Crystal Structure of the Malignant Brain Tumor (MBT) Repeats in Sex Comb on Midleg-like 2 (SCML2)*

Received for publication, June 18, 2003, and in revised form, July 29, 2003
Published, JBC Papers in Press, September 1, 2003, DOI 10.1074/jbc.M306469200

Aruna Sathyamurthy‡, Mark D. Allen, Alexey G. Murzin, and Mark Bycroft§

From the Medical Research Council Centre for Protein Engineering, Hills Road, Cambridge, CB2 2QH, United Kingdom

Sex Comb on Midleg (SCM) belongs to the Polycomb group of proteins, which are involved in transcriptional regulation in Drosophila. It is one of the components of Polycomb repressive complex 1, a multiprotein complex of Polycomb group proteins involved in the maintenance of repression and the blocking of chromatin remodeling. SCM contains two ∼100-residue malignant brain tumor (MBT) repeats at the N terminus. These repeats are also found in other proteins involved in transcriptional regulation. Here, we report the 1.78-Å crystal structure of the two MBT repeats of SCM-like 2 (SCML2), a human homologue of SCM. Each repeat consists of an extended arm and a β-barrel core. There are significant structural similarities to the Tudor, PWWP, and chromo domains, suggesting probable evolutionary relationships and functional similarities between the MBT repeats and these domains.

Polycomb group (PcG) genes are transcriptional repressors that play a key role in the establishment and maintenance of cell identity in animals. They were originally identified in Drosophila melanogaster where they are required for the silencing of homeotic loci during development (1, 2). In mammals, PcG genes have been implicated in the control of cellular proliferation and tumorigenesis (3).

PcG genes encode the components of two multiprotein complexes. Polycomb repressive complex 1 (PRC1) contains as core subunits the gene products Polycomb (PC), Polyhomeotic, Posterior Sex Combs, and dRING1 (4–6). The second complex contains Extra Sex Combs and Enhancer of Zeste (7, 8). The Extra Sex Combs-Enhancer of Zeste complex is expressed early in embryogenesis and is thought to play a role in the initiation of silencing. The Extra Sex Combs-Enhancer of Zeste complex catalyzes the modification of chromatin. It has histone deacetylase activity and can also methylate lysine residues in histones. Some of these modifications, such as the methylation of histone H3 at lysine 27, facilitate the binding of the PRC1 complex to chromatin. The PRC1 complex is involved in the maintenance of repression and can block chromatin remodeling by SWI/SNF helicases (4). The product of the PcG gene Sex Comb on Midleg (SCM) copurifies with PRC1 and colocalizes with the PSC gene product on polytene chromosomes (9). SCM plays a key role in PcG repression of gene expression, as embryos deficient in both maternal and zygotic copies of the gene die with severe homeotic transformations (10, 11). SCM protein consists of a C-terminal sterile motif (SAM) domain and two ∼100 amino acid repeats in the N-terminal portion of the protein (12). The product of the D. melanogaster tumor suppressor gene lethal (3) malignant brain tumor has a similar domain organization but contains three copies of the repeat (13). Because of their occurrence in this protein, these repeats have been termed malignant brain tumor (MBT) repeats. A third D. melanogaster protein containing four repeats has also been identified. Proteins with similar domain organizations are found in mammals.

The SAM domain of SCM has been shown to bind to the SAM domain of PC and is probably responsible for the recruitment of SCM to the PRC1 complex (14). The role of the MBT repeats is less clear. Several SCM mutant alleles map within the MBT repeats, suggesting that they have an important role (15). The MBT repeat region of human lethal (3) MBT has been shown to be required for transcriptional repression, although the molecular basis for this activity is not known (16). To try to learn more about the function of MBT repeats, we have determined the crystal structure of an N-terminal fragment, a human homologue of SCM called SCML2, containing two MBT repeats in tandem.

EXPERIMENTAL PROCEDURES

Cloning of the Gene—The SCML2 MBT repeats were cloned from a human multiple tissue cDNA library (Clontech) with 5′- and 3′-end primers containing BamHI and EcoRI restriction sites, respectively. Three clones were made, i.e. MBT 1 (Gly23–Tyr134), MBT 2 (Trp131–Ser243) and MBT 1 + 2 (Gly23–Ser243) (Gly23 is a result of having BamHI as the restriction site). The resulting PCR products were ligated into the cloning sites of modified mini-pRSET (A) with an N-terminal hexahistidine tag.

Site-directed Mutagenesis—The wild-type MBT (1 + 2) protein had an internal thrombin cleavage site. Hence, the mutant K147E was made using the QuikChange kit (Stratagene).

Protein Expression and Purification—The plasmids were transformed into Escherichia coli C-41 host cells and grown on a tryptone/yeast extract (TY)-agar plate containing 50 μg/ml ampicillin. A single colony of each clone was used to inoculate 1 ml of 2× TY medium with ampicillin. The cultures were incubated at 37 °C until they reached an OD600 of 0.8. Expressions were induced with 0.1 mM isopropyl-β-D-galactopyranoside, and the bacteria were harvested after 4 h at 37 °C by centrifugation.

