Eur. J. Biochem. *136*, 245–252 (1983)

Acetylation and methylation sites in histone H4 from *Physarum polycephalum*

Jaap H. WATERBORG, Sharon R. FRIED, and Harry R. MATTHEWS

Department of Biological Chemistry, School of Medicine, University of California, Davis

(Received May 30, 1983) – EJB 83 0542

Histone H4 has been isolated and purified from plasmodia of *Physarum polycephalum*. The four major fragments produced by hydrolysis of H4 by acetic acid were separated and the complete amino acid sequence of two of them was determined. By analogy with calf H4, these peptides are at the C-terminus and give the sequence from residue 68 to the C-terminus (residue 102). In this 35 residue sequence there are two minor differences from calf H4: (i) residue 77 is arginine in *Physarum* H4 and lysine in calf H4; (ii) lysine-79 is partially methylated in *Physarum*. Arginine occurs at position 77 in pea H4 but the occurrence of methylated lysine at position 79 has not been reported for other species. In the N-terminal region, amino acid compositions of acetic acid, tryptic and chymotryptic peptides indicate that *Physarum* H4 and calf H4 have identical sequences from the N-terminus to residue 47. There may be minor differences in the region from residue 46 to residue 67. The sites of acetylation were determined by Edman degradation of acetate-labelled peptide 4–17 of *Physarum* H4. Acetylation was observed at positions 5, 8, 12, and 16. The only other labelled peptide was the N-terminal peptide, which is not susceptible to Edman degradation and is thus probably α-*N*-acetylated as in most other organisms. The results confirm the conservation of H4 sequence and place *Physarum* H4 in an intermediate position between lower eukaryote H4, such as yeast or *Tetrahymena* H4, and higher eukaryote H4, such as mammalian H4 or pea H4.

The subunit structure in chromatin, the nucleosome, is a highly conserved feature of eukaryotic cells. Within the nucleosome the amino acid sequence of H4 is exceptionally well conserved with only two conservative replacements between histones H4 from pea and calf [1–3]. Some additional minor variations occur in other organisms for example yeast, *Tetrahymena* and sea urchin [4–8].

However, there is an extensive capacity for modification superimposed on the conserved sequence. For example, in calf histone H4 but not pea histone H4 the lysine at position 20 is methylated to a mixture of mono-ε-*N*-methyllysine and di-ε-*N*-methyllysine [1,2]. There are also four lysines that may be reversibly modified to ε-*N*-acetyllysine and, in trout histone H4 these sites are the four lysines nearest the amino terminus, at positions 5, 8, 12, and 16 [9]. It has been suggested that acetylation at these sites is required for opening out the chromatin structure for DNA replication or transcription [10–13] and we recently showed that different patterns of acetylation are correlated with these two processes [14].

An improved method was recently described for preparing histones from the true slime mold *Physarum polycephalum* [15] which has been extensively used for studies of histone modification [14]. This method has been used to prepare pure histone H4 from *Physarum* and we now report the result of a study of the amino acid sequence of *Physarum* histone H4, with an emphasis on determining the location of the sites of modification.

MATERIALS AND METHODS

Physarum *histone H4*

Histones of *Physarum polycephalum*, strain M3C, were isolated from microplasmodial cultures according to Mende et al. [15]. They were fractionated on Sephadex G75 (3.7 × 115 cm), equilibrated with 5% (v/v) acetic acid, at 40 ml/h. The fractions containing H4 (14) were pooled, lyophilized and rechromatographed on Sephadex G75-Superfine (1.2 × 115 cm), equilibrated with 5% acetic acid, at 2.3 ml/h. The center fractions of the H4 peak were homogeneous as judged by acid/urea/Triton X-100 gels [14,15]. These peak fractions were pooled and lyophilized and the amount of H4 was estimated by weight.

Histone H4 was labelled with sodium[^{3}H]acetate for 10 min in the presence of 10 μg/ml of cycloheximide in macroplasmodia in mid-G2 phase, 5 h after mitosis 2 (unpublished results). Histones were isolated according to Mende et al. [15] and added to partially purified H4 prior to Sephadex G75-Superfine chromatography. The specific activities used were 13000 counts × min^{-1} × mg H4^{-1} (Fig. 1B) and 6000 counts × min^{-1} × mg H4^{-1} (Fig. 3B).

Histone H4 was also labelled with sodium[^{3}H]acetate for 5 min in S phase, 20 min after mitosis 2. Under these conditions a limited part of the label is metabolized and incorporated as amino acids in newly synthesized histone [14]. This H4, co-purified with unlabelled H4 through Sephadex G75-Superfine, was used at a specific activity of 15500 counts × min^{-1} × mg H4^{-1} (Fig. 1A).

Acetic acid hydrolysis of H4

Histone H4 was hydrolyzed at 1 mg/ml in 0.25 M acetic acid under nitrogen in a sealed glass tube for 6 h at 109 °C according to Lewis et al. [17] or at 1 mg/ml in nitrogen-flushed 0.25 M acetic acid *in vacuo* in a sealed glass tube for 32–35 h at 104 °C according to Cary et al. [13]. Hydrolyzates of up to 13 mg H4 were lyophilized and taken into 0.5 ml 5% acetic acid. The sample was chromatographed on Sephadex G50-

Superfine (1.2 × 115 cm), equilibrated in 5 % acetic acid, at 2.3 ml/h. Fractions of 0.8 ml were collected.

Chymotryptic digestion of H4

Histone H4 was digested with chymotrypsin treated with N^{α}-tosyl-L-lysine chloromethylketone, essentially as described previously [15], at 2 mg/ml in 20 mM Tris/HCl, pH 7.4, for 20 min at 21 °C, at a weight ratio of chymotrypsin:H4 of 1:250 (Fig. 3A) or 1:50 (Fig. 3B). The digestion was stopped by acidification to 5 % acetic acid followed by lyophilization. The peptides were separated on Sephadex G50-Superfine as described for acetic acid hydrolyzates.

Preliminary peptide identification

The peptide composition of gel filtration fractions was determined by gel electrophoresis on acid/urea gels, described previously for chymotryptic digests of H4 by Mende et al. [15]. This gel system allowed a preliminary identification of amino terminal peptides that display multiple forms differing in the degree of acetylation, and other peptides, based on co-electrophoresis with calf H4 peptides of known composition. Partially digested fragments of *Physarum* H4 indicated in Fig. 1 and 3 were thus identified. Further identification was obtained through amino acid analysis of all limit peptides and most partial ones, and by sequencing of selected peptides.

