procedure. J. Biol. Chem. 247, 4866-4871.

Pisano, J.J. and Bronzert, T.J. (1969). Analysis of amino acid phenylthiohydantoins by gas chromatography. J. Biol. Chem. 244, 5597-5607.

Purkayastha, R. and Neelin, J.M. (1966). Comparison of Histones from Avian Erythroid Tissues by Zone Electrophoresis. Biochim. Biophys. Acta 127, 468-477.

Ramachandran, L.K. and Witkop, B. (1967). N-Bromosuccinimide Cleavage of Peptides. Methods in Enzymology, Vol XI.

Sautiere, P., Tyrou, D., Laine, B., Mizon, J., Ruffin, P. and Biserte, G. (1974). Covalent Structure of Calf-Thymus ALK-Histone. Eur. J. Biochem. 41, 563-576.

Schwartz, A.M. and Fasman, G.D. (1977). Interactions of DNA with Poly (L-Lys-L-Ala-Gly) and Poly (L-Lys-L-Ala-L-Pro). Circular Dichroism and Thermal Denaturation Studies. Biochemistry 16, 2287-2299.

Shires, A., Carpenter, M.P. and Chalkley, R. (1975). New histones found in mature mammalian testes. Proc. Nat. Acad. Sci. USA 72, 2714-2718.

Shires, A. and Carpenter, M.P. (1976). A Cysteine-containing H2B-like Histone found in Mature Mammalian Testis. J. Biol. Chem. 251, 4155-4158.

Snell, C.R. and Fasman, G.D. (1972). Conformational studies on copolymers of L-Leucine: Circular dichroism and potentiometric titration studies. Biopolymers 11, 1723.

Spiker, S. (1975). An Evolutionary Comparison of Plant Histones. Biochim. Biophys. Acta 400, 461-467.

Spiker, S., Key, J.L. and Wakim, B. (1976). Identification and Fractionation of Plant Histones. Arch. Biochem. Biophys. 176, 510-518.

Stark, G.R., Stein, W.H. and Moore, S. (1960). Reactions of the cyanate present in aqueous urea with amino acids and proteins. J. Biol. Chem. 235, 3177-3181.

Stauffer, C.E. (1971). The Effect of pH on Thermolysin activity. Arch. Biochem. Biophys. 147, 568-570.

Stedmann, E. and Stedmann, E. (1950). Cell Specificity of Histones. Nature 166, 780-781.

Stevely, W.S. and Stocken, L.A. (1966). Phosphorylation of rat thymus histone. Biochem. J. 100, 20c.

Strickland, W.N., Strickland, M., Brandt, W.F., Morgan, M. and von Holt, C. (1974). Partial Amino Acid Sequence of Two New Arginine-Serine Rich Histones from Male Gonads of The Sea Urchin (Parechinus angulosus). FEBS. Lett. 40, 161-166.

Strickland, M., Strickland, W.N., Brandt, W.F. and von Holt, C. (1977). The Complete Amino-Acid Sequence of Histone $H2B_{(1)}$ from Sperm of the Sea Urchin Parechinus angulosus. Eur. J. Biochem. 77, 263-275.

Strickland, W.N., Strickland, M., Brandt, W.F. and von Holt, C. (1977). The Complete Amino-Acid Sequence of Histone $H2B_{(2)}$ from Sperm of the Sea Urchin Parechinus angulosus. Eur. J. Biochem. 77, 277-286.

Strickland, M., Strickland, W.N., Brandt, W.F., von Holt, C., Wittmann-Liebold, B. and Lehmann, A. (1978). The Complete Amino Acid Sequence of Histone $H2B_{(3)}$ from Sperm of the Sea Urchin Parechinus angulosus. Eur. J. Biochem. in press.

Subirana, J.A., Cozcolluela, C., Palau, J. and Unzeta, M. (1973). Protamines And Other Basic Proteins From Spermatozoa of Molluscs. Biochim. Biophys. Acta 317, 364-379.

