Structure of the chromatin binding (chromo) domain from mouse modifier protein 1

Linda J. Ball, Natalia V. Murzina, R. William Broadhurst, Andrew R. C. Raine, Sharon J. Archer, Francesca J. Stott, Alexey G. Murzin, Prim B. Singh, Peter J. Domaille and Ernest D. Laue

Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK; 1Du Pont Merck Pharmaceutical Company, Box 80328, Wilmington, DE 19880-0328, USA; 2Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge CB2 2QH and 3The Babraham Institute, Babraham Hall, Babraham, Cambridge CB2 4AT, UK

© Oxford University Press 1997

Introduction

The chromatin organization modifier (chromo) domain has been defined as a 30–70 residue protein module found in a large number of proteins. So far more than 40 examples of the chromo domain are known (Aasland and Stewart, 1995; Koonin et al., 1995) and the list is growing rapidly. Chromo domains were first identified in two Drosophila proteins, heterochromatin associated protein 1 (HP1) and Polycomb (Pc) (Paro and Hogness, 1991). HP1 has been shown to be involved in repression of gene expression in heterochromatin (Eissenberg et al., 1990, 1992) and Pc is a member of the Polycomb group of transcriptional repressors which suppress expression of homeotic genes (Paro, 1990, 1993). Studies of both Pc (Messmer et al., 1992) and HP1 (Platero et al., 1995) have shown that chromo domains are responsible for targeting their respective proteins to their sites of action on chromatin. Similarly, the chromo domain in human retinoblastoma binding protein (RBP1) may deliver retinoblastoma protein (Rb), a global transcription factor, to particular sites (Szekely et al., 1991; Koonin et al., 1995). In addition, Drosophila male-specific lethal (MSL-3) protein, an activator, might also be localized to chromatin via its chromo domain (Koonin et al., 1995).

It is now apparent that the chromo domain family of proteins comprises at least two groups. The HP1-like proteins contain two conserved homologous domains (Clark and Elgin, 1992): an N-terminal chromo domain and a C-terminal ‘shadow’ chromo domain (Epstein et al., 1992; Aasland and Stewart, 1995). Mutational analysis has shown that the N-terminal chromo domain of HP1 is sufficient to direct heterochromatin binding of a β-galactosidase fusion protein in vivo, whereas the C-terminal shadow domain is necessary for nuclear localization and heterochromatin binding (Powers and Eissenberg, 1993; Platero et al., 1995). The two chromo domains of the HP1-like proteins may thus serve as adaptors bringing together different proteins in a multi-protein complex (Aasland and Stewart, 1995). In other cases, such as Pc, CHD-1 (Delmas et al., 1993) and SWI6 (Lorenz et al., 1994; Ekwall et al., 1995), the chromo domain is a component of a much larger protein and is typically fused to other protein modules (Aasland and Stewart, 1995).

The HP1 gene was identified as a suppressor of position-effect variegation (PEV) (Eissenberg et al., 1990). PEV occurs when a gene normally residing in euchromatin is placed next to a heterochromatic boundary after a chromosomal rearrangement. The gene becomes specifically silenced in some cells, but not in others, leading to a variegated phenotype. Suppressors or enhancers of PEV are likely to encode proteins that are involved in the formation or modification of chromatin structure (for a review see Reuter and Spierer, 1992). PEV is similar in many ways to other mechanisms of gene silencing, such as those involving the homeotic genes (Paro, 1993), mating type control in yeast (Laurenson and Rine, 1992), X-chromosome inactivation (Gartler and Riggs, 1983) and imprinting in mammalian cells (Solter, 1988).

In this paper we describe experiments that determined the domain structure of HP1-like heterochromatin-associated mouse chromatin modifier protein 1 (MoMOD1) (Singh et al., 1991). The N-terminal chromo domain (MoMOD1-N, see Figure 1) was expressed in Escherichia coli, purified to homogeneity and its solution structure determined using nuclear magnetic resonance (NMR) spectroscopy. Unexpectedly, the structure is similar to that of the Sac7d and Sso7d DNA binding proteins from the archaeabacteria Sulfolobus acidocaldarius and S. solfataricus (Baumann et al., 1994; Edmondson et al., 1995). The structural comparisons suggest how the chromo domain may function as a protein interaction motif.