Peptide nomenclature

Peptides obtained by acetic acid hydrolysis were labelled 'a' and numbered starting with the amino terminal peptide. Such peptides, when further digested with trypsin, were labelled 't' and numbered in the same way, e.g. a2t1, a2t2, etc. Peptides a1t1 through a1t5 were derived from a1 which was blocked at lysines before tryptic digestion. Peptides obtained from whole H4 by chymotryptic digestion were labelled c1 through c7.

Separation of mixtures of peptides

Limited acetic acid hydrolysis of *Physarum* H4 yielded after gel filtration a mixture of peptide a1 with the peptide equivalent to calf H4 peptide 69 – 102, as described previously for calf H4 [17]. Peptide a1 (205 nmol) was purified to homogeneity with a recovery of 85 % by a single pass of the peptide mixture in 3 ml 4 M ammonium acetate/5 % acetic acid over Octyl-Sepharose (1.5 × 15 cm) (Pharmacia) in the same solution at 20 ml/h. Peptide a1 was recovered from the flow through fractions by lyophilization. This method was developed for the purification of peptide 1 – 23 from calf H4 by M. Stolowitz (unpublished results).

Extensive chymotryptic digestion of *Physarum* H4 (Fig. 3B) yielded many peptides that were not separated by gel filtration on Sephadex G50-Superfine. These fractions were lyophilized, taken into 0.01 – 0.05 ml water und spotted onto cellulose thin layers (Polygram CEL 300, 0.1 mm, Macherey-Nagel & Co.). The thin layers were developed by ascending chromatography in pyridine/*n*-butanol/acetic acid/water (50:75:15:60, by vol.) [18]. After extensive drying, peptides were detected by fluorescamine and the spots were scraped off. The peptides were eluted from the cellulose into 5.7 M HCl/0.02 % 2-mercaptoethanol and subjected to amino acid hydrolysis.

Tryptic digestion of peptide a1 and a2

Peptide a1 (140 nmol), obtained from H4 labelled in S phase at 15500 counts × min^{-1} × mg^{-1}, was modified by maleylation according to Glazer et al. [19] at 70 nmol/ml on ice in 0.2 M Na_2HPO_4 by slow addition of a 50-fold molar excess over lysines of solid maleic anhydride while the pH was maintained between 8 and 9 by the addition of solid Na_2CO_3 followed by 60 min stirring on ice. The modified peptide was desalted on Sephadex G15 (1.0 × 46 cm), equilibrated with 0.1 M *N*-methylmorpholine acetate pH 8.0, at 36 ml/h and recovered from the void volume fractions by lyophilization. Modified a1 (115 nmol) was digested at 200 nmol/ml in 0.1 M *N*-methylmorpholine-acetate pH 8.0 with 1/100 of the weight in trypsin treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone added at the beginning and after one hour during a 4-h incubation at 37 °C. The reaction was stopped by the addition of concentrated acetic acid to pH 3.0. The peptide was then incubated for 48 h at 37 °C in a sealed tube to be demaleylated [19], and lyophilized. The peptides were fractionated on Sephadex G15 (1.0 × 46 cm), equilibrated with 5 % acetic acid, at 2.7 ml/h (Fig. 2). The fractions were pooled, lyophilized and chromatographed on thin layers as described above for chymotryptic peptides. Peptides were located by fluorescamin staining and then the thin layer was impregnated with PPO in ether [20]. Pre-flashed Kodak XAR-5 film was exposed to the thin layer at −70 °C for 15 days, and the PPO was removed from the thin layer by three washes with ether prior to peptide extraction for amino acid analysis.

Peptide a1 (200 nmol), obtained from H4 labelled in G2 phase at 13000 counts × min^{-1} × mg^{-1}, was modified with acetic anhydride exactly as described above for maleic anhydride, desalted, lyophilized and digestion with two aliquots of trypsin, 1/50 by weight. The peptide mixture was chromatographed on Sephadex G15 (1.0 × 46 cm), equilibrated with 5 % acetic acid, at 2.7 ml/h. The labelled void volume fractions containing peptide a1t2 were pooled and lyophilized prior to amino acid analysis and manual sequencing.

Peptide a2 (1.75 mg) was solubilized at 2 mg/ml in 8 M urea/0.1 M *N*-methylmorpholine acetate pH 8.0 and diluted to 0.5 mg/ml with urea-free buffer. It was digested with two aliquots of trypsin, 1/100 by weight, as described for peptide a1, and lyophilized. Partial peptide separation was achieved on Sephadex G25-Fine (1.1 × 50 cm), equilibrated with 50 mM HCl, at 3 ml/h. Fractions were lyophilized and chromatographed on cellulose thin layers as described above. The peptides, visualized by fluorescamine, were eluted and analyzed for amino acid composition.

Amino acid analysis

Peptides, produced from *Physarum* H4 as described above, were hydrolyzed *in vacuo* in 5.7 M HCl/0.02 % (w/v) 2-mercaptoethanol at 110 °C for 20 h. The dried hydrolyzate was analyzed on a Durrum D-500 amino acid analyzer. To selected samples monomethyllysine, dimethyllysine (Cyclo Chemical Co.) or trimethyllysine (Calbiochem) were added to identify the elution position of modified lysine residues in the D-500 analysis system, as previously described by Motojima and Sakaguchi [21].

Amino acid sequencing

Peptide a1t2 (55 nmol), produced from H4 labelled in G2 phase, with a specific activity of 135 counts × min^{-1}

$\times$ nmol^{-1} peptide, was subjected to manual liquid-phase Edman degradation with phenylisothiocyanate using the protocol described for 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate micromanual sequencing [22], omitting the reaction with 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate. For each degradation cycle the extracted thiazolinone derivative in *n*-butanol was dried under oil pump vacuum and the radioactivity was determined by liquid scintillation counting.