MBT 1, MBT 2, and the pseudo-wild-type MBT (1 + 2) proteins were purified by affinity chromatography using nickel-nitritotriacetic acid resin (Qiagen). Bound protein was cut with thrombin and loaded on a second nickel column. The untagged protein eluted without binding to the affinity column. The concentrated protein (12 mg/ml) was applied to
RESULTS AND DISCUSSION

Stability of the MBT Repeats—Of the three proteins expressed, only MBT (1 + 2) (GlySer2) could be purified. MBT 1 (Gly–Tyr) could be expressed but aggregated upon thrombin cleavage. MBT 2 (Trp141–Ser243) did not express well. The wild-type protein MBT (1 + 2) was prone to degradation by thrombin, which was identified by SDS gel and analyzed by mass spectrometry. The protein fragments could not be separated from the full-length protein. The mutant K147E, at the thrombin-cleavage site, gave pure protein for further studies and yielded crystals. The wild-type protein did not give any crystals, probably because of the inseparable cleavage product.

Crystal Structure—The crystal structure of the pseudo-wild-type of the two MBT repeats was solved to a resolution of 1.78Å using MIRAS. The crystallographic data are presented in Table I. Electron density for the first ten residues (Gly3–Asp12) at the N terminus and four residues (Met136–Ser139) in the linker region is missing. The final refinement gave 308 water molecules and one PEG molecule in an asymmetric unit. 92.3% of residues are present in the maximum allowed region and none in the disallowed region of the Ramachandran plot. The final structure has a Rfree value of 20.4, a Rwork value of 17.1, and an overall temperature factor of 16.09. The model has root mean square deviations in bond length of 0.004 Å and in bond angle of 1.3°.

The crystal structure of the pseudo-wild-type MBT 2 (Trp141–Ser243) did not express well. The wild-type protein MBT (1 + 2) was prone to degradation by thrombin, which was identified by SDS gel and analyzed by mass spectrometry. The protein fragments could not be separated from the full-length protein. The mutant K147E, at the thrombin-cleavage site, gave pure protein for further studies and yielded crystals. The wild-type protein did not give any crystals, probably because of the inseparable cleavage product.

Crystal Structure—The crystal structure of the pseudo-wild-type of the two MBT repeats was solved to a resolution of 1.78Å using MIRAS. The crystallographic data are presented in Table I. Electron density for the first ten residues (Gly3–Asp12) at the N terminus and four residues (Met136–Ser139) in the linker region is missing. The final refinement gave 308 water molecules and one PEG molecule in an asymmetric unit. 92.3% of residues are present in the maximum allowed region and none in the disallowed region of the Ramachandran plot. The final structure has a Rfree value of 20.4, a Rwork value of 17.1, and an overall temperature factor of 16.09. The model has root mean square deviation in bond length of 0.004 Å and in bond angle of 1.3°.

The overall structure is presented in Fig. 1A. Each repeat consists of an extended “arm” and a globular core. The arm of the first repeat packs against the core of the second repeat and vice versa. The structure of the core-interacting part of each arm consists of an N-terminal a-helix and a turn of 3₁₀ helix connected by a short 3₃ helix. The core consists of an Src homology 3-like five-stranded β-barrel followed by a C-terminal α-helix and another short β-strand. Each arm interacts...
FIG. 1. A, MOLSCRIPT (37) diagram of the overall structure of MBT repeats of SCML2 (PDB code 1OI1). The molecule is colored from the N to the C terminus following the colors of the visible spectrum (violet for the N terminus and red for the C terminus). B, model of the structure.
with its partner core in a similar way, with the orientation of the N-terminal helix relative to the barrel varying slightly. There are also extensive interactions between the two barrels. Thus, the two repeats form one inter-linked structural unit, which explains why it was not possible to express them individually. The interaction between the arm and the core units is reminiscent of “domain swapping” in homo-oligomeric proteins (24, 25).

The two barrels contact each other non-symmetrically using different sides. The residues in the barrel interface are hydrophobic and generally conserved in MBT repeats, suggesting that the formation of other multi-repeat structures may involve similar barrel interfaces and arm exchanges. These packing arrangements can easily accommodate three or four repeats and are likely to occur in all MBT repeat-containing proteins. A third repeat can readily be added to the present structure without steric clashes, and the addition of two repeats requires only minor adjustments of the observed barrel interface. The assembly of three or four repeats would form a ring-like structure with the N-terminal arm packing against the C-terminal barrel and the other arms packing against the preceding barrel. A model of the assembly of four repeats is shown in Fig. 1B. The three-repeat model is consistent with the experimental structure published after the submission of this paper (26).