Results

Expression of the intact MoMOD1 protein

The full-length MoMOD1 protein (residues 1–185) was cloned into pET16b and expressed in E.coli BL21(DE3)
Fig. 1. Amino acid sequence of the MoMOD1 chromo domain showing its homology with other representative chromo domains. Selected conserved residues are coloured yellow (core hydrophobic), green (Gly and Pro) and blue (basic). See text for details.

---

Fig. 2. Limited proteolysis of MoMOD1: 16% tricine–SDS–PAGE separation of a 0.5% trypsin or 1% subtilisin digest of the MoMOD1 protein. (a) Lane 1, intact protein; lane 2, after 5 min incubation; lane 3, after 15 min incubation; lane 4, after 30 min incubation; lane 5, after 60 min incubation with subtilisin. (b) Lane 1, intact protein; lane 2, after 5 min incubation; lane 3, after 30 min incubation; lane 4, after 60 min incubation with trypsin. On the right of these panels, the results of the microsequencing and mass spectrometric analysis are summarized. The numbers represent the N- and C-terminal amino acids of the major components identified from the bands. (c) Amino acid sequence of the MoMOD1 protein showing the protease cut sites. The sequences of the protease-resistant N- and C-terminal domains are underlined. Up arrow, trypsin cleavage site; down arrow, subtilisin cleavage site.

---

as a histidine-tagged fusion protein. The protein was purified using nickel affinity chromatography, where it copurified with small amounts of two shorter fragments of the protein. These may result from truncated transcription or translation or, perhaps, degradation by endogenous proteases. The shorter fragments did not interfere with interpretation of the limited proteolysis experiments presented below.

**Limited proteolysis of the intact MoMOD1 protein**

Before the proteolysis experiments were carried out, the stability of the protein was assessed under the conditions of the digest. This showed that the protein was not degraded. Digestion conditions were then varied over a wide range to determine which regions of the molecule are most susceptible to proteolysis. The proteolytic fragments were then separated using either tricine–SDS–PAGE (Figure 2a and b) or reversed phase HPLC (not shown) and were identified using N-terminal amino acid sequencing and electrospray ionization mass spectrometry. The N-terminus of each fragment was determined directly by N-terminal sequencing, whereas the C-terminus was identified indirectly using mass spectroscopy and the statistical program WEIGHT (W. Boucher, unpublished data); this program assigns a series of possible protein sequences to each mass spectral fragment. A summary of the results is presented in Figure 2C.

**Expression of the MoMOD1-N chromo domain fragment**

Based on the sequence alignment (Figure 1) and limited proteolysis data (Figure 2), a fragment containing residues 10–80 of MoMOD1 was expressed in *E. coli* in milligram quantities. The N-terminal sequence was HMVEEVL . . . , which is as expected with the addition of a histidine and methionine residue from the expression vector. Amino
MoMOD1-N is similar to that outlined in Clowes et al. from 1H secondary structure elements and NOEs. For and NOESY (Kumar et al.) the consensus chemical shift index for 1H and 13C assignments were identified from four-dimensional 13C-separated HCCH-TOCSY (Bax et al.) experiments. Aromatic side chain signals were determined for six Val, Ile and Thr residues position. Three-dimensional 13N- and 15N-separated NOESY spectra were determined for six Val, Ile and Thr residues position. Between residues 60 and 71 the dNN(i, i + 2), dNN(i, i + 3), dαβ(i, i + 3), dδN(i, i + 3) and dδN(i, i + 4) NOE connectivities between the residues shown.

The triple resonance assignment strategy employed for MoMOD1-N is similar to that outlined in Clowes et al. (1995). Assignments of the chemical shifts of 1H, 13C and 15N nuclei in the protein backbone were made using a combination of three-dimensional CBCA(CO)NNH and CBCANNH experiments (Grzesiek and Bax, 1992a,b) and confirmed using three-dimensional 15N-separated TOCSY-HMQC, NOESY-HMQC (Marion et al., 1989a,b) and HNHB (Archer et al., 1991) spectra. Aliphatic side chain 1H and 13C assignments were identified from four-dimensional HCC(CO)NNH (Clowes et al., 1993) and three-dimensional 13C-separated HCCH-TOCSY (Bax et al., 1990a) experiments. Aromatic side chain signals were assigned by reference to two-dimensional CBHD, CBHE (Yamazaki et al., 1993), DQF-COSY (Rance et al., 1983) and NOESY (Kumar et al., 1980) spectra.