Peptides a3 (123 nmol) and a4 (152 nmol) were sequenced completely using automated spinning cup sequencing methods employing polybrene [23]. Residues were identified as their phenylthiohydantoin derivatives on high-performance liquid chromatography according to Bhown et al. [24], and in addition by either two-dimensional thin-layer chromatography according to Kulbe [25] or by gas-liquid chromatography according to Pisano et al. [26]. Both peptides were essentially free of aspartic acid at their aminoterminal position. Both sequences were determined for two steps past the carboxyterminal of the peptides.

RESULTS

Partial amino acid sequence

The amino acid composition of *Physarum* H4 is very similar to that of calf histone H4 [15]. The patterns of digestion of these H4 histones by chymotrypsin are also very similar, as judged by gel electrophoresis of the peptides [15]. Hydrolysis of calf histone H4 at aspartic acid residues by acetic acid, as described by Lewis et al. [17], gives upon peptide separation on Sephadex G50 a very distinctive pattern of four limit peptides and several incompletely hydrolyzed fragments of H4. Fig. 1A shows that a similar digestion of *Physarum* H4 gave the same characteristic pattern. The more complete hydrolysis, as suggested for calf H4 by Cary et al. [13], gave near quantitative yields of the four limit peptides from *Physarum* H4 (Fig. 1B). These overall similarities confirm previous identifications of *Physarum* H4 [27,28,15]. We have taken these similarities together with the known sequence conservation of histone H4 as the basis for assigning peptides to positions within the full sequence.

The acetic acid peptides (a1 through a4) were isolated, as described in detail in Materials and Methods, by chromatography on Sephadex G50-Superfine (Fig. 1) and subjected to amino acid analysis. The amino acid compositions of these peptides are given in Table 1. These data give unambiguous compositions for the three smaller peptides, except for aspartic acid values. Aspartic acid residues remain partially on carboxy terminal and to a lesser degree on amino terminal sides of a hydrolyzed peptide bond and they are to a certain degree lost completely especially in more extensive hydrolyses [22]. The carboxy terminal peptide from *Physarum* H4, a4, was identical in amino acid composition to the equivalent peptide from calf. Automated liquid phase amino acid sequencing confirmed the complete sequence identity of peptide a4 with calf H4 residues 85 through 102 (Table 3). The next peptide, a3, showed small differences from the corresponding calf peptide. These were: (a) one less lysine residue, (b) one more arginine residue, and (c) partial methylation of lysine. Automated sequence determination of all residues of peptide a3 confirmed that this peptide was identical to the calf histone H4 residues 68 through 85, except for a change of residue 77 from lysine in calf H4 to arginine in *Physarum* H4, and the presence

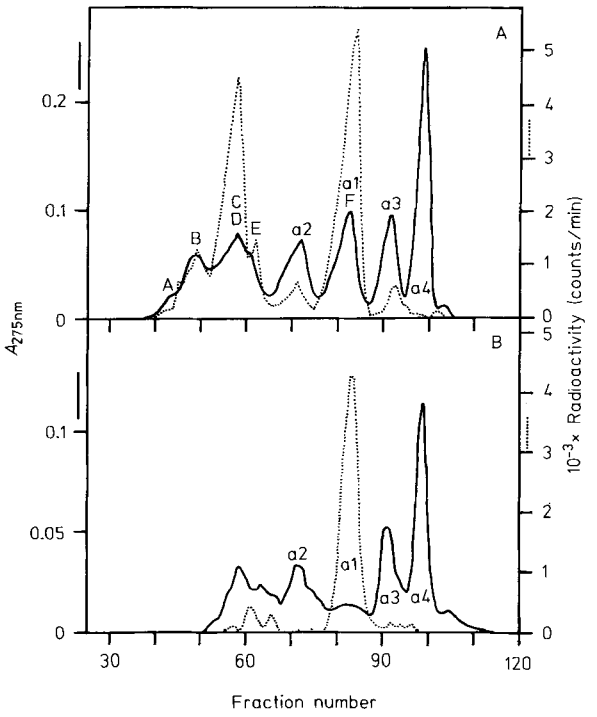


Fig. 1. *Sephadex G50 chromatography of* Physarum *H4 hydrolyzed with acetic acid.* (A) 5.9 mg H4, S phase labelled at 15500 counts $\times$ min^{-1} $\times$ mg , was hydrolyzed in acetic acid for 6 h at 109 °C. (B) 2.5 mg H4, G2 phase labelled at 13000 counts $\times$ min^{-1} $\times$ mg^{-1}, was hydrolyzed in acetic acid for 34.5 h at 104 °C. a1 through a4 are defined limit peptides (Table 1) and A through F are incompletely digested fragments of H4. A: intact H4 (1–102); B: 1–84; C: 1–67; D: 25–102; E: 25–84; F: 69–102. Hydrolysis and peptide separation conditions are described in detail in Materials and Methods

Table 1. *Amino acid composition of selected peptides*

Amino acid	Number in peptide					
	a1 1–23/24	a2 24/25–67/68	a3 68/69–84/85	a4 85/86–102	a1t2 4–17	c1 1–37
Asx	0.60	3.18	0.48	0.16	0.17	2.03
Thr	0.05	3.78	3.66	1.11	<0.10	1.02
Ser	0.87	1.13	<0.01	0.10	0.21	1.24
Glx	0.07	3.78	1.24	1.10	0.18	1.23
Pro	<0.02	1.07	0.05	0.03	<0.10	1.01
Gly	8.28	4.24	0.23	3.85	7.34	9.73
Ala	1.05	2.23	2.79	1.08	1.07	2.17
Val	1.01	2.65	1.99	1.30	<0.01	1.07
Met	<0.02	0.09	0.72	0.05	<0.10	<0.05
Ile	0.02	4.84	0.14	0.06	<0.10	2.68
Leu	1.93	3.06	0.23	2.01	1.00	2.72
Tyr	<0.02	1.13	0.99	1.91	<0.10	<0.05
Phe	<0.02	1.00	0.10	0.96	<0.10	<0.05
His	0.85	0.24	0.94	0.06	0.17	1.00
Lys	5.13	4.18	1.37[a]	1.21	3.84	5.83
Arg	4.02	6.90	2.18	2.12	0.96	5.71
Total	23.9	43.5	17.1	17.1	14.9	37.4

[a] Assuming identical color constants for lysine and its monomethyl, dimethyl and trimethyl derivatives.

