The EMBO Journal vol.7 no.1 pp.69-75, 1988

# $\alpha$-Helix in the carboxy-terminal domains of histones H1 and H5

**David J.Clark, Caroline S.Hill, Stephen R.Martin[1] and Jean O.Thomas**

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW and [1]National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Communicated by J.O.Thomas

**Although the carboxy-terminal domains of histones H1 and H5 exist as random-coil in aqueous solution, secondary structure prediction suggests that this region has a high potential for $\alpha$-helix formation. We have measured CD spectra in various conditions known to stabilize $\alpha$-helices, to determine whether this potential can be realized in an appropriate environment. Trifluoroethanol increases the helix contents of H1, H5 and their carboxy-terminal fragments, presumably through promotion of axial hydrogen bonding. Sodium perchlorate is also effective and better than sodium chloride, suggesting stabilization by binding of bulky perchlorate ions rather than simple charge screening. Extrapolating from these measurements in solution, and taking into account the occurrence of proline residues throughout the carboxy-terminal domain, we propose that binding to DNA stabilizes helical segments in the carboxy-terminal domains of histones H1 and H5, and that it is this structured form of the domain that is functionally important in chromatin.**

*Key words:* Histones H1 and H5/C-terminal domain/$\alpha$-helix/circular dichroism

## Introduction

Histone H1 is required for the ordered, salt-dependent folding of the 10 nm chromatin filament to form the 30 nm filament *in vitro* (Thoma *et al.*, 1979) and is probably involved in the control of the accessibility of genes for transcription. Histone H5, an extreme variant of H1, is associated *in vivo* with the general repression of gene transcription during the maturation of avian erythrocytes (Neelin *et al.*, 1964; Johns, 1969) and *in vitro* with more stable 30 nm filaments (Bates *et al.*, 1981). Similarly, sperm H1 from sea urchin is associated with the complete repression of gene activity in mature sperm (see Poccia *et al.*, 1987 and references therein). The structures of H1 and H5 when bound to DNA in chromatin are therefore of central relevance.

H1 and all its variants have a tripartite domain structure (Bradbury *et al.*, 1975): a central globular domain which binds where the linker DNA enters and leaves the nucleosome, sealing the two turns of DNA (Allan *et al.*, 1980), flanked by amino- and carboxy- (N- and C-) terminal 'tail' domains. The C-terminal domain, which is rich in lysine, proline and alanine, probably binds to the linker, partially neutralizing its charge, and is required for chromatin condensation (Allan *et al.*, 1986). It appears to exist as random-coil in solution (Bradbury *et al.*, 1975), probably due to the rather regular distribution of proline residues and to repulsions between the numerous positively charged side-chains of lysine and arginine, which constitute about 30% of the total residues. However, secondary structure prediction suggests an $\alpha$-helical conformation for much of the C-terminal domain, which might be manifest in particular conditions, such as when bound to DNA, as noted previously (Fasman *et al.*, 1977; van Helden, 1982).

Direct determination of the helix content of H1 when bound to DNA from circular dichroism (CD) difference spectra is precluded because the DNA spectrum in the complex is altered (Tinoco *et al.*, 1980). In an alternative approach we have determined the $\alpha$-helicities of the C-terminal

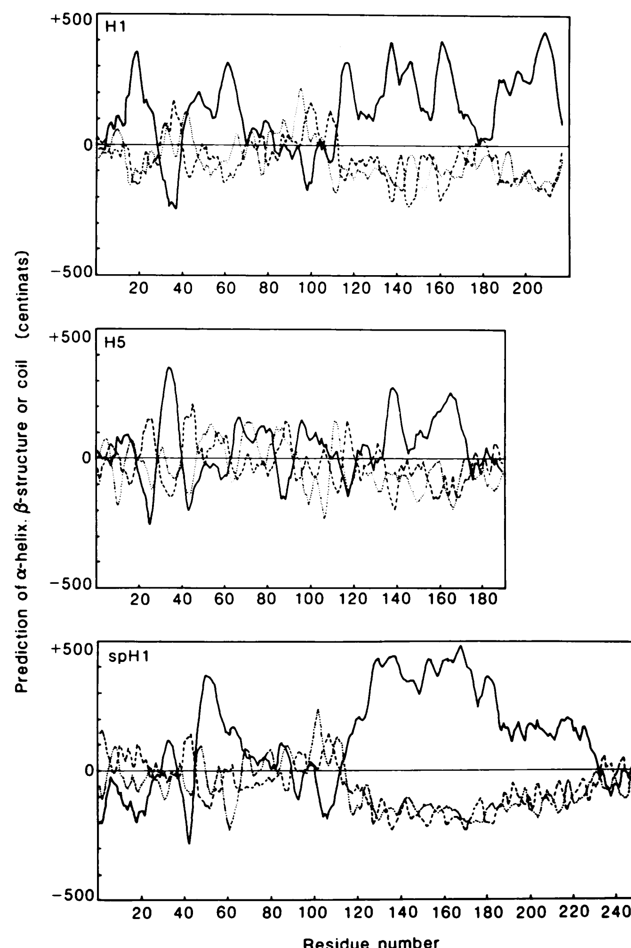


**Fig. 1.** Secondary structure predictions for H1, H5 and spH1 variants using the Robson algorithm (Garnier *et al.*, 1978). (——), $\alpha$-helix; (·····), $\beta$-structure; and (---), coil. **(a)** Chicken H1 (using the sequence of Sugarman *et al.*, 1983); **(b)** chicken H5 (Briand *et al.*, 1980); **(c)** sea urchin (*Parechinus angulosus*) sperm H1 (Strickland *et al.*, 1980).