Through-space nuclear Overhauser effect (NOE) connectivities from backbone amide to side chain protons were identified from a four-dimensional 13C/15N-separated NOESY experiment (Kay et al., 1990). Amide-amide and side chain–side chain connections were obtained from three-dimensional 15N- and 13C-separated NOESY spectra (Marion et al., 1989b; Ikura et al., 1990; Zuiderweg et al., 1990). Vicinal coupling constants between amide and 1H, spins, 3JHN, were determined by least squares fitting to the intensities of crosspeaks from a set of two-dimensional J-modulated 13N-1H-COSY experiments (Billeter et al., 1992), leading to estimates of the backbone angle φ by reference to a Karplus curve (Pardi et al., 1984). Slowly exchanging amide protons were identified by recording 1H,13N HSQC (Bax et al., 1990b; Norwood et al., 1990) spectra on a sample immediately after the H2O solvent had been exchanged for D2O.

The elements of secondary structure were then identified from 1Hα, 13Cα and 15N chemical shifts (Wishart and Sykes, 1994), patterns of short and intermediate range inter-residue NOEs, 3JHN couplings and amide exchange rates (Wuthrich, 1986), as illustrated by the connectivity summary in Figure 3. The presence of strong dNN NOEs, retarded amide exchange rates, positive values for the consensus chemical shift index (CSI) and large values of 3JHN indicate that there are β-strands from residues 23 to 31, 36 to 42 and 51 to 54. Between residues 60 and 71 the dNN(i, i + 3) and dαβ(i, i + 3) NOEs, the strong dNN and weak dNN NOEs, the slow amide exchange, the negative values for the consensus CSI and the small values of 3JHN reveal a C-terminal α-helix.

**Determination of the tertiary structure**

In addition to the 329 sequential and 218 medium range inter-residue NOEs that were used to locate the elements of secondary structure in the previous section, 910 intra-residue NOEs were identified. For certain side chains it was possible to identify the most populated χ1 rotamer and make stereospecific assignments of 4H or γ-methyl resonances. 3JCC and 3JCN couplings were determined for six Val, Ile and Thr residues pos-
sensing resolved γ-methyl signals from 15N and 13CO spin-echo difference experiments (Grzesiek et al., 1993; Vuister et al., 1993). Stereospecific 1H assignments for a further nine residues could be made by analysis of peak intensities in three-dimensional 15N-separated HMQC-TOCSY, HMQC-NOESY and HNHB spectra (Powers et al., 1993).

Three-dimensional structures of the chromo domain were produced essentially as described before (Kraulis et al., 1994). Initial structures were computed using simulated annealing calculations from 1207 1H–1H distance restraints that could be determined unambiguously. These computations used the program X-PLOR (Brunger, 1992), starting from random globular initial structures (Nilges et al., 1988, 1991). This set of preliminary structures was used to assist in assignment of the remaining NOE crosspeaks for which the assignments were ambiguous and in selection of acceptor carbonyl oxygen atoms for hydrogen bonds to the slowly exchanging backbone amide protons. The final round of calculations used a restraint list comprising 1937 distance restraints, of which 1786 were unambiguous and 151 semi-ambiguous (Nilges, 1995), 22 φ and 15 χi dihedral angle restraints and 17 hydrogen bonds; the selected ensemble contained all of the 27 converged structures out of 40 calculated. The structures are well defined (see Figure 4A) except for the first 11 residues at the N-terminus and the last nine residues at the C-terminus, which are genuinely flexible in solution, as judged by NMR studies of 15N relaxation in the backbone amides (data not shown). The chromo domain therefore corresponds to residues 21–71 of MoMOD1. The average r.m.s. differences from the mean structure are 0.66 Å (± 0.14) for the backbone atoms and 1.14 Å (± 0.17) for all the non-hydrogen atoms in the well-defined core of the molecule. If the less well-ordered loop region (residues 43–47) is included, these values increase to 0.96 Å (± 0.27) and 1.52 Å (± 0.29) respectively. No distance restraint was violated by >0.5 Å and no dihedral angle restraint by >5.0°. The structures have good covalent geometry as judged by PROCHECK (Laskowski et al., 1993) and non-bonded contacts (see Table I).