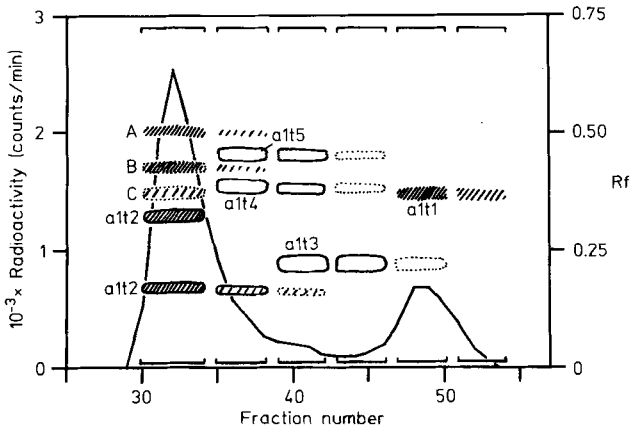


Fig. 2. *Separation of tryptic peptides from lysine-blocked a1.* S phase labelled peptide a1, trypsin digested and demaleylated, was fractionated on Sephadex G15 followed by thin-layer chromatography. Peptides detected on the thin layer are outlined by a solid line (heavy spots) or a broken line (faint spots). Unlined spots were not detected by fluorescamine. Radioactive peptides, detected by fluorography, are indicated by cross hatching simulating differences in amounts of radioactivty detetected. The amounts of peptides recovered from the spots A, B and C were not sufficient for unambiguous identification by amino acid analysis. R_f indicates the relative mobility of the peptides relative to the running front. Fraction size: 0.45 ml

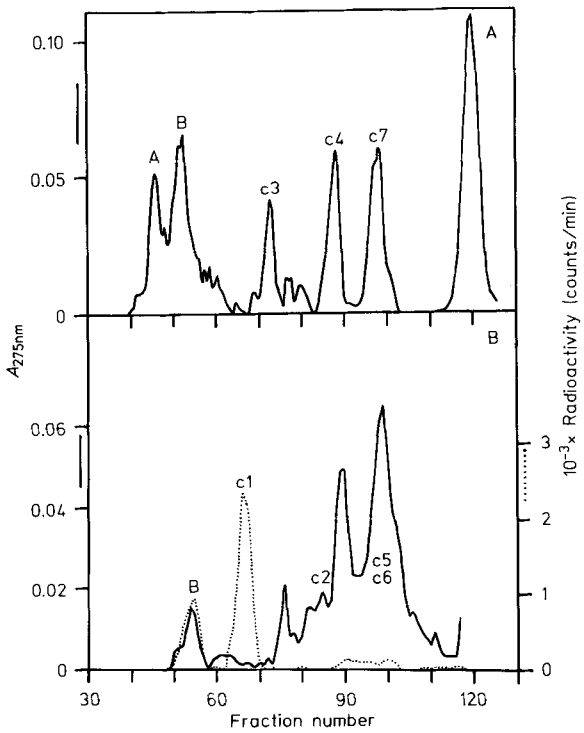


Fig. 3. *Sephadex G50 chromatography of* Physarum *H4 digested with chymotrypsin.* (A) 4.6 mg H4 was digested with chymotrypsin at a weight ratio of 1:250. (B) 2.2 mg H4, G2 phase labelled at 6000 counts × min^{-1} × mg , was digested with chymotrypsin at a weight ratio of 1:50. c1 through c7 are peptides identified by amino acid analysis (Table 2). A: undigested H4 (1 – 102). B: partially digested H4, peptide 1 – 72. Fraction size: 0.8 ml. Digestion and peptide separation are described in detail in Materials and Methods

of several forms of lysine at residue 79 differing in their degree of methylation.

Peptide a2 was too large (about 44 amino acids) and contained too many slowly hydrolyzing hydrophobic residues to give an unambiguous amino acid composition (Table 1). It was further digested with trypsin and the peptides were separated by gel filtration followed by thin-layer chromatography. The major peptides were eluted from the thin layer and subjected to amino acid hydrolysis. The results are summarized in Table 2 which shows that tryptic peptides a2t1 through a2t6 correspond to the peptides expected from the amino terminus of a2 if its sequence is identical with that of residues 25 through 45 of calf H4. No tryptic peptide(s) corresponding to the expected carboxy terminus of a2 were isolated, possibly due to insolubility or hydrophobic adsorption and loss. Some loss clearly occurred because, for example, the phenylalanine residue seen in the amino acid composition of a2 was not found in the tryptic peptides. Peptide a2t7 has an amino acid composition similar to but not identical with the expected peptide corresponding to the region 46 through 55 of H4. It lacks glycine and leucine but contains asparagine and threonine residues. These may be the amino acids observed at residues 48 and 49 that are also variant from calf H4 in *Tetrahymena* [6,7]. Peptide a2t8 corresponds to the expected peptide 56 – 59. Data displayed in Table 2 suggest that most of the residues observed in calf H4 sequence 60 through 67 are also present in *Physarum* H4. However, some differences are possible within this region of *Physarum* H4 compared to calf. Table 3 summarizes our current model for the sequence of *Physarum* H4 in this region with the sequence from residue 46 through 59 containing two possible variations from calf H4 and the sequence from 60 through 67 unknown.

Peptide a1 corresponded to the amino terminus of H4. It displayed heterogeneity on acid/urea gels based on differences in the extent of acetylation of lysines, as also seen for intact H4 [15]. In chromatography (Fig. 1) it was detected by the presence of label in acetylated amino acids since it is undetectable by absorbance measurements at 275 nm because it lacks tyrosine or phenylalanine residues. The amino acid composition of peptide a1, with a reduced yield of carboxy-terminal aspartic acid (Table 1) was identical to the corresponding aminoterminal peptide of calf H4, residues 1 through 24. This identity was further confirmed by tryptic digestion at arginines of peptide a1 in which all lysines were modified, followed by separation of the resulting peptides (Fig. 2). Each of the expected peptides a1t1 (1 – 3), a1t2 (4–17), a1t3 (18–19), a1t4 (20–23), and a1t5 (20–24) was isolated and its identity confirmed by amino acid analysis (Table 2). The full amino acid composition of a1t2 (4 – 17) is given in Table 1 because of the importance of this region with respect to acetylation sites.

