Precise elimination of the *N*-terminal domain of histone H1

Lothar BÖHM,*§ Pierre SAUTIERE,† Peter D. CARY‡ and Colyn CRANE-ROBINSON‡

*Department of Pharmacology, University of Stellenbosch Medical School, Cape Town, South Africa,

†Unité 124, INSERM, Lille, France, and ‡Biophysics Laboratories, Portsmouth Polytechnic, Portsmouth PO1 2DT, U.K.

(Received 25 November 1981/Accepted 22 January 1982)

The proteinase from mouse submaxillary gland was used to cleave total calf thymus histone H1 between residues 32 and 33. The *C*-terminal peptide, comprising residues 33 to the *C*-terminus, was purified and identified by amino acid analysis and Edman degradation. Spectroscopic characterization by n.m.r. for tertiary structure and by c.d. for secondary structure shows the globular domain of the parent histone H1 to be preserved intact in the peptide. It has therefore lost only the *N*-terminal domain and is a fragment of histone H1 comprising the globular plus *C*-terminal domains only. Precise elimination of only the *N*-terminal domain makes the fragment suitable for testing domain function in histone H1.

Structural studies of histone H1 from calf thymus, sea-urchin sperm and a fruit fly and of the homologous protein histone H5 from chicken erythrocytes indicate that in free solution these lysine-rich histones consist of three domains: an unstructured *N*-terminal sequence (approx. 35 residues in calf thymus H1), a folded central region that exhibits a high degree of sequence conservation (approx. 86 residues in calf thymus H1) and a long unstructured *C*-terminal domain of very basic composition comprising almost half the molecule (approx. 93 residues in calf thymus H1) (Chapman *et al.*, 1976; Hartman *et al.*, 1977; Puigdomenech *et al.*, 1980; Barbero *et al.*, 1980; Aviles *et al.*, 1978). By reconstituting several H1 peptides on to chromatin depleted of its original H1, it was shown that the central globular domain alone is able to close two full turns of nucleosomal DNA. It presumably thereby defines the exit points of the DNA strands from the core particle and also locates the H1 molecule (Allan *et al.*, 1980). But this does not explain how H1 is able to condense the nucleofilament into the thick fibre, particularly since the globular domain of H1 does not self-associate (E. Russo, V. Giancotti, S. Cosimi, C. Crane-Robinson & G. Geraci, unpublished work). Clearly one or both of the two terminal domains must be involved in the condensation process. To date we have been able to make a peptide consisting of the *N*-terminal plus globular domains (Chapman *et al.*, 1976), and a test of its ability to condense chromatin

Abbreviation used: SDS, sodium dodecyl sulphate.

§ To whom reprint requests should be addressed.

(Allan *et al.*, 1980) showed it to be only half as efficient as intact H1. The conclusion drawn was that the longer *C*-terminal domain is the principal agent of condensation, a role which fitted well with its high basicity. The next step was clearly to test an H1 peptide consisting of the globular plus *C*-terminal domains for its ability to condense chromatin, but no method was available to prepare such a fragment. This paper describes the production of such a fragment, together with its physical and biochemical characterization so as to be certain that the *N*-terminal domain has been essentially fully eliminated while the globular domain has remained intact. The ability of the fragment to condense chromatin requires further study. The availability of H1 fragments that consist of the globular domain linked to either the *N*-terminal or *C*-terminal domains will in the future allow a separate assessment of the role of phosphorylation sites in the *N*- and *C*-terminal domains. For these reasons the preparation of a peptide consisting of the globular and *C*-terminal domains is of first-order importance.

Mammalian H1 molecules appear to have three or four arginine residues. Those at positions 53 and 78 (Cole, 1977; R. D. Cole, personal communication) are highly conserved, even outside mammals (MacLeod *et al.*, 1977), and are within the folded globular domain. The remaining arginine residue is at the border between the *N*-terminal and globular domains: in rabbit thymus RTL-3 subfraction there is a single arginine at residue 34 (Cole, 1977), and in calf thymus CTL-1 subfraction there are two adjacent arginine residues at what seem to be homo-

logous positions 31 and 32. We have therefore used the proteinase from mouse submaxillary gland, which is said to be arginine-specific (Levy *et al.*, 1970), and digested H1 at pH 8.5 and at sufficiently high ionic strength to fold the globular domain in the hope that arginine-53 and -78 would be protected from digestion and cleavage would occur only at the *N*-terminal/globular-domain boundary. In fact, the enzyme preparation used exhibited a somewhat wider spectrum of activity, but nevertheless the required peptide, residues 37 to the *C*-terminus from calf thymus H1, was isolated from the digestion products. Physical characterization showed that the globular and tertiary structure of the central domain is preserved in the peptide, indicating that it contains just two intact domains.