**Analysis of the tertiary structure**

The structure of the chromo domain is shown in Figures 4 and 5. The protein consists of an N-terminal three-stranded anti-parallel β-sheet, which packs against the C-terminal α-helix (Figure 4A). A hydrophobic core is formed by residues 23, 26, 38, 40, 42, 58, 60, 63, 64 and 67 (see Figure 4B), all of which are highly conserved in different chromo domains (Figure 1). Several of these residues (23, 40, 42, 58, 60 and 63) also lie at the bottom of a hydrophobic groove (see below) on the surface of the β-sheet below the α-helix. The first strand of the β-sheet is rather irregular and contains two bulges at residues 24 and 27 (Figure 6). The insertions and deletions in the sequences of the different chromo domains are all found in the loops. Insertions and deletions are found in the first and at the end of the second loop in the β-sheet (residues 33–34 and 47–48). Insertions are also found in the loop connecting the β-sheet to the α-helix (residues 59–60) (Aasland and Stewart, 1995). It is therefore likely that all the chromo domains will have similar three-dimensional structures.

**Discussion**

Limited proteolysis of MoMOD1 revealed two regions that are resistant to digestion; these correspond to the conserved sequences in the HP1-like proteins (Clark and Elgin, 1992). The results suggest that the MoMOD1 protein consists of two structural domains, residues 10–78 and 104–171, connected by an exposed linker. Our NMR studies of the MoMOD1-N chromo domain show that the structured region lies between residues 21 and 71, although it is possible that, in the intact MoMOD1 protein, residues 10–20 and 71–78 are protected.

The structure of the MoMOD1-N chromo domain explains the results of mutations in the chromo domains of both HP1 (Platero et al., 1995) and Pc (Messmer et al., 1992), which have been shown to disrupt gene silencing and/or chromatin binding. Three of these mutations, V23M in HP1, I26F and removal of residues 64/65 in Pc (all MoMOD1 numbering), are to residues in the hydrophobic core and they may therefore disrupt the structure of the chromo domain (Figure 4A). One mutation, Y21F in HP1, should not disrupt the structure, but this residue lies to one side of the hydrophobic groove, which we suggest mediates protein–protein interactions (see below). This tyrosine may therefore be important for the function of at least the HP1-like chromo domains. We note that the V23M mutation may additionally disrupt function, because this residue is also exposed in the hydrophobic groove.

Comparisons with other protein structures from the Brookhaven Protein Databank were carried out by eye and using the computer program DALI (Holm and Sander, 1993). Homologies to two different classes of protein were found. The first group includes the α and β chemokines, e.g. IL-8, MGSA/GRO and MIP-1β, RANTES and MCP1, all of which are involved in protein–protein interactions with receptors (Leonard and Yoshimura, 1990). Most interestingly, the second group comprises small histone-like DNA binding proteins found in the archaebacteria *S.acidocaldarius* and *S.solfataricus*. These proteins are also involved in the formation of chromatin structure (Dijk and Reinhardt, 1986) and they are more similar to the MoMOD1-N chromo domain. The r.m.s. deviations for the 31 Cα atoms, which comprise all of the conserved secondary structure, are 1.44 and 2.01 Å, when the chromo domain is compared with the DNA binding domains of Sac7d and Sso7d respectively (Baumann et al., 1994; Edmondson et al., 1995). For the DNA binding domains of Sac7d and Sso7d themselves, which are 70% identical, the r.m.s. deviation is 1.72 Å over the same region. (By comparison, the Cα r.m.s. deviations over the same region are 2.25 and 1.96 Å respectively.) In Figure 5a, a comparison of the structure of Sac7d with that of the chromo domain is shown. The structures are very similar, but the chromo domain lacks the N-terminal hairpin which binds in a hydrophobic groove below the α-helix in Sac7d. We believe that there may be a distant evolutionary relationship between the chromo domain and Sac7d, for the following reasons. First, in addition to conservation of the three-dimensional folds, there are detailed structural similarities as well. For example, the first bulge at residue 24 in the chromo domain is conserved in Sac7d, to facilitate...
Fig. 4. (a) Stereoviews of the backbone (N, Cα and C′) of the 27 superimposed structures of MoMOD1-N (residues 20–73) that converged out of 40 computed from the NMR data. (b) Structure closest to the mean, showing the side chains of the conserved hydrophobic residues in the core and hydrophobic groove of the protein. The side chains are colour coded according to their r.m.s. deviation from the mean structure of all the non-hydrogen atoms in the set of calculated structures: red $< 0.75$ Å; light blue $< 1.5$ Å.