The results were further confirmed by isolating seven chymotryptic peptides c1 through c7 (Fig. 3) and determining their amino acid compositions (Table 2). The full amino acid analysis data are given in Table 1 for peptide c1 (1 – 37) which contained all the sites labelled with [^{3}H]acetate (Fig. 3). These data support the conclusions obtained with the acetic acid peptides, summarized in Table 3.

Acetylation sites

Physarum histone H4 was labelled *in vivo* with [^{3}H]acetate during a 5-min pulse in S phase of the cell cycle. Under these conditions the majority of the label appears in acetate groups, both in ε-*N*-acetyllysine and in α-*N*-acetylserine, while part of the label is metabolized and appears in internal positions in amino acids [14]. Labelled H4 was partially digested with

Table 2. *Amino acid composition of all peptides*
Peptides a2t3/a2t4 and a2t5/a2t6 were peptides differing in R_f values after thin layer chromatography. Sequence 60–67 in calf H4: Asx Glx Val_2 Ile Leu Phe Arg. In *Physarum* H4 deduced from a2 minus a2t2, a2t4, a2t6, a2t7 and a2t8 peptides: Thr Glx $Val_{1(2)}$ Ile Leu Lys Arg. In *Physarum* H4 deduced from c3 minus c2, residue 59 from at28 and residues 68 through 72 from the sequenced amino terminal part of peptide a3: Asx Glx Val_{1-2} Ile Leu Phe Lys Arg

Peptide name	Number of residues	Amino acid composition	Differences between H4 peptides of *Physarum* and calf thymus	Sequence
a1	23.6	$Asx_{0.6}$ Ser Gly_8 Ala Val Leu_2 His Lys_5 Arg_4	none	1--23/24
a2	43.0	$Asx_{3.0}$ Thr_4 Ser Glx_4 Pro Gly_4 Ala_2 Val_3 Ile_5 Leu_3 Tyr_1 Phe_1 Lys_4 Arg_7	unknown	24/25--67/68
a3	16.5	$Asx_{0.5}$ Thr_4 Glx Ala_3 Val_2 Met Tyr His Lys^a Arg_2	+1 Arg -1 Lys	68/69--84/85
a4	17.2	$Asx_{0.2}$ Thr Glx Gly_4 Ala $Val_{1(2)}{}^a$ Leu_2 Tyr_2 Phe Lys Arg_2	none	85/86--102
a1t1	3	Ser Gly Arg	none	1--3
a1t2	14	Gly_7 Ala Leu Lys_4 Arg	none	4--17
a1t3	2	His Arg	none	18--19
a1t4	4	Val Leu Lys^b Arg	none	20--23
a1t5	5	Asx Val Leu Lys^b Arg^a	none	20--24
a2t1	11	Asx Thr Glx Pro Gly Ala Ile_3 Lys Arg	none	25--35
a2t2	12	Asx Thr Glx Pro Gly Ala Ile_3 Lys Arg_2	none	25--36
a2t3	4	Ala Leu Arg_2	none	36--39 or 37--40
a2t4	4	Ala Leu Arg_2	none	36--39 or 37--40
a2t5	5	Gly_2 Val Lys Arg	none	40--44 or 41--45
a2t6	5	Gly_2 Val Lys Arg	none	40--44 or 41--45
a2t7	10	Asx Thr_2 Ser Glx_2 Ile_2 Tyr Arg	+1 Asx -1 Gly +1 Thr -1 Leu	46--55
a2t8	4	Gly Val Leu Lys	none	56--59
c1	37	Asx_2 Thr Ser Glx Pro Gly_9 Ala_2 Val Ile_3 Leu_3 His Lys_6 Arg_6	none	1--37
c2	19-21	Asx Thr_2 Ser Glx_2 Gly_3 Ala Val_{1-2} Ile_2 Leu Tyr Lys Arg_{3-4}	+1 Asx -1 Gly +1 Thr -1 Leu	38--58
c3	33-36	Asx_3 Thr_3 Ser Glx_3 Gly_3 Ala_2 $Val_{3(4)}$ $Ile_{2(3)}$ $Leu_{2(3)}$ Tyr_2 Phe Lys_3 Arg_5	unknown	37--72
c4	16	Asx Thr_3 Glx Ala_2 Val_2 Met Tyr_2 His Lys^c Arg_2	+1 Arg -1 Lys	72--88
c5	12	Asx Thr Glx Gly Ala Val Leu_2 Tyr Lys Arg_2	none	85--97
c6	10	Thr Glx Gly Ala Leu_2 Tyr Lys Arg_2	none	88--97
c7	13	Thr Glx Gly_2 Ala Leu_2 Tyr_2 Phe Lys Arg_2	none	8--100