©IRL Press Limited, Oxford, England

**Chicken H1**

```
         *           *   *  *    *      *         *     *           *        *
106 KLNKKPGETKEKATKKKPAAKPKKPAAKKPAAAAKKPKKAAAVKKSPKKAKKPAAAATKKAAKSPKKATKAGRP-
          *       *         *
    KKTAKSPAKAKAVKPKAAKSKAAKPKAAKAKKAATKKK 217
```

**Chicken H5**

```
               *            *     *     *   *              *     *                    *
94  RLAKSDKAKRSPGKKKKAVRRSTSPKKAARPRKARSPAKKPKATARKARKKSRASPKKAKKPKTVKAKSRKASKAKKVKRSKP-
             *
    RAKSGARKSPKKK 189
```

**Sea urchin sperm H1**

```
           *                                                         *          *    *
112 RVGAVAKPKKAKKTSAAAKAKKAKAAAAKKARRAKAAAKRKAALAKKKAAAAKRKAAAKAKKAKKPKKKAAKKAKKPAKKSP-
    ----------------------E--R-RAA-KK-----------------------T-----------T-A..........
         *    *        *           *        *        *
    KKAKKPAKKSPKKKKAKRSPKKAKKAAGKRKPAAKKARRSPRKAGKRRSPKKARK 248
    .......................................................
```

**Fig. 2.** Amino acid sequences of the C-terminal fragments of H1, H5 and sea urchin (*P.angulosus*) sperm H1 (Sugarman *et al.*, 1983; Briand *et al.*, 1980; Strickland *et al.*, 1980). The proline residues (P) are marked with * for emphasis. Note that the spacing varies from 2 to 57 residues; for segments <4 residues, α-helix is only possible if intra-helical hydrogen bonds are formed between adjacent segments (see text). H1 from chicken erythrocytes is in fact a mixture of six subfractions (Dupressoir and Sautière, 1984; Harborne and Allan, 1986; Shannon and Wells, 1987); the amino acid sequences of the complete set of six H1s have recently been determined from their gene sequences (Coles *et al.*, 1987). The C-terminal regions are virtually identical (in particular, 11 of the 13 proline residues are conserved in all cases), and interpretation of our results with reference to one sequence (Sugarman *et al.*, 1983) does not affect our conclusions, although small differences between subtypes may, of course, turn out to be relevant to function. Part of the sequence of the C-terminal fragment of *Echinus esculentus* sperm H1 (C.S.Hill, unpublished results) used in this work is shown below the sequence of sperm H1 from *P.angulosus;* (---) identical residues; (·····) residues not determined; a 57-residue proline-free segment is present in both species (see text).

domains of chicken erythrocyte H1 and its variant, H5, and of sea urchin sperm H1 (spH1) from CD spectra of the fragments free in solution, under various conditions known to be conducive to α-helix formation in the model compounds poly-L-lysine and poly-L-arginine (Greenfield *et al.*, 1967; Greenfield and Fasman, 1969; Rifkind, 1969; Saitô *et al.*, 1978). Alpha-helix can be induced in H1 and H5 under conditions expected to lead to stabilization of intra-helix hydrogen bonding and neutralization of positive charges, just as found for poly-L-lysine and poly-L-arginine. Because of the proline distribution in H1 and H5 this helix must occur in segments. Clearly, structuring of the C-terminal domain of H1 would have important implications for the organization and condensation of linker DNA in chromatin.

## Results

### *Secondary structure predictions for H1 and its variants*

Although protein secondary structure predictions apply best to 'normal' globular proteins in solution, and the H1 histones are highly atypical and positively charged, predictions can nonetheless reveal potential for secondary structure that might be manifest in an appropriate environment—in the case of histones, when bound to DNA. Prediction of secondary structure for chicken erythrocyte H1 and H5, and spH1, using the algorithm of Garnier *et al.* (1978), indicates a high helical potential of the C-terminal domains of all three molecules (Figure 1), as does the method of Chou and Fasman (1974) used by Fasman *et al.* (1977) and van Helden (1982). The C-terminal domains are, however, punctuated by proline residues (Figure 2) which have a low probability of being found in α-helix, the high overall predicted helicity of these regions (Figure 1) being due to their high alanine and lysine contents, since alanine, particularly, and lysine have a high probability of occurrence in helix. The helix may be continuous (i.e. including proline residues) but divided into segments demarcated by proline residues, each of which would introduce a bend centred on the residue just N-terminal to the proline residue; such proline-containing α-helices are found in glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al.*, 1974), tobacco mosaic virus protein (Bloomer *et al.*, 1978), and 3-phosphoglycerate kinase (Banks *et al.*, 1979). Alternatively, the helix may be broken up into distinct α-helical segments demarcated by proline residues, with no intra-helical hydrogen bonds between adjacent segments; because of the proline distribution (Figure 2), some of the putative helical segments would be very short and therefore relatively unstable.

### *Realization of the helix potential of the C-terminal domains of H1 and its variants: circular dichroism measurements*

The intact chicken histones H1 and H5, and spH1 from *Echinus esculentus*, and fragments (generated by chymotryptic cleavage) containing their C-terminal domains (Figure 3) were examined in various conditions known to be conducive to α-helix formation in the model compounds poly-L-lysine and poly-L-arginine. The conformation of the polypeptide backbone was monitored by measurement of the CD in the 260−195 nm range.

