Structure of the Chromo Barrel Domain from the MOF Acetyltransferase*

Received for publication, February 4, 2005, and in revised form, June 15, 2005 Published, JBC Papers in Press, June 17, 2005, DOI 10.1074/jbc.M501347200

Peter R. Nielsen†, Daniel Nietlispach‡, Alessia Buscaino§, Rosemary J. Warner‡, Asifa Akhtar§, Alexey G. Murzin*, Natalia V. Murzina†, and Ernest D. Laue†,‡

From the*Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge, CB2 1GA, United Kingdom, ‡Gene Expression Program, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany, and †Centre for Protein Engineering, Medical Research Council Hills Road, Cambridge, CB2 2QH, United Kingdom

We report here the structure of the putative chromo domain from MOF, a member of the MYST family of histone acetyltransferases that acetylates histone H4 at Lys-16 and is part of the dosage compensation complex in Drosophila. We found that the structure of this domain is a β-barrel that is distinct from the α + β fold of the canonical chromo domain. Despite the differences, there are similarities that support an evolutionary relationship between the two domains, and we propose the name “chromo barrel.” The chromo barrel domains may be divided into two groups, MSL3-like and MOF-like, on the basis of whether a group of conserved aromatic residues is present or not. The structure suggests that, although the MOF-like domains may have a role in RNA binding, the MSL3-like domains could instead bind methylated residues. The MOF chromo barrel shares a common fold with other chromatin-associated modules, including the MBT-like repeat, Tudor, and PWWP domains. This structural similarity suggests a probable evolutionary pathway from these other modules to the canonical chromo domains (or vice versa) with the chromo barrel domain representing an intermediate structure.

The chromo domain was originally defined as a region of homology between heterochromatin protein 1 (HP1) and Polycymph (Pc) and is found in many other chromatin-associated proteins (2, 3). Later, it was recognized that the HP1 family of proteins also contain a second highly related chromo shadow domain (2–4), which unlike the chromo domain, proves to be dimeric (5, 6). Structure determination of the chromo domain suggests that they might act as protein interaction modules, including the MBT-like repeat, Tudor, and PWWP domains. This structural similarity supports a probable evolutionary pathway from these other modules to the canonical chromo domains (or vice versa) with the chromo barrel domain representing an intermediate structure.

The dosage compensation mechanism in Drosophila is different from that in mammals. Genes in both female X chromosomes are actively expressed, but the single male X chromosome is transcribed at a 2-fold higher level to give an equal level of transcription in both sexes. This results from the differential expression of the Drosophila dosage compensation complex (see Scheme 1), which is known to include the MLS1, MLS2, MLS3, MLE, and MOF proteins, as well as two non-translated RNAs, roX1 and roX2 (22). MOF is a histone acetyltransferase that acetylates histone H4 at Lys-16, and this modification is specific for the dosage compensation mechanism (23–25). In Schneider cells, association of both MOF and MLS3 to the X chromosome is sensitive to RNase treatment (21, 26), and point mutations in the MOF chromo-related domain severely affect the interaction of MOF with RNA (21). Here we present the solution structure of the putative chromo domain from MOF, show that this new type of domain (chromo barrel) does not by itself bind RNA, and explore its structural and functional relationships with other proteins.

MATERIALS AND METHODS

Expression and Purification of the MOF Chromo Barrel Domain—The DNA sequence corresponding to amino acids 367–454 of Drosophila MOF was inserted into the NdeI and BamHI sites of the pET15b vector (Novagen). The vector was transformed into the expression strain Escherichia coli TUNER (DE3) pLacI (Novagen) and the protein purified using standard methods (see supplemental information).

* This work was supported by The Wellcome Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† The on-line version of this article (available at http://www.jbc.org) contains supplemental text and references.

The atomic coordinates and structure factors (code 2bud) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

To whom correspondence may be addressed. E-mail: nm@mole.bio.cam.ac.uk.