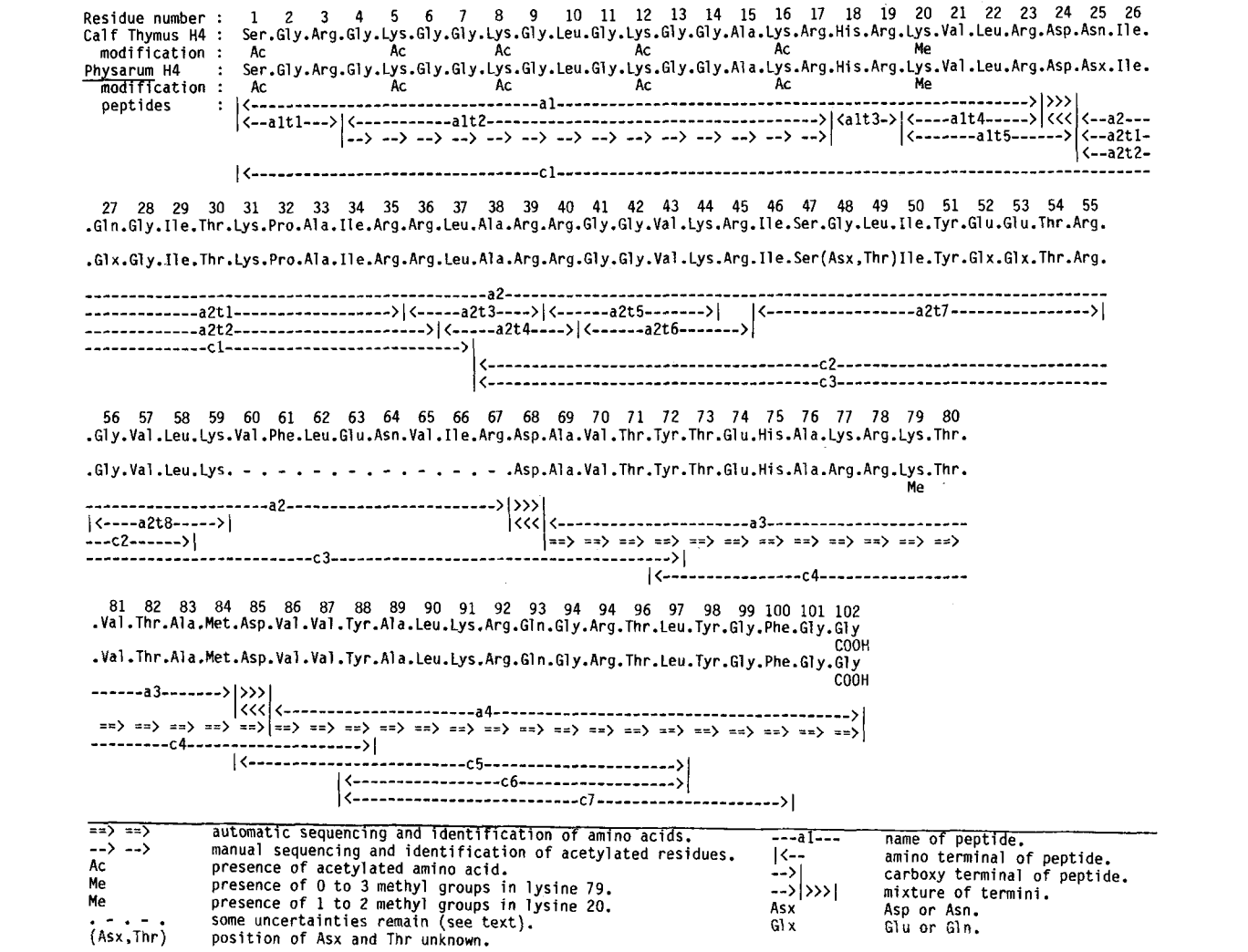
[a]Low value due to incomplete hydrolysis of Val(86)-Val(87), as previously observed by Ogawa et al. [29] and DeLange et al. [43].

[b]Lysine is dimethyllysine and some monomethyllysine.

[c]Lysine is a mixture of trimethyl, dimethyl and monomethyllysine with unmodified Lysine.

acetic acid, as described above, and Fig. 1A shows that ^3H label was associated with most of the peptides but that the major labelling occurred in the peaks containing large partial digest peptides and in the peak containing peptide a1 (1–23) and the partial digest peptide F. Peptide a1 was separated from peptide F by chromatography on Octyl-Sepharose, its lysines were blocked by maleylation and it was digested with trypsin. After incubation to regenerate free lysines the peptides, a1t1 through a1t5, were separated as shown in Fig. 2 and identified by amino acid composition. The multiple forms of a1t2

Table 3. *Histone H4 sequence*



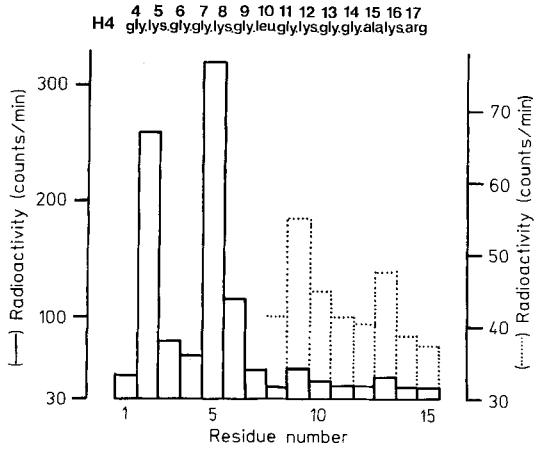
probably differ in their degree of remaining lysine maleylation. ^{3}H label was found in peptides a1t1 and a1t2. The amino acid composition of a1t1 was (Ser, Gly, Arg). In a subsequent experiment, in which *Physarum* H4 was labelled during G2 phase (and in the presence of cycloheximide) the peptide a1t1 was not labelled. The data are fully consistent, with the interpretation that peptide a1t1 is the amino terminus of *Physarum* histone H4, with the sequence Ac-Ser-Gly-Arg as in most other histones H4.

The locations of acetylation sites in peptide a1t2 were determined using a different preparation of *Physarum* histone H4 which was labelled with [^{3}H]acetate in G2 phase in the presence of cycloheximide. The labelled H4 was hydrolyzed almost to completion with acetic acid, yielding one major labelled peptide, a1 (1 – 23), as shown in Fig. 1B. The lysines of peptide a1 were blocked with acetic anhydride and the peptide was digested with trypsin. The digest was chromatographed on Sephadex G15 and the excluded peak was used for sequencing by manual Edman degradation. Fig. 4 shows the amount of label recovered at each degradation step in the thiazolinone-derivatized amino acid. Label was released after cycles 2, 5, 9 and 13 corresponding to residues 5, 8, 12 and 16 of intact H4. These results identify the sites of acetylation in *Physarum* H4 but do not provide a quantitative measure of the acetate turnover rate at each site because of variable yields during the Edman degradation cycles. The sites are the same as those previously identified in trout testis H4 [9].

Methylation sites

The lysine at position 20 is methylated in calf histone H4 but not in pea histone H4. The amino acid analysis of peptides alt4 and alt5 (Fig. 2) showed the presence of a mixture of dimethyllysine and probably lower amounts of monomethyllysine.