The interaction of a basic residue in the β-sheet (residue 29) with residues 64 and 67 in the α-helix is also conserved (Figure 6), as is the loop between the third β-strand and the α-helix (Figure 5a). Second, both the HP1-like chromo domains and the archaeabacterial DNA binding proteins have a cluster of negatively charged residues adjacent to the N-terminus of the structured part of the chromo domain. Finally, distinct similarities in the control of transcription in archaeabacteria and eukaryotic cells have been found (Baumann et al., 1995b) and both proteins are involved in the formation of chromatin structure (Dijk and Reinhardt, 1986).

The surface charge of MoMOD1-N and Sac7d is compared in Figure 5b. It will be seen that the charge distribution on the exterior of the three-stranded β-sheet of Sac7d, which has been shown to be involved in binding non-specifically to the major groove of DNA (Baumann et al., 1995a), is not conserved in the chromo domain. There are two clusters of basic residues on the surface of the chromo domain (residues 25, 41 and 43 and residues 29, 30, 33 and 35) but overall the charge is negative, suggesting that this class of chromo domain would not bind to DNA or RNA by itself. In support of this, MoMOD1-N shows no appreciable affinity for DNA–cellulose (data not shown). It is also interesting to note that mutations in one of the clusters (R29QR30Q and K31Q, both MoMOD1 numbering) had no influence on the activity of HP1 (Platero et al., 1995). Instead, we believe that the structural homologies suggest that the
chromo domain is a protein interaction motif and that the structure of Sac7d provides a model for how the chromo domain might bind other proteins. In the structure of the chromo domain there is a hydrophobic groove around the protein, where the N-terminal hairpin would bind in Sac7d, and we speculate that protein–protein interactions occur via this groove (Figure 5c). This hypothesis is supported by previous studies which suggest that mutation of residues at the edge of this groove (Y21F) and within the groove itself (V23M) disrupt gene silencing in HP1 (Platero et al., 1995). It has been proposed that gene silencing might be controlled by tyrosine phosphorylation at this site (Platero et al., 1995; Madireddi et al., 1996) and phosphorylation should also affect protein–protein interactions in this groove. Strong support for the notion that the chromo domain mediates protein–protein interactions also comes from the results of two-hybrid screens, where both Le Douarin et al. (1996) and ourselves (N.Murzina, K.Johnson, A.De Smet and E.D.Laue, submitted for publication) have isolated different proteins that interact with the chromo domains of either the HP1α or MoMOD1 proteins respectively.

It has been proposed that the C-terminal domain may be related to the N-terminal chromo domain in HP1-like proteins (Epstein et al., 1992) and the term ‘shadow’ domain has been suggested (Aasland and Stewart, 1995). Sequence comparisons based on the structure of the MoMOD1-N chromo domain provide direct experimental support for this hypothesis. If the N- and C-terminal domains have a similar fold we would expect that residues in the hydrophobic core and key residues required to make the turns in the β-sheet would be conserved between the N-terminal chromo and C-terminal shadow domains. Examination of Figure 1 shows this to be the case. The residues in the hydrophobic core and the glycine residues in loops 1 and 2 (Gly34 and Gly44) are indeed conserved between the N-terminal chromo and many of the

<table>
<thead>
<tr>
<th>Table I. Structural statistics for the final ensemble of 27 refined structures of MoMOD1-N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root mean square deviations from restraints and idealized geometry</td>
</tr>
<tr>
<td>&lt;SA&gt;</td>
</tr>
<tr>
<td>NOE distances (Å)</td>
</tr>
<tr>
<td>Dihedral angles (°)</td>
</tr>
<tr>
<td>Bonds (Å)</td>
</tr>
<tr>
<td>Angles (°)</td>
</tr>
<tr>
<td>Impropers (°)</td>
</tr>
<tr>
<td>Final energy $E_{L-J}$ (kJ/mol)</td>
</tr>
<tr>
<td>Assessment of backbone quality according to the Ramachandran plot</td>
</tr>
<tr>
<td>All structures</td>
</tr>
<tr>
<td>Most favoured region</td>
</tr>
<tr>
<td>Additionally allowed region</td>
</tr>
</tbody>
</table>

aRepresents the average r.m.s. deviations for the ensemble.  
bRepresents values for the final structure that is closest to the mean.