To whom correspondence may be addressed. E-mail: e.d.laue@bioc.cam.ac.uk.

The abbreviations used are: HP1, heterochromatin protein 1; Pc, Polycymph; NOE, nuclear Overhauser effect; MBT, malignant brain tumor.
NMR Spectroscopy and Structure Calculations—All NMR experiments were recorded at 25 °C on Bruker DRX 600 and 800 MHz spectrometers equipped with 5-mm triple resonance H/C/N z-gradient probes. Samples used for structural work were all at pH 5 and 1 mM concentration and contained 10% D2O. Structure calculations were performed using crystallography NMR software, version 1.0 (27) and ARIA, version 1.1.2 (28). Further details are provided in the supplemental information.

RESULTS AND DISCUSSION

Definition of the Structured Domain—The fragment of MOF used in the structural studies (Fig. 1a) was chosen based on an alignment of canonical chromo domain amino acid sequences together with a sequence alignment and structure prediction of the new family of chromo barrel domains. The chromo barrel domain contains the conserved chromo box motif (Fig. 1b), but in addition, it is possible to recognize a pattern of aliphatic residues that define a strand equivalent to β1 in the chromo domain. N- and C-terminal to this, there are other clusters of aliphatic residues that we thought might also be part of the structure. This sequence, residues 367–440, was then extended at the C terminus until it included a good number of closely spaced charged residues (which were unlikely to be structured), resulting in a final construct composed of MOF residues 367–454.

This construct was used to produce protein for NMR experiments. The domain was found to be monomeric by analytical ultracentrifugation (data not shown). Due to its high level of stability, it was possible to obtain a large number of structural restraints, and the resulting ensemble of structures has a high degree of precision (TABLE ONE). Fig. 2a shows an overlay of the 30 lowest energy structures aligned over the secondary structural elements; the root mean square deviation over the backbone atoms of the well defined part of the construct (residues 373–443) was 0.7 Å. The backbone dynamics of the structure was investigated by measuring the 15N T1 and T2 relaxation rates and the 1H-15N heteronuclear NOE. These experiments defined the structured region as residues 374–437, with residues N-terminal to 370 and C-terminal to 441 being highly unstructured (data not shown). A model free analysis of the T1 and T2 relaxation rates suggested that the domain had an overall tumbling time of 7.2 ns, and the S2 values also suggested that the well defined region comprised residues 374–438 (S2 > 0.8 for data recorded on a 600-MHz spectrometer). However, a careful analysis of the NOE data showed that unambiguous NOEs that connected residues 439–443 to the core of the structure could be observed. It could be concluded, therefore, that these residues were probably only in this conformation transiently. Likewise, at the N terminus, there were NOEs connecting residues 370, 371, and 373 to the core of the structure, and these residues also probably only interacted transiently. In conclusion, the structured domain comprised residues 370–443, but the N and C termini only interacted transiently with the core of the protein.

The Fold of the Chromo Barrel Domains—The structure of the chromo barrel domain is virtually all β. It consists of five β-strands (β1, residues 383–386; β2, residues 392–401; β3, residues 411–416; β4, residues 425–428; and β5, residues 432–433) organized into two sheets of three strands with one strand, β2, spanning the two sides of the domain. Strands β5, β1, and β2 form one side of the barrel and strands β2, β3, and β4, the other. A bulge in strand β2 (residues 398–399) helps provide the required geometry (Figs. 1b and 2). The β-barrel comprises amino acids 383–433, and the alignment clearly shows that the β-strands are conserved among the chromo barrel domains (Fig. 1b). However, as discussed above, the NOE data suggest that residues 370–443 are structured. The residues N-terminal to the β-barrel form a turn-like structure, with Ile-373 involved in forming a hydrophobic lid capping the β-barrel (Fig. 2c). This turn only forms transiently in the MOF domain, but analysis of the sequence alignment of the chromo barrel domains shows that a pair of large hydrophobic residues and one with a small side chain are conserved in the N terminus (see residues in green in Fig. 1b). This hydrophobic pair is typically separated by only two other residues, compared with the six in MOF, predicting a more stable hairpin turn in the chromo barrel domains from other proteins (e.g. MSL3). The residues C-terminal to strand β5 form one turn of the helix in the MOF domain but are not conserved in the chromo barrel family.