Materials and methods

Calf thymus histone H1 was prepared from 0.15 M-NaCl-washed calf thymus tissue by extraction with 0.5 M-HClO₄. The extract was dialysed against 1 mM-HCl and finally against water and then freeze-dried. The crude H1 preparation was fractionated on Bio-Gel P-60 (Böhm *et al.*, 1973) to give a mixture of H1 subfractions, consisting of three major electrophoretic bands: CTL-3a + CTL-3b, CTL-2 and CTL-1, as indicated in Fig. 1, lane 3.

This H1 mixture, which was free of DNA, served as the substrate for proteolytic digestion. A solution of 140 mg of histone H1 in 20 ml of water was adjusted to pH 2 with 1 M-HCl. After centrifugation at 1000 rev./min for 10 min the clear supernatant was slowly neutralized with 3 M-NaOH, and 10 ml of 0.3 M-(NH₄)₂CO₃ buffer, containing 1 mg of glycine/ml, was added. The solution was then incubated for 90 min at 37°C with a total of 10 units of proteinase from the submaxillary glands of mice (Boehringer 269590). Proteolysis was terminated with phenylmethanesulphonyl fluoride. The solution was then made 8 M with respect to urea and adjusted to pH 2.8 with 5 M-HCl. It was then fractionated on Bio-Gel P-60 as described by Böhm *et al.* (1973). Fraction A I (Fig. 2a) was dialysed against 1 mM-HCl and freeze-dried. A solution of 19 mg of fraction A I in 5 ml of 0.05 M-sodium acetate, pH 5.8 (elution buffer), was then fractionated on Whatman CM-52 CM-cellulose, by eluting first with 50 ml of elution buffer followed by 200 ml of a 0.1–0.4 M-NaCl linear gradient and then 200 ml of a 0.4–0.6 M-NaCl linear gradient made up in elution buffer. Fraction B IV was dialysed and freeze-dried, and 0.2 mg of this fraction was subjected to amino acid analysis. Edman degradation was as previously described (Böhm *et al.*, 1981). SDS/polyacrylamide-gel electrophoresis was as described by Laemmli (1970).

High-resolution n.m.r. spectra were obtained on a

Bruker 270 MHz spectrometer by using protein concentrations of approx. 7 mg/ml in 5 mm tubes. C.d. spectra of the B IV peptides were obtained with a Cary 61 instrument by using a 10 mm cell and a protein concentration of approx. 0.07 mg/ml in 1 M-NaCl/25 mM-sodium phosphate buffer, pH 7.0. Protein concentrations for the c.d. measurement were obtained from amino acid analysis of the stock solution, with 2-aminoguanidinopropionic acid as internal standard.

Results and discussion

Identification of cleavage products

Digestion of calf thymus H1 with mouse submaxillary proteinase generates five major and eight to ten minor bands on SDS/polyacrylamide gels (Fig. 1, lane 4). Five major bands travel ahead of histone H5, but are slower than histone H3. Subfractionation of this digest on Bio-Gel P-60 (Fig. 2a) yields a mixture of high-molecular-weight

Fig. 1. SDS/polyacrylamide-gel electrophoresis of the fragments produced by proteolysis of calf thymus H1 with mouse submaxillary-gland proteinase

For full details see the text. Lanes 1 and 2, whole chicken erythrocyte standards. Lane 3, total calf thymus H1 (subfractions are indicated in lane 5). Lane 4, complete digest of total calf thymus H1 with submaxillary-gland proteinase. Lane 5, fraction A I isolated by chromatography on Bio-Gel P-60 (see Fig. 2a). Lanes 6–9, fraction B IV isolated from fraction A I by CM-cellulose ion-exchange chromatography (loadings 2, 5, 10 and 20 µg respectively).

Fig. 2. (a) Fractionation on Bio-Gel P-60 of total peptide mixture resulting from digestion of calf thymus H1 with submaxillary-gland proteinase, and (b) re-fractionation of fraction A I on Whatman CM-52 CM-cellulose. Elution was with 50mM-NaCl/20mM-HCl, pH 2.7, in (a), and with 50mM-sodium acetate buffer, pH 5.8, and an NaCl gradient (—) as illustrated in (b). For full details see the text. Shaded areas represent the fractions taken.

fractions in peak A I (Fig. 1, lane 5). Further fractionation of fraction A I on CM-cellulose provides a major fraction, B IV (Fig. 2b). SDS/polyacrylamide-gel electrophoresis shows that fraction B IV is free of undigested H1 and consists of two bands, which travel just ahead of H5 (Fig. 1, lanes 6–9). With histone H5, cytochrome c and chymotrypsinogen A as mobility standards it is estimated that the peptides in fraction B IV consist of approx. 175 residues, i.e. about 35 residues have been lost and differ by about six residues. The occurrence of arginine residues at positions 30–35 (depending on the subfraction) immediately suggests that cleavage has occurred at this point. However, the B IV fraction contains two peptides, and the enzyme has clearly cleaved at several points throughout the molecule, i.e. not only at arginine.