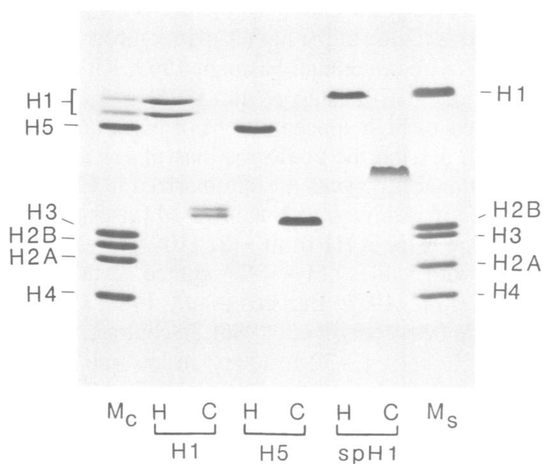


**Fig. 3.** Electrophoretic analysis of histones (H) and chymotryptic fragments (C) used in circular dichroism experiments. H1 and H5 were salt-extracted from chicken erythrocyte nuclei; spH1 was extracted from *E.esculentus* sperm. The C-terminal fragments (C.H1, C.H5 and C.spH1) were prepared by chymotryptic cleavage. The gel contains SDS/18%-polyacrylamide (Thomas and Kornberg, 1978). $M_c$, total marker histones from chicken erythrocyte nuclei; $M_s$, total marker histones from sea urchin sperm nuclei.

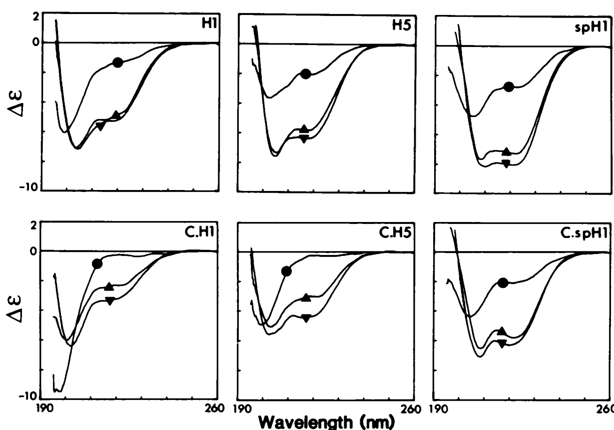


**Fig. 4.** CD spectra of H1, H5, spH1 and their C-terminal fragments: the effect of TFE. The CD spectrum of each histone in low salt buffer (i.e. 1 mM Na phosphate, 0.2 mM $Na_3EDTA$, pH 7.4) (●) is shown with each set of spectra for ease of comparison. (▲) 50% TFE, (▼) 65% TFE, both in the low salt buffer. For analysis of the spectra, see Tables I and II.

To calculate helix contents we have used an empirical relationship (S.R.Martin, unpublished):

$$\% \text{ helix} = (\Delta\epsilon_{220} - 0.25)/0.105$$

which is derived from a plot of the α-helix content of proteins of known (crystallographic) structure against the measured circular dichroism at 220 nm ($\Delta\epsilon_{220}$) and which holds for many (although not necessarily all) proteins in aqueous solution. Helix contents calculated using the CONTIN procedure (Provencher and Glöckner, 1981) are in close agreement with those derived using the above equation (not shown).

*Secondary structure of the H1, H5, spH1 and their C-terminal fragments in low salt buffer.* Although the C-terminal fragments of H1 and H5 are essentially random-coil in low

**Table I.** CD (Δε) measured at 220 nm for H1, H5, spH1 and their C-terminal fragments under various conditions[a]

| | H1 | C.H1 | H5 | C.H5 | spH1 | C.spH1 |
|---|---|---|---|---|---|---|
| Buffer | 1.3 | 0.3 | 2.0 | 0.5 | 2.7 | 1.9 |
| 1 M NaCl | 1.9 | 0.6 | 2.2 | 0.6 | 4.0 | 3.1 |
| 1 M $NaClO_4$ | 2.4 | 1.2 | 2.9 | 1.6 | 5.0 | 4.7 |
| pH 11.2 | 1.8 | 0.8 | 2.1 | 0.7 | 4.5 | 4.1 |
| 50% TFE | 4.8 | 2.4 | 5.6 | 2.9 | 7.3 | 5.9 |
| 65% TFE | 5.1 | 3.3 | 6.2 | 4.3 | 8.1 | 6.5 |

[a]From the spectra in Figures 4 and 5.

**Table II.** Numbers of residues in α-helix for H1, H5, spH1 and their C-terminal fragments under various conditions[a]

| | H1 | C.H1 | H5 | C.H5 | spH1 | C.spH1 |
|---|---|---|---|---|---|---|
| Buffer | 22 | 1 | 32 | 2 | 58 | 21 |
| 1 M NaCl | 34 | 4 | 35 | 3 | 89 | 37 |
| 1 M $NaClO_4$ | 44 | 10 | 48 | 12 | 112 | 58 |
| pH 11.2 | 32 | 6 | 33 | 4 | 100 | 50 |
| 50% TFE | 94 | 23 | 96 | 24 | 166 | 73 |
| 65% TFE | 100 | 32 | 107 | 37 | 186 | 81 |

[a]Calculated from the Δε values in Table I assuming % helix = ($\Delta\epsilon_{220}$ − 0.25)/0.105. The numbers of residues in helix were calculated using the sequences determined for H1 by Sugarman *et al.* (1983) (total 217 residues; C.H1, 112 residues), for H5 by Briand *et al.* (1980) (total 189 residues; C.H5, 96 residues); and for spH1 by Strickland *et al.* (1980) (248 residues; C.spH1, 137 residues). *E.esculentus* spH1 and its C-terminal fragment were identical in length to those from *P.angulosus* (Strickland *et al.*, 1980) as determined by SDS and acid/urea PAGE (C.S.Hill, unpublished); the amino acid composition and a partial amino acid sequence for *E.esculentus* spH1 are also very similar to those of *P.angulosus* spH1 (see, e.g., Figure 2).