One other methylation site has been located in *Physarum* H4 at a position, lysine 79, not reported to be modified in any other species. Fig. 5 shows the lysine region of the amino acid analysis of peptide a3 (residues 68/69 through 84/85) from *Physarum* H4 and from peptide 73 through 88 produced from calf H4 by chymotryptic digestion [1]. Note the absence of methylated lysine in the calf peptide and the mixture of forms of lysine with 0 through 3 methyl groups in the *Physarum* one. Sequencing of peptide a3 gave at step 11, corresponding to lysine 79, four different phenylhydantoinyl amino acid derivatives. These were separated by high-performance liquid chromatography [24] and identified by co-chromatography with phenylhydantoinyl derivatized lysine with 0 through 3



Fi.g. 4. *Acetylated residues in peptide a1t2 (4–17).* Peptide a1t2 (55 nmol) obtained from H4 labelled in G2 phase and fully acetylated with acetic anhydride, was manually sequenced and the label liberated at each step was determined. For comparison purposes the known sequence of calf H4 residues 4 through 17 is included in the figure

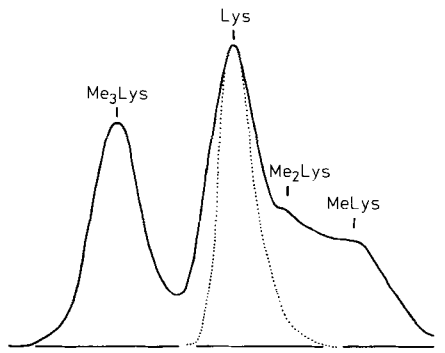


Fig. 5. *Methylation of lysine 79 in histone H4.* Amino acid analysis of peptide a3 on the Durrum D-500 column showed multiple peaks of lysine that were identified by co-chromatography with known methylated lysines as trimethyllysine, unmodified lysine, dimethyllysine, and monomethyllysine. For comparison the hydrolysis pattern of calf H4 chymotryptic peptide 73–88 is included. The ordinate shows elution time and the abscissa is absorbance after staining with ninhydrin. The elution time axis is the same for both samples but the absorbance scales are different. For a given amount of peptide hydrolyzate, the absolute area under the calf thymus lysine peak was two times the absolute area under the *Physarum* lysines

methyl groups. The approximate amounts present in lysine and its monomethyl, dimethyl and trimethyl derivatives were 12 %, 25 %, 20 % and 43 %, respectively.

DISCUSSION

Complete histone H4 sequences based on sequencing of overlapping peptides have only been determined for calf [1,29] and pea [2]. For the sequence determination of histone H4 in many organisms, amino acid analysis of peptides in combination with positioning of peptides on the basis of homology with calf H4 has been used, often supported to a limited extent by some overlapping peptides and sequencing of selected peptides for a few residues. This approach has been employed for the amino acid sequence determination of H4 from human spleen [7], pig [30], rat [31] Novikoff hepatoma [32], cuttle fish [33], sea urchin [4,5], *Tetrahymena* [6,7] and yeast [8]. For *Physarum* H4 we used this approach also and obtained sequencing data and amino acid compositions of overlapping peptides comparable to the studies cited above. For sea urchin and *Xenopus* H4 the amino acid sequence has been determined from nucleotide sequencing data [34–37]. This approach has not been employed for *Physarum* but hybridization under stringent conditions of sea urchin H4 to *Physarum* histone H4 genes [38] supports the high degree of sequence conservation for *Physarum* H4 compared to higher eukaryotes. The presence of irreversible acetylation of the aminoterminal serine residues in H4 and the precise location of the four possible sites for internal acetylation of lysine residues has been reported for H4 of trout [9,39], calf [1,29], pea [2], *Tetrahymena* [6], sea urchin [5], human spleen [7] and chicken and cuttle fish [33]. Complete methylation to dimethyllysine and monomethyllysine has been observed for lysine 20 in H4 of calf [1,29], chicken [40], Novikoff hepatoma [32] and human spleen [7]. Partial methylation occurs in trout H4 [39] while lysine 20 in pea has been reported to be completely unmodified [2]. Yeast and *Tetrahymena* H4 with 10 and 20 variant amino acids, compared to calf H4, are clearly more distantly related to calf and pea than *Physarum*. Compared to calf and pea, residues 48 and 49 are probably variant in *Physarum* while residue 77 is, like pea, arginine. The unknown sequence 60 through 67 may contain one to three variant residues (Table 2), which could include residues 60, 64 and 66 variant in pea, yeast and *Tetrahymena*[2,7,8].

We are particularly interested in the modification sites of histone H4 since *Physarum* is being used as a model system for studies of histone acetylation, particularly using the naturally occurring nuclear division cycle [14, 41] (unpublished results). We have shown that the sites of modification occur at positions 5, 8, 12 and 16, i.e. on the four lysines nearest the amino terminus of H4. The amino terminal serine is also *N*-acetylated, giving a blocked amino terminus. Thus the amino terminal region from residue 1 through 23 appears to be identical between *Physarum* and calf. Previously we have used acetylated peptide 1–23 from calf H4 as a substrate for *Physarum* histone deacetylase studies in the cell cycle of *Physarum* [41,42] on the assumption, now more clearly established, that the equivalent peptides would be identical in both organisms.

In further studies on histone acetylation [14] we have shown different patterns of histone acetylation correlated with transcriptionally active chromatin in G2 phase and replicative chromatin in S phase. It has been suggested that the acetylation of amino terminal lysines to multi-acetylated levels (3–4 acetates per H4) will abolish interaction between this part of histone H4 and DNA phosphates [13] allowing the chromatin to maintain a more open structure in which transcription could take place. The acetate turnover in G2 phase on H4 displayed the expected emphasis for multi-acetylation. Replication is associated with acetylation of newly synthesized H4 to 1 or 2 acetates per molecule (unpublished results).

The function of methylation at lysine-20 is unknown. It occurs in a very basic sequence, -Lys-Arg-His-Arg-Lys-, in which the first lysine, lysine-16, can be acetylated. The second methyllysine, so far only observed in *Physarum*, at position 79 also occurs together with other basic residues in the sequence -Arg-Arg-Lys- which is part of a larger hydrophilic sequence -Thr-Glu-His-Ala-Arg-Arg-Lys-Thr-. It has been stated that histidine-75 is subject to phosphorylation by a protein kinase

from rat tissues [44]. These post-translational modifications and the hydrophilic character of this sequence suggest that it occurs on the outside of the globular domain of H4 and that it may be involved in binding DNA to the core particle.

We thank Drs E. M. Bradbury, L. M. Mende, and M. L. Stolowitz for their continuous support and ideas during the course of this study. We would like to thank S. Van Patten and D. A. Walsh for their willingness to share the amino acid analysis data of calf H4 chymotryptic peptide 73–88 with us. We thank Mr A. J. Smith for the determinations of amino acid composition, the automated sequencer runs and his help with identifying the methyllysines. This study was supported by the National Institute of Health (GM 26901 and GM 30917). JHW was supported by a North Atlantic Treaty Organisation fellowship from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) during part of this project.