The amino acid composition of the B IV peptide mixture is given in Table 1. Comparison with that of the H1 mixture used in the digestion shows that the arginine content falls by 28% and that of glutamic acid by 17%. There is a smaller loss of alanine and proline. As expected, residues located exclusively in the central part of the chain increase their molar percentage (e.g. leucine, aspartic acid). The lysine content even increases slightly. These compositional data clearly indicate that the major loss of residues is at the N-terminus: only in this way could the fall in glutamic acid and arginine be accounted for.

The B IV peptide mixture was therefore subjected

Table 1. Amino acid composition of calf thymus H1 and its peptides

Results are from 22h hydrolysates, uncorrected for destruction of serine, threonine and tyrosine and for incomplete hydrolysis at valine, leucine and isoleucine. The H1 substrate used for fragmentation by submaxillary-gland proteinase was a mixture of subfractions CTL-1, CTL-2, CTL-3a and CTL-3b in the weight ratio 4:4:1:1, as estimated from gels. Thus only 40% of the peptides in the B IV mixture could actually be derived from CTL-1.

	Calf thymus H1 mixture used (mol %)	CTL-1 from sequence* (mol %)	CTL-1 (residues/molecule)	CTL-1 (33–210)		B IV peptide mixture (mol %)
				(residues/molecule)	(mol %)	
Asp	2.03	1.90	4	4	2.24	2.3
Thr	5.53	5.71	12	10	5.61	5.4
Ser	5.87	6.19	13	12	6.74	6.3
Glu	3.63	3.33	7	5	2.80	3.0
Pro	9.04	10.47	22	16	8.99	8.5
Gly	6.55	6.19	13	12	6.74	7.4
Ala	24.85	26.67	56	45	25.28	23.9
Val	5.28	5.24	11	10	5.61	5.0
Cys	—	0	0	0	0	—
Met	—	0	0	0	0	—
Ile	0.93	0.95	2	2	1.12	1.1
Leu	4.43	4.28	9	9	5.06	5.2
Tyr	0.55	0.48	1	1	0.56	0.5
Phe	0.63	0.48	1	1	0.56	0.6
Lys	28.89	26.19	55	49	27.52	29.5
His	—	0	0	0	0	—
Arg	1.77	1.90	4	2	1.12	1.2
Total no. of residues ...			210	178		

* R. D. Cole, unpublished work.

to automatic Edman degradation, and the results are given in Table 2. It is immediately apparent that the sequence is Lys-Ala-Ser-Gly-Pro-Val-Ser-Glu-Leu-Ile-Thr. The published sequence of peptide CTL-1 (Liao & Cole, 1981) shows that this corresponds to residues 33–44. The data of Table 2 also show the presence of a single *N*-terminal sequence. This indicates that the two peptides of the B IV mixture differ elsewhere in the molecule and is in accord with sequence data (R. D. Cole, unpublished work), which show that subfractions 1, 2, 3a and 3b from calf thymus are virtually identical over this region between residues 33 and 44. Thus they either represent different subfractions cut at the same point in the chain and/or a small amount of *C*-terminal loss. Although it is often thought (on the basis of partial sequence data) that H1 subfractions differ most characteristically at the *N*-terminus, it is found that the peptide mixture generated by thrombin digestion (Chapman *et al.*, 1976) (which represents residues 122 to the *C*-terminus from a mixture of fractions) consists of peptides with a wide spread of mobilities. The two

peptides of the B IV mixture could both be intact at the *C*-terminus. That the *C*-terminus is intact is shown by the high lysine content of the B IV mixture. Of the last 10 residues of subfraction CTL-1, six are lysine and the subfraction ends with the sequence Lys-Lys-Lys. With the proviso that the other subfractions might be different right at the *C*-terminus, on the basis of the CTL-1 sequence it seems that no *C*-terminal residues have been lost. (No sequence data are available for other calf H1 subfractions, but rabbit H1, subfraction RTL 3, also has six lysine residues in the last ten residues and is closely homologous to CTL-1 in this region.) Table 1 gives the molar composition and number of residues in intact CTL-1(1–210) and CTL-1(33–210). The molar composition of peptide CTL-1(33–210) is in close agreement with that of peptide mixture B IV. The significant discrepancies between the two are presumably a result of the subfraction heterogeneity in the B IV mixture.