ionic strength buffer (1 mM Na phosphate, 0.2 mM $Na_3EDTA$, pH 7.4), intact H1 and H5 contain about 10% helix (~22 residues) and 17% helix (~32 residues), respectively (Figure 4; Tables I and II). These values are higher than those reported in earlier studies (Bradbury *et al.*, 1975; Crane-Robinson *et al.*, 1976), probably due to the presence in our solutions of phosphate ions, which stabilize the globular domains of both H1 and H5 at low concentrations (de Petrocellis *et al.*, 1986). spH1 contains ~23% helix (~58 residues), about half of which may be assigned to the C-terminal fragment (C.spH1), which has ~21 residues in helix even under these conditions, and the remainder presumably to the globular domain. It is likely that the helix stabilized in C.spH1 is part of the long proline-free segment of 57 residues (Figure 2).

Although chicken H1 and H5 have similar helix contents, the intensity of the low wavelength extremum (~200 nm) is much greater for H1 (Figures 4 and 5), suggesting that the non-helical regions of H1 and H5 are quite different. This is also true of C.H1 and C.H5; only the former shows a spectrum typical of an almost totally random-coil polypeptide. Consistent with this observation, a $^{13}$C-NMR study showed that the polypeptide backbones of the N- and C-terminal domains in H5 are less mobile than those in H1 (Shimidzu *et al.*, 1985).

*Effect of hydrogen bond stabilization on the stability of helical segments.* Intra-molecular hydrogen bonds are stabilized in organic solvents such as 2,2,2-trifluoroethanol (TFE) presumably due primarily to its relatively poor hydrogen

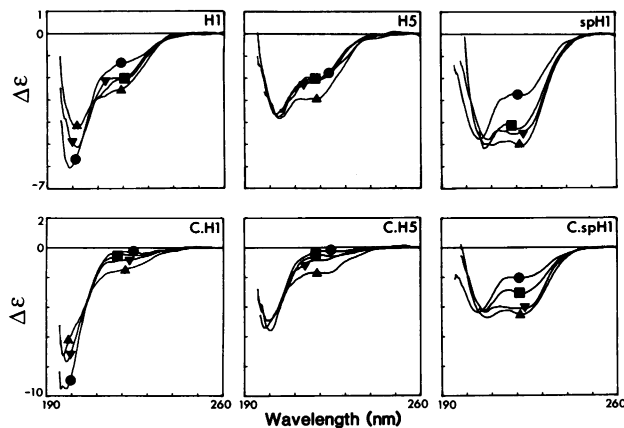


**Fig. 5.** CD spectra of H1, H5, spH1 and their C-terminal fragments: the effect of salt and high pH. The CD spectrum of the histone in low salt buffer (1 mM Na phosphate, 0.2 mM $Na_3EDTA$, pH 7.4) (●) is shown with each set of spectra for ease of comparison. (■) 1 M NaCl; (▲) 1 M $NaClO_4$; (▼) high pH (11.2), all in the low salt buffer. For analysis of the spectra, see Tables I and II.

bonding capacity (Tanford, 1962). For poly-L-lysine and poly-L-arginine, the coil–helix transition can be induced on addition of dioxane (Rifkind, 1969) even when the amino acid side-chains are fully protonated, suggesting that the stabilization of hydrogen bonds outweighs the potentially destabilizing increase in electrostatic repulsion between side-chains (although the latter may be offset by increased counterion binding in the organic solvent).

TFE has a dramatic effect on the conformation of all three histones (Figure 4; Tables I and II). The helix content, determined from $\Delta\epsilon_{220}$, increases with TFE concentration, reaching a maximum at 65%; the apparent decrease in $\Delta\epsilon_{220}$ above this value (not shown) was probably due to precipitation of the protein. In 65% TFE, histone H1 has ~100 residues in helix (46%), H5 has ~107 (57%) and spH1 has ~186 (75%). This concentration of TFE has a similar effect on the conformation of the C-terminal fragments: C.H1 has ~32 of a total of 112 residues in helix (29%), compared with no helix in the absence of TFE; C.H5 has ~37 of a total of 96 residues in helix (39%) (no helix in the absence of TFE); and C.spH1 has ~81 of a total of 137 residues in helix (60%) (although in this case ~21 residues are already in helix in the absence of TFE). The measured helix contents (% residues in helix) of the C-terminal fragments in 65% TFE are thus in the order C.spH1 > C.H5 > C.H1; their proline contents (% proline) show the reverse trend, consistent with the disruptive effect of proline on intra-helical hydrogen bonding.

*Effect of charge screening and neutralization on the stability of helical segments.* The coil–helix transition is induced in the model compounds poly-L-lysine and poly-L-arginine by sodium perchlorate but not by sodium chloride. This has been attributed to the binding of bulky $ClO_4^-$ anions to guanidinium groups in poly-L-arginine or to ε-amino groups in poly-L-lysine, thereby neutralizing helix-destabilizing positive charge repulsions; the relatively small chloride anions are not bound, their general charge-screening effect being insufficient to stabilize helix (Rifkind, 1969). Poly-L-lysine and poly-L-arginine also become helical at pH values sufficiently high to deprotonate side chains, although in the case of poly-L-arginine the picture is confused by the onset of precipitation at the higher pH required (Greenfield *et al.*, 1967; Greenfield and Fasman, 1969; Rifkind, 1969; Boublik *et al.*, 1970; Saitô *et al.*, 1978).