Fig. 5. Comparison of the structure of the MoMOD1-N chromo domain with that of the DNA binding domain of Sac7d (Edmondson et al., 1995). (a) MOLSCRIPT (Kraulis, 1991) plots, in the same orientation as those shown in Figure 4, showing a comparison, of the structures closest to the mean, of the chromo domain from MoMOD1 (left) with that of the DNA binding domain from Sac7d (right); the side chains of residues where mutation affects gene silencing and/or chromatin binding (Messmer et al., 1992; Platero et al., 1995) are shown. (b) Electrostatic potential at the protein surfaces contoured and colour coded at –4.8 kT (red) and +4.8 kT (blue). The potential was calculated and displayed with the program GRASP (Gilson et al., 1988; Nicholls et al., 1991). Compared with the plots in Figures 4 and 5a and c, the structures in (b) have been rotated by ~130° so that the β-sheet is now facing out of the page. (c) A similar surface, but now showing the putative peptide binding site in the chromo domain and binding of the N-terminal hairpin in Sac7d.
understanding the role of chromo domain proteins in formation of chromatin structure and control of transcription.

Materials and methods

Plasmids and bacterial strains

The MoMOD1 cDNA, cloned into the Bluescript SK+ vector as an EcoRI fragment (Singh et al., 1991), was used for construction of both pET-MoMOD1 and pET-MoMOD1-N (see below). Bacterial strains used were E. coli TG1 recO [genotype K12 (lac-pro supE thi hsd D55F’ tra D38 proA+ B+ lacF’ lac Z AM15 recO::Tn5] (constructed by Dr POliver, Department of Genetics, Cambridge, UK) and BL21(DE3) (Novagen).

Construction of pET-MoMOD1 and pET-MoMOD1-N

PCR was used to subclone the regions of the MoMOD1 cDNA corresponding to amino acids residues 1–185 (for MoMOD1) and 10–80 (for MoMOD1-N). Primers 1 (5′-GGATCCCAT ATG GGG AAA AAG CA-3′) and 2 (5′-GGATCCCAT ATG GTG GAG GAG-3′) introduced an Ndel site encompassing the ATG start codon at the 5′-end of the coding sequence; primers 3 (5′-GGATCCGAATTC TCA ATT CTT GTC-3′) and 4 (5′-GGATCCGAAATTC TCA GTT ATC TG-3′) introduced a stop codon and either an EcoRI or BamHI restriction site at the 3′-end of the fragment. The PCR products were digested with Ndel and either EcoRI or BamHI and cloned into pET16b (Novagen). The ligation products were transformed into TG1 recO and the DNA sequence of the insert verified.

Expression and purification of MoMOD1 and MoMOD1-N

Escherichia coli BL21(DE3) cells were transformed with either pET-MoMOD1 or pET-MoMOD1-N and grown at 37°C in 2×TY medium containing 100 µg/ml ampicillin. Expression was induced at an OD 600 of 0.70 with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and the cells were incubated for a further 3 h. All the following procedures were carried out at 0–4°C and were monitored using 16% tricine–SDS–PAGE (Schagger and von Jagow, 1987). Cell pellets were resuspended in lysis buffer (0.5 M NaCl, 20 mM Tris, pH 7.9, and 1 mM phenylmethylsulfonyl fluoride (PMSF)) and lysed with a French press at 1000 p.s.i. The lysate was centrifuged at 35 000 r.p.m. for 1 h in a Beckman 42.1 Ti rotor and the supernatant was then loaded onto a 10 ml nickel–IDA column (Sigma Fast Flow) equilibrated with buffer A (100 mM NaCl, 20 mM Tris, pH 7.9, and 0.05% NaNa). The column was washed with buffer A and then with 60 mM imidazole in buffer A. The protein was eluted using a step gradient of imidazole (40 ml 60–500 mM, 30 ml 500 mM, 20 ml 500 mM–1 M) at a flow rate of 2 ml/min and monitored using UV spectroscopy at 254 nm. The protein eluted between 300 and 500 mM imidazole. The pooled peak fractions containing the protein were concentrated using a stirred cell (Amicon) with a YM3 ultrafiltration membrane. The buffer was exchanged for 100 mM NaCl, 50 mM Tris, pH 8.0, and 1 mM CaCl2 using an Econo-Pack 10DG column (BioRad). The N-terminal histidine tag was removed using 1:100 (w/w) Factor Xa and either EcoRI or BamHI restriction site at the 3′-end of the fragment. The digestion products were then loaded onto a nickel–IDA affinity column to remove the histidine tag and any remaining uncleaved protein.