Physical characterization of B IV peptides

Published data indicate that the globular domain of histone H1, which is protected from trypsin digestion, starts at about residue 36 (Hartman *et al.*, 1977). The sequence residues 33–210 should thus have lost almost all of the *N*-terminal domain, but the globular domain should be intact. This has been checked spectroscopically. In the high-resolution p.m.r. spectrum, perturbed resonances in both the high-field region of ring-current shifted methyl groups (+1 to –1 p.p.m.) and in the aromatic region (+6 to +9 p.p.m.) are characteristic of the tertiary folding of a protein. Fig. 3 compares these regions in denatured and native H1 with three peptides that comprise (a) the globular + *C*-terminal domains (GCH1), i.e. the B IV peptide mixture prepared here, (b) the *N*-terminal + globular domains (NGH1) and (c) the globular domain alone (GH1). Since all three peptides exhibit spectra perturbed identically with that for intact H1, we conclude that they all contain the same intact element of tertiary folding. Secondary structure can be probed by c.d., and the ellipticity at 222 nm is a good indicator of helicity, particularly in the absence of β -structure, as appears to be the case for histone H1 (Bradbury *et al.*, 1975). Table 3 gives the ellipticities at 222 nm of H1 and the three fragments GCH1, NGH1 and GH1. The helicity has been calculated by taking that of random coil as -1000° (Moss *et al.*, 1976) and a fully helical chain as -30000° (Chen *et al.*, 1974). The number of helical residues follows from the estimated chain lengths. Bearing in mind the considerable errors involved in both the measurement of ellipticity and the concentration of dilute protein solutions, all three peptides appear to contain the same amount of helix as the parent H1, i.e. the secondary structure is preserved

Table 2. *Edman degradation of B IV peptide mixture*
For details see the text. Approx. 100 nmol of mixture was sequenced. Abbreviations: N.Q., not quantified; R–1, previous residue; R–2, last but one residue etc.

Cycle no.		Residues	Yield (nmol)
1		Lys	78.7
2		Ala	66.0
	R–1	Lys	28.8
3		Ser	3.5
	R–1	Ala	26.9
4		Gly	51.6
	R–1	Ser	0.6
5		Pro	87.7
	R–1	Gly	18.8
6		Pro	116.5
	R–1	Gly	2.2
7		Val	N.Q.
	R–1	Pro	42.2
8		Ser	5.4
	R–1	Pro	8.3
9		Glu	N.Q.
	R–1	Ser	2.7
	R–3	Pro	5.0
10		Leu	48.9
	R–4	Pro	5.4
11		Ile	36.3
	R–1	Leu	28.4
		Pro	5.9
12		Thr	9.5
	R–1	Ile	15.0
	R–2	Leu	5.8
		Pro	7.5

Fig. 3. 270 MHz p.m.r. spectra of calf thymus H1 and its peptide fragments consisting of intact domains: GCH1, globular + C-terminal domains; NGH1, N-terminal + globular domains; GH1, globular domain alone. The spectrum of denatured H1 was obtained in ²HCl, pD2.0, with no added salt, and the other spectra were obtained in 25 mM-phosphate buffer, p²H(pD)7.0, with added 0.25 M-K₂SO₄ to induce folding of the globular domain. The aromatic spectrum (6.5–7.5 p.p.m.) comes from a single tyrosine and a single phenylalanine residue located at positions 70 and 104 respectively in subfraction CTL-1, i.e. within the globular domain. In the folded state the peaks at 6.7 and 6.9 p.p.m. can be assigned to tyrosine and those between 7.0 and 7.5 p.p.m. to phenylalanine (Bradbury *et al.*, 1975). Resonance at 7.7 p.p.m. is from exchangeable NH protons. Upfield spectra (+1.0 to –0.5 p.p.m.) show aliphatic methyl resonances, and all are presented with identical convolution difference parameters (Campbell *et al.*, 1973) to display the ring-current-shifted methyl groups labelled 1–6 (Chapman *et al.*, 1976). The GCH1 fragment was prepared as described in the Materials and methods section, and the NGH1 fragment was prepared as described by Campbell *et al.* (1976), by using thrombin. The globular domain, GH1, was prepared by using trypsin as in Hartman *et al.* (1977), but with the following modifications. The digestion was started by adding trypsin to an unbuffered solution of H1 in 1 M-NaCl at pH 10. The pH was monitored continuously during proteolysis and fell approximately linearly for 70 min, after which it decreased only slowly. This period of 70 min was assumed to be the time taken to digest away most of the N- and C-terminal domains to leave the resistant globular peptide. Digestion was allowed to proceed for a further 20 min before inhibition of trypsin, to ensure complete removal of the flanking domains. The GH1 fragment was then separated on Sephadex G50 in 10 mM-HCl to give a final yield of 27%.