The effects of high concentrations of NaCl and $NaClO_4$, and of high pH, on the conformations of H1 and H5 and their C-terminal fragments are summarized in Figure 5 and Tables I and II. Addition of NaCl to 1 M raises the number of residues in helix in H1 from ~22 (10%) to ~34 (16%), consistent with values (11–18%) quoted in the literature (Bradbury *et al.*, 1975; Barbero *et al.*, 1980; Giancotti *et al.*, 1981; de Petrocellis *et al.*, 1986). In contrast, H5 already contains 17% helix (~32 residues) in low salt buffer and NaCl induces little more, the observed helix content (19%) being consistent with values of 14–16% suggested by earlier studies (Garel *et al.*, 1975; Crane-Robinson *et al.*, 1976; Aviles *et al.*, 1978). Since no helix is formed in C.H1 or C.H5 in 1 M NaCl, the helix in H1 and H5 probably occurs in the globular domain, as shown in the studies cited above and consistent with a two-dimensional NMR study of the globular domain of H5 (Clore *et al.*, 1987). This suggested 32 out of the total 79 residues in helix, corresponding to 17% for the whole molecule if there is no helix in the C-terminal domain. In the low ionic strength phosphate buffer the globular domain of H5 is therefore completely folded, whereas that of H1 is only partially folded.

In low salt buffer spH1 has ~58 residues in α-helix. Addition of NaCl stabilizes ~31 more residues in helix, consistent with the observation of some salt-induced helix in the tail domain(s) of *Sphaerechinus granularis* sperm H1 (Giancotti *et al.*, 1981). The helix content of C.spH1 increases by ~16 residues (from ~21 to ~37) on addition of NaCl (presumably in the 57-residue proline-free segment), thus accounting for about half of the extra residues in helix in the intact molecule. The increased helix content in 1 M NaCl indicates a contribution of positive charge-screening to helix stabilization; that this is not the case for poly-L-lysine and poly-L-arginine in 1 M NaCl (see above) is possibly due to their much higher charge density.

More helix is induced in all three H1 variants by 1 M $NaClO_4$ than by 1 M NaCl, suggesting that anion binding may have some effect on their conformations. This effect is greatest in C.spH1, in which ~21 additional residues take up an α-helical conformation to give a total of ~58 residues in helix (possibly all in the long proline-free segment); in C.H1 and C.H5, ~10 and ~12 residues respectively are in helix in 1 M $NaClO_4$.

Increasing the pH from 7.4 to 11.2 induces considerably more α-helix in H1 and spH1 and a little more in H5 (Figure 5; Tables I and II). The helix contents of H1 and H5 at pH 11.2 are essentially the same as those in 1 M NaCl, suggesting that high pH and high NaCl concentration have similar effects on their secondary structures, i.e. stabilization of helix in the globular domain. In contrast very little helix is induced in C.H1 and C.H5 at high pH, as in 1 M NaCl, showing that removal of most of the side-chain positive charges alone is not sufficient for the stabilization of helix in these regions. However, in C.spH1 about 50 residues are in α-helix at pH 11.2, again probably part of the very long proline-free segment, suggesting that neutralization of side-chain positive charges is sufficient to stabilize α-helix if the intra-helical hydrogen bonding pattern is not disrupted by proline residues.

## Discussion

This systematic analysis shows that the C-terminal domains of H1 and its variants possess latent α-helix-forming potential some of which at least can be realized in an appropriate environment. It seems reasonable to assign helix to the longest segments demarcated by proline residues. The stability of individual segments is probably determined by the number of residues able to participate fully in intra-helical hydrogen bonding, by the degree of repulsion between positively charged side-chains in the helix and by stabilizing interactions between the various side-chains.

Clearly the important question concerns the structure of the C-terminal domains of H1 and its variants bound to DNA. Unfortunately, CD difference spectra cannot be used to answer this directly as the binding to DNA of H1, H5, protamine and indeed the model compounds poly-L-lysine and poly-L-arginine (Carroll, 1972; Wagner *et al.*, 1977; Liao and Cole, 1981) results in an inversion of the B-DNA spectrum to a strong negative CD band, the '*psi*-DNA' spectrum (for a review, see Tinoco *et al.*, 1980); in the case of H1 this is due to binding of the C-terminal domain (Moran *et al.*, 1985) not the globular domain (Russo *et al.*, 1983; Moran *et al.*, 1985). Similarly, we found that Fourier transform infra-red (FTIR) spectra could not be used to determine directly whether helix is induced in the C-terminal domains of H1 and its variants when they bind to DNA. However, extrapolation from structure induced in DNA-binding peptides in TFE to structure induced on binding to DNA in aqueous solution seems justified from studies of model peptides (Walters and Kaiser, 1985). In the case of two 23-residue DNA-binding peptides that essentially lacked secondary structure free in aqueous solution, CD spectra showed 65% and 81% α-helix in 80% TFE, and FTIR spectroscopy showed a shift in the amide I absorption maximum on binding to DNA, consistent with α-helix.

There is, moreover, a precedent for the induction of α-helix in a highly basic, random-coil protein on binding to a nucleic acid. The X-ray crystal structure shows that protamine bound to the shallow groove of tRNA contains helical segments of up to ten residues, such that the protonated guanidino side chains of arginine residues form hydrogen bonds to the phosphates on either side of the groove (Warrant and Kim, 1978). The structure of the C-terminal domain of H1 when bound to DNA could well be similar, the rigidity of double-helical DNA having a stabilizing effect on helical segments 'anchored' to the DNA phosphates by the side-chains of lysine and arginine residues.