REFERENCES

1. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1969) *J. Biol. Chem. 244*, 319–334.
2. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1969) *J. Biol. Chem. 244*, 5669–5679.
3. von Holt, C., Strickland, W. N., Brandt, W. F. & Strickland, M. S. (1979) *FEBS Lett. 100*, 201–218.
4. Strickland, M., Strickland, W. N., Brandt, W. T. & von Holt, C. (1974) *FEBS Lett. 40*, 346–348.
5. Wouters-Tyrou, D., Sautiere, P. & Biserte, G. (1976) *FEBS Lett. 65*, 225–228.
6. Glover, C. V. C. & Gorovsky, M. A. (1979) *Proc. Natl Acad. Sci. USA, 76*, 585–589.
7. Hayashi, T., Ohe, Y., Hayashi, H. & Iwai, K. (1982) *J. Biochem. (Tokyo) 92*, 1995–2000.
8. Brandt, W. F., Patterson, K. & von Holt, C. (1980) *Eur. J. Biochem. 110*, 67–76.
9. Candido, E. P. M. & Dixon, G. H. (1971) *J. Biol. Chem. 246*, 3182–3188.
10. Marushige, K. & Dixon, G. H. (1969) *Dev. Biol. 19*, 397–414.
11. Allfrey, V. G. (1977) in *Chromatin and Chromosome Structure* (Li, H. J. & Eckhardt, R. A., eds), pp. 167–191, Academic Press, New York.
12. Allan, J., Harborne, N., Rau, D. C. & Gould, H. (1982) *J. Cell Biol. 93*, 285–297.
13. Cary, P. D., Crane-Robinson, C., Bradbury, E. M. & Dixon, G. H. (1982) *Eur. J. Biochem. 127*, 137–143.
14. Waterborg, J. H. & Matthews, H. R. (1983) *Biochemistry, 22*, 1489–1496.
15. Mende, L. M., Waterborg, J. H., Mueller, R. D. & Matthews, H. R. (1983) *Biochemistry, 22*, 38–51.
16. Reference deleted.
17. Lewis, P. N., Bradbury, E. M. & Crane-Robinson, C. (1975) *Biochemistry, 14*, 3391–3400.
18. Mende, L. M. (1979) Ph. D. Thesis, Berlin, FRG.
19. Glazer, A. N., DeLange, R. J. & Sigman, D. S. (1976) in *Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S. & Work, E., eds), pp. 1–204, North Holland, Amsterdam.
20. Randerath, K. (1970) *Anal. Biochem. 34*, 188–205.
21. Motojima, K. & Sakaguchi, K. (1981) *FEBS Lett. 132*, 334–336.
22. Allen, G. (1981) in *Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S. & Burdon, R. H., eds), pp. 1–327, North Holland, Amsterdam.
23. Niall, H. D. (1973) *Methods Enzymol. 27*, 942–1010.
24. Bhown, A. S., Mole, J. E., Weissinger, A. & Bennett, J. C. (1978) *J. Chromatogr. 148*, 532–535.
25. Kulbe, K. D. (1974) *Anal. Biochem. 59*, 564–573.
26. Pisano, J. J. & Bronzert, T. J. (1972) *Anal. Biochem. 45*, 43–59.
27. Mohberg, J. & Rusch, H. P. (1969) *Arch. Biochem. Biophys. 134*, 577–589.
28. Chahal, S. S., Matthews, H. R. & Bradbury, E. M. (1980) *Nature (Lond.) 287*, 76–79.
29. Ogawa, Y., Quagliarotti, G., Jordon, J., Taylor, C. W., Starbuck, W. C. & Busch, H. (1969) *J. Biol. Chem. 244*, 4387–4392.
30. Sautiere, P., Lambelin-Breynaert, M. D., Moschetto, Y. & Biserte, G. (1971) *Biochimie (Paris) 53*, 711–715.
31. Sautiere, P., Tyrou, D., Moschetto, Y. & Biserte, G. (1971) *Biochimie (Paris) 53*, 479–483.
32. Wilson, R. K., Starbuck, W. C., Taylor, C. W., Jordan, J. & Busch, H. (1970) *Cancer Res. 30*, 2942–2951.
33. Wouters-Tyrou, D., Martin-Ponthieu, A., Sautiere, P. & Biserte, G. (1981) *FEBS Lett. 128*, 195–200.
34. Moorman, A. F. M., De Boer, P. A. J., De Laaf, R. T. M., Van Dongen, W. M. A. M. & Destree, O. H. J. (1981) *FEBS Lett. 136*, 45–52.
35. Busslinger, M., Rusconi, S. & Birnstiel, M. L. (1982) *Eur. Mol. Biol. Org. J. 1*, 27–33.
36. Grunstein, M., Diamond, K. E., Knoppel, E. & Grunstein, J. E. (1981) *Biochemistry, 20*, 1216–1223.
37. Turner, P. C. & Woodland, H. R. (1982) *Nucleic Acids Res. 10*, 3769–3780.
38. Wilhelm, M. L., Wilhelm, F. X., Toublan, B. & Jalouzot, R. (1982) *FEBS Lett. 250*, 439–444.
39. Honda, B. M., Dixon, G. H. & Candido, E. P. M. (1975) *J. Biol. Chem. 250*, 8681–8685.
40. Bohm, L., Briand, G., Sautiere, P. & Crane-Robinson, C. (1981) *Eur. J. Biochem. 119*, 67–74.
41. Waterborg, J. H. & Matthews, H. R. (1982) *Exp. Cell Res. 138*, 462–466.
42. Waterborg, J. H. & Matthews, H. R. (1982) *Anal. Biochem. 122*, 313–318.
43. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1968) *J. Biol. Chem. 243*, 5906–5913.
44. Fujitaki, J. M., Fung, G., Ohe, Y. & Smith, R. A. (1981) *Biochemistry, 20*, 3658–3664.

J. H. Waterborg, Department of Biochemistry, School of Biological Sciences, University of Sussex, Biological Building, Falmer, Brighton, Sussex, Great Britain BN1 9QJ

S. R. Fried and H. R. Matthews, Department of Biological Chemistry, School of Medicine of the University of California at Davis, Davis, California, USA 95616