Table 3. Secondary structure of calf thymus H1 and its fragments obtained from c.d. spectra
For details see the text.

Protein or fragment	Chain residues	No. of residues	θ_{222} (degrees·cm ² ·dmol ⁻¹)	Helix (%)	No. of helical residues	Reference
$N\bullet\text{---}C$ H1	1-210	210	-4700	12.6	26.8	Giancotti <i>et al.</i> (1981)
$\bullet\text{---}$ GCH1	33-210	178	-5100	14.1	25.2	The present work
$\text{---}\bullet$ NGH1	1-121	121	-7700	23.1	28.2	Chapman <i>et al.</i> (1976)
$\bullet$ GH1	36-121	86	-9000	27.6	23.2	Giancotti <i>et al.</i> (1981)

in all three peptides. Even if the method of calculating helicity is imprecise, the observed ellipticities vary essentially as expected for conserved secondary structure. Spectroscopic comparison of the new fragment GCH1 with intact H1 and with the other two fragments clearly indicates that the N-terminal domain has been effectively eliminated in GCH1 without affecting the globular domain. Since the C-terminal domain is also intact, the peptide is appropriate for further testing of the function of H1 domains.

L. B. gratefully acknowledges the technical assistance of Helgard Böhm and Tanya Mitchell and financial support of the Medical Research Council and National Cancer Association of South Africa. P. S. is supported by grants from the Institut National de la Santé et de la Recherche Médicale and from the CNRS. C. C. R. and P. D. C. are supported by the S.E.R.C. of Great Britain. We are particularly grateful to Professor R. D. Cole for communicating histone-H1 sequence data before publication.

References

- Allan, J., Hartman, P. G., Crane-Robinson, C. & Aviles, F. X. (1980) *Nature (London)* **288**, 675–679
- Aviles, F. J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C. & Bradbury, E. M. (1978) *Eur. J. Biochem.* **88**, 363–371
- Barbero, J. L., Franco, L., Montero, F. & Moran, F. (1980) *Biochemistry* **19**, 4080–4087
- Böhm, L., Strickland, W. N., Strickland, M., Thwaites, B. G., van der Westhuyzen, D. R. & von Holt, C. (1973) *FEBS Lett.* **34**, 217–221
- Böhm, L., Briand, G., Sautière, P. & Crane-Robinson, C. (1981) *Eur. J. Biochem.* **119**, 67–74
- Bradbury, E. M., Cary, P. D., Chapman, G. E., Crane-Robinson, C., Danby, S. E., Rattle, H. W. E., Boublik, M., Palau, J. & Aviles, F. J. (1975) *Eur. J. Biochem.* **52**, 605–613
- Campbell, I. D., Dobson, C. M., Williams, R. J. P. & Xavier, A. V. (1973) *J. Magn. Reson.* **11**, 172–181
- Chapman, G. E., Hartman, P. G. & Bradbury, E. M. (1976) *Eur. J. Biochem.* **61**, 69–75
- Chen, Y. H., Yang, J. T. & Chau, K. H. (1974) *Biochemistry* **13**, 3350–3359
- Cole, R. D. (1977) in *The Molecular Biology of the Mammalian Genetic Apparatus* (Ts'o, P., ed.), p. 93, Elsevier, Amsterdam
- Giancotti, V., Russo, E., Cosimi, S., Cary, P. D. & Crane-Robinson, C. (1981) *Biochem. J.* **197**, 655–660
- Hartman, P. G., Chapman, G. E., Moss, T. & Bradbury, E. M. (1977) *Eur. J. Biochem.* **77**, 45–51
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Levy, M., Fishman, L. & Schenkein, I. (1970) *Methods Enzymol.* **19**, 672–681
- Liao, L. W. & Cole, R. D. (1981) *J. Biol. Chem.* **256**, 3024–3029
- MacLeod, A. R., Wong, N. C. & Dixon, G. H. (1977) *Eur. J. Biochem.* **78**, 281–291
- Moss, T., Cary, P. D., Abercrombie, B. D., Crane-Robinson, C. & Bradbury, E. M. (1976) *Eur. J. Biochem.* **71**, 337–350
- Puigdomenech, P., Palau, J. & Crane-Robinson, C. (1980) *Eur. J. Biochem.* **104**, 263–270