In summary, there is strong circumstantial evidence that the C-terminal domain of H1 might assume a segmented α-helical conformation on binding to DNA. H1 is positioned at its binding site on the nucleosome by the globular domain; at this site linker DNA enters and leaves the nucleosome and the N- and C-terminal tails emerge from the globular domain. One possibility is that rigid helical segments of the C-terminal domain track the phosphate backbone, kinking around the linker DNA by virtue of proline-induced bends or breaks in the helix, following one or other of the grooves, possibly the major rather than the minor groove, since early studies revealed no protection of the minor groove from chemical modification (Mirzabekov *et al.*, 1977). Alternatively, helical segments of the C-terminal domain may lie on the face of the DNA, binding across rather than in the groove, but still kinking around the DNA; in particular, the very long segment in spH1 is likely to bind in this way because it is far too long to be accommodated as a straight rod within a groove.

Structuring of the C-terminal domain into α-helical segments is likely to have repercussions for the organization of the linker DNA, which has different lengths in different chromatins. The organization of the linker in the 30 nm filament is unknown, but it is very likely that it is bent (see, for example, Butler, 1984; Widom *et al.*, 1985), and that the C-terminal domain plays some part in this. The linker lengths associated with H1, H5 and spH1 are 30, 46 and 74 bp respectively, calculated from measured repeat lengths of 196 bp for chicken oviduct (Compton *et al.*, 1976), 212 bp for chicken erythrocytes (Morris, 1976) and 240 bp for *E.esculentus* sperm (Thomas *et al.*, 1986). If an H1 molecule is associated with a particular linker length it presumably has helical segments designed to guide that length of linker DNA into the next nucleosome in the 30 nm filament. Important factors would probably be the length, stability and charge density of the helical segments. Further studies of the interaction of the C-terminal domains of H1 variants with DNA, and possibly molecular modelling, may provide new insights.

## Materials and methods

### *Preparation of the C-terminal fragments of chicken erythrocyte H1 and H5*

H1 and H5 were salt-extracted (0.65 M NaCl) from chicken erythrocytes exactly as described (Clark and Thomas, 1986). H1 at 0.75 mg/ml in 10 mM Na phosphate, pH 7.0 was digested for 5 min at 37°C with chymotrypsin (Sigma) (enzyme:histone 1:500, w/w). Digestion was stopped by addition of phenylmethylsulphonyl fluoride (PMSF) from a 50 mM solution in propan-2-ol to 0.5 mM and cooling to 4°C. The digest was fractionated by ion-exchange chromatography (Whatman CM52), using a linear gradient of 0–0.7 M NaCl in 10 mM Na phosphate, pH 7.0, 0.5 mM PMSF. The eluate was monitored at 230 nm and protein fractions analysed by SDS/18%-PAGE (Thomas and Kornberg, 1978). The N-terminal fragment of H1 and its digestion products eluted first and the C-terminal fragment in a second peak. Pooled fractions were dialysed into 10 mM Na phosphate, 0.25 mM PMSF, pH 7.0, concentrated on CM52 and eluted with 0.6 M NaCl in the same buffer. The fragments were dialysed into 1 mM Na phosphate, 0.2 mM $Na_3EDTA$, 0.25 mM PMSF, pH 7.4 and stored at −20°C in aliquots. The C-terminal fragment of H5 was prepared in essentially the same way.

### *Preparation of sea urchin sperm H1 and its C-terminal fragment*

H1 was extracted from *E.esculentus* sperm nuclei (isolated as described by Thomas *et al.*, 1986) with 5% (v/v) perchloric acid (PCA) at 0°C (Johns, 1964). The PCA-soluble fraction was immediately neutralized with 1 M triethanolamine and dialysed at 4°C against 10 mM Na phosphate buffer containing 0.25 mM PMSF ('column buffer'). H1 was purified by ion-exchange chromatography (Whatman CM52) as described for chicken erythrocyte H1 and H5 (Clark and Thomas, 1986) using a linear gradient of NaCl (0.3–1 M) in column buffer. Fractions containing pure H1 were pooled and dialysed against 0.1% (w/v) ammonium bicarbonate, pH 7.8. After lyophilization H1 was dissolved at 1 mg/ml in 1 mM Na phosphate, 0.2 mM $Na_3EDTA$, pH 7.4 and the concentration determined accurately by amino acid analysis.

To prepare the C-terminal fragment, the neutralized PCA-soluble extract of nuclei (see above) containing predominantly H1 was dialysed overnight at 4°C against 10 mM Na phosphate buffer pH 7, and digested with chymotrypsin (1:100, w/w), as described above for chicken H1 and H5, at 23°C for 7 min. This gave three fragments derived from the C-terminal domain of H1 and an N-terminal fragment (as indicated by SDS–PAGE); these were separated on carboxymethyl cellulose as described above, using a linear gradient of 0.1–1 M NaCl. The three C-terminal fragments eluted together and were dialysed against 0.1% (w/v) ammonium bicar-

bonate, pH 7.8, lyophilized, and separated by reverse phase HPLC. The peptide mixture was dissolved in 100 μl aqueous trifluoroacetic acid (TFA; 0.3%, v/v) and separated on a C-18 column [Spherisorb 3ODS(II), 3 μm, 25 × 0.46 cm, fully capped, from HPLC Technology] with a guard column (Waters), using an LKB Ultrachrom GTi HPLC system. The fragments were eluted (flow rate 0.7 ml/min) with a gradient of acetonitrile (15–50% in 60 min) in 0.3% (v/v) TFA and 0.7 ml fractions were collected. The eluate was monitored at 226 nm and analysed by SDS gel electrophoresis. The three C-terminal fragments, which were obtained pure, corresponded to cleavages at Tyr-75, Leu-90 and Phe-111 in spH1, as shown by comparison of their amino acid compositions with those of chymotryptic fragments expected from the sequence of *P.angulosus* spH1 (Strickland *et al.*, 1980). The fragment arising from cleavage at Phe-111, which corresponded to the C-terminal fragment generated from chicken erythrocyte H1 and H5, was lyophilized and dissolved in 1 mM Na phosphate, 0.2 mM $Na_3EDTA$, pH 7.4.