Proteolytic digestion

Proteolytic reactions contained 50 µg protein in 1 mM triethanolamine, pH 7.5. Either trypsin or subtilisin (Sigma) was added to a final concentration of either 20, 10, 5, 1 or 0.2% w/w. The reactions were carried out at 26°C and aliquots of the reaction were taken at 5, 30, 60 and 120 min and mixed with PMSF and tricine–SDS–PAGE loading buffer. Proteolytic reactions were then repeated using selected enzyme concentrations and incubated for a single chosen time point. The digestion products were separated using tricine–SDS–PAGE or reversion phase HPLC.

Electroblotting and N-terminal sequencing

Protein fragments were separated using tricine–SDS–PAGE and transferred by semi-dry electroblotting (Pharmacia LKB) to a Problott membrane. The running buffer included 2 mM mercaptoacetic acid and a Trizma transfer buffer of 25 mM Tris, 190 mM glycine, 10% methanol and 0.02% SDS. The membrane was stained with 0.1% Coomassie R250 in 50% methanol and 1% acetic acid for 5 min, destained with 50% methanol and washed in water. The regions of the membrane containing

**Chromo domain structure**

![Chromo domain structure](image_url)
the protein bands of interest were subjected to automated N-terminal amino acid microsequence analysis.

**Reverse phase HPLC (RP HPLC) and mass spectrometry**

RP HPLC of peptide samples for electrospray ionization mass spectrometry was performed using a Hewlett Packard 1090 liquid chromatograph. The peptide samples were purified using a G108 Octadexyl (C18) column and eluted in a gradient of 0–100% acetonitrile containing 0.1% (v/v) trifluoroacetic acid.

**NMR spectroscopy**

Uniformly $^{13}$C/15N- and 15N-labelled MoMOD1-N proteins were prepared for the NMR experiments by growing E. coli BL21(DE3) (Novagen) harbouring the expression plasmid in a MOPS minimal medium containing either 10% $^{15}$N- or $^{13}$C/15N-labelled Celltone (Marteck) and $^{15}$NNaCl, either with or without [13C]glucose. The protein samples were purified as described above. Four protein samples of MoMOD-N were prepared at pH 6.0 in a buffer containing 10 mM KH$_2$PO$_4$, 25 mM NaCl and 0.05% NaN$_3$; a 2.0 mM 15N-labelled sample, a 2.7 mM 13C/15N-labelled sample, a 3.5 mM unlabelled sample, all in 90% H$_2$O, 10% D$_2$O, and a 1.0 mM 13C/15N-labelled sample in 100% D$_2$O.

All spectra were acquired using a Bruker AMX600 spectrometer at 298 K as described in either Kraulis et al. (1994) or the original references. For the identification of slowly exchanging amide protons, the 15N-labelled 90% H$_2$O, 10% D$_2$O sample was buffer exchanged into 99.8% D$_2$O by 20-fold dilution followed by concentration using a stirred cell Amicon with a YM3 ultrafiltration membrane. All the data were processed using the AZARA software package (W.Boucher, unpublished data) and analysed with the program ANSIG 3.3 (Kraulis et al., 1994) on Silicon Graphics Iris workstations. Both of these programs are available by anonymous ftp to ftp.bio.cam.ac.uk in the directory “ftp/pub.”

**Structure calculations**

The structures were calculated from the NMR data using the program X-PLOR (Brunger, 1992), but with minor modification to the files sa.inp and refine.inp from the X-PLOR 3.1 release (M.Nilges, personal communication). Coarse structures correlating backbone 15N with sidechain H$_y$ resonances in larger proteins, J. Magn. Reson., 95, 636–641.


Chromatin domain structure


Received on December 24, 1996; revised on January 27, 1997

Note added in proof

After this paper had been accepted we became aware of work that shows that HP1 also interacts with the lamin B receptor [Ye,Q. and Worman,H.J. (1996) J. Biol. Chem., 271, 14653–14656].