Comparison of the Chromo and Chromo Barrel Domains—It is clear that the fold of the MOF chromo barrel domain is different from that of the chromo domain. The C-terminal helix of the chromo domain is absent in the chromo barrel, whereas two of the five β-strands are not present in the chromo domain (see Figs. 1b and 3). The residues conserved between the two types of domains, the originally defined chromo box motif, correspond to the β-hairpin formed by strands β2 and β3 in the chromo domain and strands β3 and β4 in the chromo barrel. The structural relationship between the chromo and chromo barrel domains is supported by a further feature: Arg-401 in MOF and Trp-426 in Pc, mouse M334 chromo domains, the structures reveal exactly the same interaction. The tryptophan is conserved in most of the domains, but the arginine is often absent (Fig. 1b). In most cases where this arginine is absent, the amino acid corresponding to Ser-400 is a large hydrophobic residue, and this may instead pack against the Trp. It is also possible that interactions with the conserved Tyr/Phe at position 412 compensate for the loss of the Arg to Trp stacking.

The three histidine residues in the domain are all involved in stabilizing the structure. The chemical shifts of the Nε2 proton in the imidazole side chains show that, at pH 5, His-393 and -430 are still partly deprotonated. The chemical shift changes of 1H-15N pairs in heteronuclear single quantum correction spectra recorded at pH values of 5, 6, 7, and 8 also allowed an estimation of the pKε values of the histidines:

4 S. M. Southall, P. R. Nielsen, D. Nettlispach, N. V. Murzina, and E. D. Laue, unpublished results.
His-393, pKₐ ~ 6; His-430, pKₐ ~ 6.5; and His-415, pKₐ ~ 7.5. His-415 and Asp-424 form a salt bridge that fixes the ends of the loop between residues 3 and 4. This salt bridge is likely to be conserved among most of the chromo barrel and some of the chromo domains (Fig. 1b).

Strikingly, however, when the chromo barrel domain is compared with that of the HP1 chromo domain in complex with the Lys-9-methylated N-terminal tail of histone H3 (H3-K9), there is a much higher degree of structural identity. In the MOF chromo barrel domain, the strand corresponding to the histone peptide in the chromo domain complex is provided by the N terminus of the protein, similar to the situation with the archaeal proteins Sac7d/Sso7d (29, 30) (see Fig. 4). This interaction prevents “chromo domain-like” peptide interactions. When the structures are aligned over the five beta-strands, the root mean square deviation over the backbone atoms is only 1.3 Å. In effect, therefore, the chromo barrel structure is an “auto-inhibited” chromo domain, where methylated peptide binding is prevented by interactions of the protein with its own N-terminal tail.

**Function of Chromo Barrel Domains**—The structure of the chromo domain-histone H3-K9 complex (14, 15) shows that binding of the dimethylamino group of the lysine requires a cluster of aromatic residues on the surface of the protein. However, these three critical aromatic residues in the chromo domain are not conserved in the MOF chromo barrel domain; MOF has His-393, Tyr-416, and Leu-419, whereas HP1 has Tyr-21, Trp-42, and Phe-45 (see Figs. 1b and 3b). In addition, superposition of the two structures shows that the MOF residue Arg-387 would clash with the methylated Lys-9 side chain in the chromo domain-histone H3 complex (Fig. 3b). Consequently, binding of methylated residues is prevented in the MOF chromo barrel domain.