### Protein concentrations

Protein concentrations were determined by amino acid analysis using norleucine as internal standard, and the known amino acid sequences of the histones and their fragments as described previously (Clark and Thomas, 1986). For chicken H1 an average amino acid composition was calculated from the compositions of the six known subtypes, which are very similar (Coles *et al.*, 1987). For H5 the sequence determined by Briand *et al.* (1980) was used, and for spH1 the experimentally determined amino acid composition (C.S.Hill, unpublished results) assuming 248 residues (see footnote to Table II).

### CD spectra

Stock solutions of H1, H5, spH1 and their C-terminal fragments were prepared at 1 mg/ml and the concentrations determined accurately by amino acid analysis. Dilutions were such that the final protein concentration was 0.12 mg/ml in 1 mM Na phosphate, 0.2 mM $Na_3EDTA$ with varying concentrations of trifluoroethanol (Sigma), NaCl or $NaClO_4$ as necessary, and the final pH was 7.4, except in the experiments carried out at high pH where the final pH was 11.2. CD spectra were recorded from 260–195 nm on a JASCO J41C spectropolarimeter equipped with a model J-DPY data processor, at sensitivities in the range 0.5–2.0 mdeg/cm and with an instrument time constant of 4 s. Cuvettes with a 1 mm path length were used and the spectra are the average of at least four scans. Digitized spectra collected by the data processor were transferred to a PDP 11/23 computer and subsequently processed for base-line substraction, normalization and smoothing according to the method of Savitzky and Golay (1964). All spectra are presented as molar circular dichroism, $\Delta\epsilon$, based on mean residue weights (mrw) of 100 for H1, 103 for C.H1, 109 for H5, 110 for C.H5, 105 for spH1 and 104 for C.spH1, calculated from the published amino acid sequences of H1 (Sugarman *et al.*, 1983) and H5 (Briand *et al.*, 1980); for *E.esculentus* sperm H1, the experimentally determined amino acid composition was used (C.S.Hill, unpublished results) assuming 248 residues for spH1 and 137 residues for C.spH1 (see footnote to Table II). Values of molar ellipticity may be obtained using the equation $[\Theta]_{mrw} = 3330.\Delta\epsilon$. The values of $\Delta\epsilon_{220}$ reported in Tables I and II are the average of at least two separate determinations. From the experimental signal-to-noise ratio we estimate the error in values of $\Delta\epsilon_{220}$ to be less than 0.3.

### Protein secondary structure prediction

The RPREDICT programme on the Cambridge University IBM 3084 mainframe computer was used to predict secondary structure according to the Robson algorithm (Garnier *et al.*, 1978).

## Acknowledgements

We thank Dr P.M.Bayley for making available the CD spectropolarimeter at NIMR, Professor C.von Holt for providing a sample of *P.angulosus* spH1 for comparison with *E.esculentus* spH1, Dr C.W.Wharton (University of Birmingham) for assistance with measurements of FTIR spectra, and Drs P.M.Bayley and P.J.G.Butler for helpful comments on the manuscript. Amino acid analyses were carried out by J.B.Lester, and sequence analysis of *E.esculentus* sperm H1 peptides by Dr L.Packman on an Applied Biosystems 470A gas phase sequencer at the Protein Sequencing Facility, Department of Biochemistry, Cambridge. This work was supported by a project grant from the SERC (GR/D/05134) and by equipment grants from the SERC (GR/D/71214) and the Wellcome Trust. D.J.C. and C.S.H. acknowledge the award of Research Studentships form the MRC and SERC, respectively.

## References

Allan,J., Hartman,P.G., Crane-Robinson,C. and Aviles,F.X. (1980) *Nature*, **288**, 675–679.

Allan,J., Mitchell,T., Harborne,N., Böhm,L. and Crane-Robinson,C. (1986) *J. Mol. Biol.*, **187**, 591–601.

Aviles,F.J., Chapman,G.E., Kneale,G.G., Crane-Robinson,C. and Bradbury,E.M. (1978) *Eur. J. Biochem.*, **88**, 363–371.

Banks,R.D., Blake,C.C.F., Evans,P.R., Haser,R., Rice,D.W., Hardy,G.W., Merrett,M. and Phillips,A.W. (1979) *Nature*, **279**, 773–777.

Barbero,J.L., Franco,L., Montero,F. and Moran,F. (1980) *Biochemistry*, **19**, 4079–4087.

Bates,D.L., Butler,P.J.G., Pearson,E.C. and Thomas,J.O. (1981) *Eur. J. Biochem.*, **119**, 469–476.

Bloomer,A.C., Champness,J.N., Bricogne,G., Staden,R. and Klug,A. (1978) *Nature*, **276**, 362–368.

Boublik,M., Bradbury,E.M., Crane-Robinson,C. and Rattle,H.W.E. (1970) *Eur. J. Biochem.*, **12**, 258–263.

Bradbury,E.M., Cary,P.D., Chapman,G.E., Crane-Robinson,C., Danby,S.E., Boublik,M., Palau,J. and Aviles,F. (1975) *Eur. J. Biochem.*, **52**, 605–613.

Briand,G., Kmiécik,D., Sautière,P., Wouters,D., Borie-Loy,O., Biserte,G., Mazen,A. and Champagne,M. (1980) *FEBS Lett.*, **112**, 147–151.

Buehner,M., Ford,G.C., Moras,D., Olsen,K.W. and Rossman,M.G. (1974) *J. Mol. Biol.*, **90**, 25–49.