Sung, M.T. and Dixon, G.H. (1970). Modification of Histones during Spermiogenesis in Trout: A Molecular Mechanism for Altering Histone Binding to DNA. Proc. Nat. Acad. Sci. USA 67, 1616-1623.

Tsai, Y.H. and Hnilica, L.S. (1975). Tissue - Specific Histones in the Erythrocytes of Chicken and Turtle. Exp. Cell. Res. 91, 107-112.

van der Westhuyzen, D.R., Bohm, E.L. and von Holt, C. (1974). Fractionation of Chicken Erythrocyte Whole Histone into the six main components by Gel Exclusion Chromatography. Biochim. Biophys. Acta 359, 341-345.

van Helden, P.D., Strickland, W.N., Brandt, W.F. and von Holt, C. (1978a). The complete amino-acid sequence of histone H2B from the mollusc Patella granatina. Submitted for publication.

van Helden, P.D., Strickland, W.N., Brandt, W.F. and von Holt, C. (1978b). Histone H2B variants from the erythrocytes of an amphibian, a reptile and a bird. Biochim. Biophys. Acta 533, 278-281.

Vanhoutte-Durand, G., Mizon, J., Sautiere, P. and Biserte, G. (1977). Histones From Gonads of The Star-Fish Asterias Rubens. Comp. Biochem. Physiol. 57B, 121-126.

Van Orden, H.O. and Carpenter, F.H. (1964). Hydrolysis of phenylthiohydantoins of amino acids. Biochem. Biophys. Res. Comm. 14, 399-403.

Vaughn, J.C. and Thomson, L.A. (1972). A Kinetic Study of DNA and Basic Protein Metabolism during Spermatogenesis in The Sand Crab, Emerita Analoga. J. Cell Biol. 52, 322-336.

von Holt, C. and Brandt, W.F. (1977). Fractionation of Histones on Molecular Sieve Matrices. Methods In Cell Biology 14, 205-225.

Warrant, R.W. and Kim, S-H. (1978). α-Helix-double helix interaction shown in the structure of a protamine-transfer RNA complex and a nucleoprotamine model. Nature 271, 130-135.

Weintraub, H., Palter, K. and van Lente, F. (1975). Histones H2a, H2b, H3 and H4 Form a Tetrameric Complex in Solutions of High Salt. Cell 6, 85-110.

Williams, R.E. (1976). Phosphorylated Sites in Substrates of Intracellular Protein Kinases: A Common Feature in amino acid Sequences. Science 192, 473-474.

Wittmann-Liebold, B., Geissler, A.W. and Marzinzig, E. (1975). Studies on the primary structure of 14 proteins from the large subunit of Escherichia coli ribosomes with an improved protein sequenator and with mass spectrometry. J. Supramol. Struct. 3, 426-447.

Wright, E.B. and Olins, D.E. (1975). Histone Stoichiometry in Chicken Erythrocyte Nuclei. Biochem. Biophys. Res. Comm. 63, 642-650.

Yaron, A., Tal, N. and Berger, A. (1972). Synthesis and conformation of poly (L-Lysyl-L-Alanyl-L-Alanyl), a sequence-ordered water-soluble copolymer. Biopolymers 11, 2461.

Yeaman, S.J., Cohen, P., Watson, D.C. and Dixon, G.H. (1977). The Substrate Specificity of Adenosine 3':5' - Cyclic Monophosphate-Dependent Protein Kinase of Rabbit Skeletal Muscle. Biochem. J. 162, 411-421.

Yokotsuka, K., Kikuchi, A. and Shimura, K. (1971). Fractionation and Characterization of Silk Gland Histones. J. Biochem. 70, 1055-1059.

Yoshida, M., Yokotsuka, K. and Shimura, K. (1966). Amino Acid Composition and Heterogeneity of Histones from Posterior Silkgland. J. Biochem. 60, 586-588.

Zweidler, A. (1976). Tissue - Specific Multiplicity of Histones F2b and F2a2. Abstract 593. Abstracts Twelfth Annual Meeting American Society for Cell Biology, 297a.