The residues corresponding to the methyllysine binding aromatic box are, however, present in a subset of the chromo barrel domains (Fig. 4).
The loop between strands β3 and β4 is less well ordered than the rest of the structure. The cross peaks corresponding to residues 403 and 410 in the 1H,15N heteronuclear single quantum correlation spectrum are broadened indicative of slow time scale motions, and although the loop appears well defined in the ensemble, the Ramachandran plot reveals that several of the residues are constrained to unfavorable conformations, probably due to averaging of measured NMR parameters. It is however, likely that this loop is not completely flexible, because the values of the cross-correlated cross-relaxation rates (see “Materials and Methods”) are not averaged, as is the case for residues in the N- or C-terminal regions.

### Structure of the Chromo Barrel Domain

#### FIGURE 2. Structure of the MOF chromo barrel domain. a, stereo view of the backbone structure (N, Cα, CO) from the 30 (of 100 computed) lowest energy structures obtained from the final set of calculations using ARIA. The well defined part of the structure (residues 373–438, root mean square deviation over the backbone atoms of 0.7 Å) is shown in blue, whereas residues 370–373 and 439–443, which have transient structure, are shown in yellow. b and c, schematic representations of the MOF chromo barrel domain in two different orientations. The structure in c is rotated 45° around x relative to b. Strands β1, β2, β3, β4, and β5 are shown in blue, cyan, green, orange, and red, respectively.
Structure of the Chromo Barrel Domain

FIGURE 3. Comparison of the structures of the MOF chromo barrel and the HP1 chromo domain-histone H3 peptide complex. a, superposition (using residues in the \( \beta \)-strands) of the MOF chromo barrel and the HP1 chromo domain (residues 17–75)-histone H3 peptide (residues 4–11) complex. b, close up view of the methyl lysine binding site in a (where the structures are now superimposed using residues in the aromatic site) showing the three residues that form the aromatic pocket. Side chains from the MOF domain are shown in blue, from the HP1 chromo domain, in red, and from histone H3, in yellow. N, N terminus; C, C terminus.

FIGURE 4. The family of structures related to the chromo domain. The five strands in the MOF chromo barrel are color coded as in Fig. 2 with the remainder of the backbone in gray, apart from the loop connecting strands \( \beta 2 \) and \( \beta 3 \), which is yellow. In the HP1 chromo domain-histone H3 complex, the H3 peptide is black, and the methyl-K9 is purple. In the HP1 chromo shadow domain-CAF-1 p150 complex, one of the monomers is shaded in softer tones, and the CAF-1 peptide is pink. Sso7 is also shown in its complex with DNA. Strands that are equivalent to those in the MOF chromo barrel are shown using the same color scheme, and the positions of the conserved aromatic residues are indicated by spheres. The Protein Data Bank accession numbers for the structures used in the Fig. are as follows: MBT domain, 1oi1; Tudor domain, 1mhn; HP1 chromo domain-histone H3 complex, 1guw; HP1 chromo shadow domain-CAF-1 p150 complex, 1s4z; and the Sso7-DNA complex, 1bf4.

uous peptide contexts, and they therefore need not (and likely do not) share a common peptide-binding site. They do, however, share a putative functional site with each other and with other superfamily members.

Despite the fact that they have now diverged, the chromo and chromo-barrel domains, as well as the MBT repeat and Tudor domains, all contain a similar cluster of aromatic residues that, in the HP1 and Pc chromo domains, recognize methylated lysine (see Figs. 1b and 4). In each of these families, these aromatic residues are not always conserved suggesting that they are not essential for structure or stability. However, in each family, many members retain aromatic residues in all three of the crucial methyllysine-binding positions found in the HP1 and Pc chromo domains. In addition, two of the three aromatic sites are conserved in some PWWP domains, and these have an alternative site for a third conserved aromatic residue. Furthermore, a fourth aromatic position, which has recently been shown to be important for methyllysine binding in yeast CHD1 (32), is also conserved in the MSL3-like chromo barrels, the Pc chromo domains, and some of the MBT repeats.