Butler,P.J.G. (1984) *EMBO J.*, **3**, 2599–2604.

Carroll,D. (1972) *Biochemistry*, **11**, 421–433.

Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry*, **13**, 222–245.

Clark,D.J. and Thomas,J.O. (1986) *J. Mol. Biol.*, **187**, 569–580.

Clore,G.M., Gronenborn,A.M., Nilges,M., Sukumaran,D.K. and Zarbock,J. (1987) *EMBO J.*, **6**, 1833–1842.

Coles,L.S., Robins,A.J., Madley,L.K. and Wells,J.R.E. (1987) *J. Biol. Chem.*, **262**, 9656–9663.

Compton,J.L., Bellard,M. and Chambon,P. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 4382–4386.

Crane-Robinson,C., Danby,S.E., Bradbury,E.M., Garel,A., Kovacs,A., Champagne,M. and Daune,M. (1976) *Eur. J. Biochem.*, **67**, 379–388.

de Petrocellis,L., Quagliarotti,G., Tomei,L. and Geraci,G. (1986) *Eur. J. Biochem.*, **156**, 143–148.

Dupressoir,T. and Sautière,P. (1984) *Biochem. Biophys. Res. Comm.*, **122**, 1136–1145.

Fasman,G.D., Chou,P.Y. and Adler,A.J. (1977) In Ts'o,P.O.P. (ed.), *The Molecular Biology of the Mammalian Genetic Apparatus.* North-Holland, Amsterdam, pp. 1–45.

Garel,A., Kovacs,A.M., Champagne,M. and Daune,M. (1975) *Biochim. Biophys. Acta*, **395**, 5–15.

Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.

Giancotti,V., Russo,E., Cosimi,S., Cary,P.D. and Crane-Robinson,C. (1981) *Biochem. J.*, **197**, 655–660.

Greenfield,N., Davidson,B. and Fasman,G.D. (1967) *Biochemistry*, **6**, 1630–1637.

Greenfield,N. and Fasman,G.D. (1969) *Biochemistry*, **8**, 4108–4115.

Harborne,N. and Allan,J. (1986) *FEBS Lett.*, **194**, 267–272.

Johns,E.W. (1964) *Biochem. J.*, **92**, 55–59.

Johns,E.W. (1969) In Wolstenholm,G.E.W. and Knight,J. (eds), *Homeostatic Regulation, Ciba Found. Symp.* Churchill, London, p. 128.

Liao,L.W. and Cole,R.D. (1981) *J. Biol. Chem.*, **256**, 6751–6755.

Mizabekov,A.D., Sanko,D.F., Kolchinsky,A.M. and Melnikova,A.F. (1977) *Eur. J. Biochem.*, **75**, 379–389.

Moran,F., Montero,F., Azorin,F. and Suau,P. (1985) *Biophys. Chem.*, **22**, 125–129.

Morris,N.R. (1976) *Cell*, **9**, 627–632.

Neelin,J.M., Callahan,P.L., Lamb,D.C. and Murray,K. (1964) *Can. J. Biochem.*, **42**, 1743–1752.

Poccia,D.L., Simpson,M.V. and Green,G.R. (1987) *Dev. Biol.*, **121**, 445–453.

Provencher,S.W. and Glöckner,J. (1981) *Biochemistry*, **20**, 33–37.

Rifkind,J.M. (1969) *Biopolymers*, **8**, 685–688.

Russo,E., Giancotti,V. and Crane-Robinson,C. (1983) *Int. J. Biol. Macromol.*, **5**, 366–368.

Saitô,H., Ohki,T., Kodama,M. and Nagata,C. (1978) *Biopolymers*, **17**, 2587–2599.

Savitzky,A. and Golay,M.J.E. (1964) *Anal. Chem.*, **36**, 1627–1639.

Shannon,M.F. and Wells,J.R.E. (1987) *J. Biol. Chem.*, **262**, 9664–9668.

Shimidzu,M., Shindo,H., Matsumoto,U., Mita,K. and Zama,M. (1985) *Eur. J. Biochem.*, **148** 431–436.

Strickland,W.N., Strickland,M., Brandt,W.F., von Holt,C., Lehmann,A. and Wittmann-Liebold,B. (1980) *Eur. J. Biochem.*, **104**, 567–578.
Sugarman,B.J., Dodgson,J.B. and Engel,J.D. (1983) *J. Biol. Chem.*, **258**, 9005–9016.
Tanford,C. (1962) *J. Am. Chem. Soc.*, **84**, 4240–4247.
Thoma,F., Koller,T. and Klug,A. (1979) *J. Cell Biol.*, **83**, 403–427.
Thomas,J.O. and Kornberg,R.D. (1978) *Methods Cell Biol.*, **18**, 429–440.
Thomas,J.O., Rees,C. and Butler,P.J.G. (1986) *Eur. J. Biochem.*, **154**, 343–348.
Tinoco,I., Bustamante,C. and Maestre,M.F. (1980) *Annu. Rev. Biophys. Bioeng.*, **9**, 107–141.
van Helden,P.D. (1982) *J. Theor. Biol.*, **96**, 327–336.
Wagner,T.E., Hartford,J.B., Serra,M., Vandegrift,V. and Sung,M.T. (1977) *Biochemistry*, **16**, 289–290.
Walters,L. and Kaiser,E.T. (1985) *J. Am. Chem. Soc.*, **107**, 6422–6424.
Warrant,R.W. and Kim,S.-H. (1978) *Nature*, **271**, 130–135.
Widom,J., Finch,J.T. and Thomas,J.O. (1985) *EMBO J.*, **4**, 3189–3194.

*Received on October 26, 1987*