Strands \( \beta 2 \) and \( \beta 3 \) in the chromo domain or \( \beta 3 \) and \( \beta 4 \) in the chromo barrel domain together with the connecting loop that carries two of the three key aromatic residues correspond to the originally defined chromo box motif (Fig. 1b). The structure of this chromo box motif has been retained in evolution, because it contains a number of conserved interactions, in particular the conserved stacking interaction involving Trp-426 (Fig. 1b). In the MBT repeat and Tudor domain, the loop has a somewhat different conformation, but the aromatic residues are in similar positions (Fig. 4). This suggests that this aromatic residue pocket was present in a common ancestor, but is not now required in all of these proteins. The hypothesis that members of all branches of this superfamily of proteins might be involved in binding methylated residues is supported by a recent study of the 53BP1 protein (31). There the conserved aromatic clusters in the tandem Tudor domains in 53BP1 were found to have a role in the recognition of methylated Lys-79 of histone H3 (H3-K79). Although H3-K79 probably binds in the cleft between the two Tudor domains, i.e. not as H3-K9 does to the chromo domain, its methyl group(s) does likely bind the aromatic site in the N-terminal domain. An interaction of this type has also recently been identified in the lethal (3) MBT protein.3

In addition to the aromatic site, in the chromo barrel, MBT repeat, and Tudor domains, the C terminus forms an extended interaction down one side of the barrel, whereas the N terminus forms a stabilizing hook that caps the \( \beta \)-barrel (Fig. 4). It is these common structural features, together with the conserved aromatic site, that suggests that these proteins are evolutionarily related to each other and which sets them apart from other similar \( \beta \)-barrel folds, e.g. the SH3 domains. As well as revealing the fold of a distinct family of chromo barrel domain proteins, the structure also suggests a probable evolutionary pathway between the chromo barrel and canonical chromo domains involving a dramatic perturbation of structure, the extension of the C-terminal \( \alpha \)-helix and loss of the N-terminal \( \beta \)-strand, resulting in the formation of the chromo domain peptide-binding site. By contrast, the dimeric chromo shadow domain, with its dimerization interface harboring a new peptide-binding site, has probably evolved from the canonical chromo
domain by a more conventional pathway of point mutation, insertion, and deletion. This has resulted in a change in conformation of the loop carrying the two critical aromatic residues, as well as an insertion of two residues in the loop connecting the β-sheet to the α-helix, which disrupts the binding site for methylated peptides in chromo shadow domains. The similarity of the chromo barrel and chromo domains to the common fold of the MBT repeat, Tudor, and PWWP domains supports the hypothesis that all of these “Royal Family” domains share a common ancestry (33, 34, 35).

Acknowledgments—We thank Prof. P. Becker for advice and encouragement and amino acid analysis.

Addendum—The solution structure of a Tudor domain from a mouse hypothetical protein has recently been determined by the RIKEN Structure Genomics/Proteomics Initiative (H. Li, K. Saito, M. Inoue, T. Kigawa, and S. Yokoyama, unpublished data; Protein Data Bank entry 1WGS). On the basis of its sequence similarity, this domain belongs to the chromobarrel family, and it is structurally most similar to the MOF domain.

REFERENCES

1. Paro, R., and Hogness, D. S. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 263–267
2. Koonin, E. V., Zhou, S., and Lucchesi, J. C. (1995) Nucleic Acids Res. 23, 4229–4233
3. Aasland, R., and Stewart, A. F. (1995) Nucleic Acids Res. 23, 3168–3174
4. Epstein, H., James, T. C., and Singh, P. B. (1992) J. Cell Sci. 263–267
5. Broadhurst, R. W., Ball, L. J., Murzina, N. V., and Laue, E. D. (2000) EMBO J. 19, 1587–1597
6. Cawson, P. N., Partridge, J. F., Allshire, R. C., and McLaughlin, P. J. (2000) Curr. Biol. 10, 517–525
7. Ball, L. J., Murzina, N. V., Broadhurst, R. W., Raine, A. R., Archer, S. J., Stott, F. J., Murzin, A. G., Singh, P. B., Domaillé, P. J., and Laue, E. D. (1997) EMBO J. 16, 2473–2481
8. Bannister, A. J., Zegerman, P., Partridge, J. F., Misca, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001) Nature 410, 120–124
9. Lachner, M., O’Carroll, D., Mechtler, K., and Jenuwein, T. (2001) Nature 410, 116–120
10. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002) Science 298, 1039–1043
11. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002) Genes Dev. 16, 2893–2905
12. Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O’Connor, M. B., Kingston, R. E., and Simon, J. A. (2002) Cell 111, 197–208
13. Li, Y., Kirschmann, D. A., and Wallrath, L. L. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 16462–16469
14. Nielsen, P. R., Nielitsch, M., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2002) Nature 416, 103–107
15. Jacobs, S. A., and Khorasanizadeh, S. (2002) Science 295, 2080–2083
16. Thiru, A., Nielitsch, M., Mott, H. R., Okuwaki, M., Lyon, D., Nielsen, P. R., Hirshberg, M., Verreault, A., Murzina, N. V., and Laue, E. D. (2004) EMBO J. 23, 489–499
17. Brehm, A., Tufteland, K. R., Allshere, R. C., and Becker, P. B. (2004) BioEssays 26, 133–140
18. Bouzourene, K., Mitterweger, A., Langat, G., Imhof, A., Akhtar, A., Becker, P. B., and Brehm, A. (2002) EMBO J. 21, 2430–2440
19. Nielsen, P. R., Callaghan, J., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2004) Methods Enzymol. 376, 148–170
20. Martin, I., and Baker, R. S. (2000) Mol. Biol. Evol. 17, 1240–1250
21. Akhtar, A., Zink, D., and Becker, P. B. (2000) Nature 407, 405–409
22. Akhtar, A. (2003) Curr. Opin. Genet. Dev. 13, 161–169
23. Akhtar, A., and Becker, P. B. (2000) Mol. Cell 5, 367–375
24. Hiffiker, A., Hiffiker-Kleiner, D., Pannuti, A., and Lucchesi, J. C. (1997) EMBO J. 16, 2054–2060
25. Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D., and Lucchesi, J. C. (2000) Mol. Cell. Biol. 20, 312–318
26. Buscaino, A., Kocher, T., Kind, J. H., Holz, H., Taipale, M., Wagner, K., Wilm, M., and Akhtar, A. (2003) Mol. Cell 11, 1265–1277
27. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998). Acta Crystallogr. Sect. D 54, 905–921
28. Linge, I. P., O’Donoghue, S. J., and Nilges, M. (2001) Methods Enzymol. 339, 71–90
29. Gao, Y. G., Su, S. Y., Robinson, H., Padmanabhan, S., Lim, L., McCray, B. S., Edmondson, S. P., Shriver, J. W., and Wang, A. H. (1998) Nat. Struct. Biol. 5, 782–786
30. Robinson, H., Gao, Y. G., McCray, B. S., Edmondson, S. P., Shriver, J. W., and Wang, A. H. (1998) Nature 392, 202–205
31. Huyen, Y., Zghieb, O., Delteil, R. A., Jr., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., Shenton, E. A., Mellert, H. S., and Stavridi, E. S. (2004), Nature 432, 406–411
32. Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R., and Grant, P. A. (2005) Nature 433, 434–438
33. Sathyamurthy, A., Allen, M. D., Murzin, A. G., and Bycroft, M. (2003) J. Biol. Chem. 278, 46968–46973
34. Wang, W. K., Tersukho, V., Boccuni, P., MacGrogan, D., Nimer, S. D., and Patel, D. J. (2003) Structure (Camb.) 11, 775–789
35. Maurer-Stroh, S., Dickens, N. J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C. P. (2003) Trends Biochem. Sci. 28, 69–74