Conserved residues are distributed throughout the repeat within both the barrel and the N-terminal extension. There are a number of conserved glycine and proline residues in the extensions situated where the polypeptide chain changes direction, indicating that the conformation of these regions is conserved (Fig. 1C). Conserved hydrophobic residues are present both within the core of the barrel and in the N-terminal and C-terminal extensions. The high degree of sequence conservation in the N-terminal extension suggests that it is present in all MBT repeats. There is a highly conserved glutamate in the first strand of the barrel (see alignment in Figs. 1C and 2), which was suggested previously to have a functional role. This residue hydrogen bonds to the N terminus of the C-terminal helix and appears to play a critical structural role in the formation of the arm-binding surface. Most of other residues in the repeats also appear to be conserved for structural reasons; for example, a lysine residue at the start of the first strand of the barrel and an aspartate residue at the end of the third strand form a salt bridge that stabilizes the barrel. The exceptions are Asp73 and Asp105, which have no apparent structural role and are presumably conserved for function (Fig. 1C).

SCML2 has a high degree of sequence similarity with Drosophila SCM within the MBT repeats. Three Scm mutant alleles have been identified that map within the first MBT repeat of the protein (Fig. 3). One of these is an in-frame deletion that removes four amino acids at the C terminus of the repeat. This deletion is likely to have a significant effect on the packing of the repeats. The other two are missense mutations. In one of these, a conserved aspartate residue is changed to an asparagine. This is a very conservative mutation of a solvent exposed residue that does not appear to have a structural role. This strongly suggests that this residue is functionally important. In the other mutation, a buried valine in the hydrophobic core of the β-barrel is substituted to glutamate. This is likely to have a destabilizing effect on the domain and could cause a local disruption of the structure.

**Structural Similarities—Comparison of the structure of the MBT repeats to other protein structures by eye and by using**

![Crystal Structure of SCML2 MBT Repeats](Image 263x442 to 563x737)

**Fig. 2.** MOLSCRIPT representation of some conserved residues in the hydrophobic core, the conserved glutamate Glu^{179} (E179) and the two extra residues, Arg^{222} (R222) and Asp^{223} (D223) present in the turn between the fourth and fifth strand. Other residues shown are Trp^{35} (W35), Leu^{39} (L39), Pro^{55} (P48), Phe^{52} (F52), Phe^{172} (F172), Gly^{175} (G175), Met^{176} (M176), Lys^{177} (K177), Leu^{178} (L178), and Ala^{180} (A180).
the DALI (27) server reveals a significant similarity within the β-barrel portion of the structure to the Tudor domain of the survival motor neuron (SMN) protein (53 residues) superimposed with an root mean square deviation of 1.4 Å and to the PWWP domain with 77 residues overlapping at an root mean square deviation of 3.1 Å (28–30). This degree of structural similarity is in a good agreement with the prediction that the MBT, Tudor, and PWWP domains are all members of a homologous protein superfamily (29, 31). The Tudor and PWWP domains are not typically found in tandems, and the similarity does not extend to the N- and C-terminal extensions present in the MBT repeat. The most significant difference within the β-barrel of the MBT repeat compared with the Tudor and PWWP barrels is an insertion of two extra residues in the turn between the fourth and fifth strand (Fig. 2). This insertion widens the top of the MBT repeat barrel, resulting in the formation of a hydrophobic pocket that accommodates the conserved aromatic residue in the arm. The conservation of this insertion in the MBT repeats reiterates the importance of the MBT repeat assembly for their biological function. Although there is no identified ligand for MBT repeat proteins, it can be suggested as well. Also, there is an important structural similarity between the Tudor and PWWP barrels and the complex of the chromo domain with a methylated lysine peptide from the histone H3 tail (Fig. 4) (32, 33). The histone peptide is bound in a β-strand conformation and occupies the site corresponding to the first strand of the common Tudor/PWWP/MBT barrel. The methylated lysine binds to a conserved aromatic pocket in the equivalent site to the proposed methylated arginine-binding site of the SMN Tudor domain. In the chromo domain complex, the conserved loop forming a part of the methylated lysine-binding site adopts a conformation close to that of the equivalent loop between the third and fourth strand of the MBT repeat barrel. In the MBT loop, there are three conserved aromatic residues, two of which are also conserved in the chromo and Tudor domains. These aromatic residues are conserved in some but not all MBT repeats; in SCML2 they are found in the second repeat only (Fig. 1C). In all MBT containing proteins, however, at least one of the repeats has aromatic residues in all these positions. The highly conserved aspartate residues that have been found to be important for SCM function in vivo are also located close to this site (Fig. 3). This would suggest that this binding site is important in the MBT domain as well.

Although genetic results implicate the MBT repeats of SCM in the in vivo function of the protein, little is known about their biochemical role. Proteins containing mutations within the MBT repeat region are associated with their normal sites on polytene chromosomes. This suggests that the repeats are not involved in SCM localization. It is more likely that they provide a biochemical activity required for transcriptional silencing. Based on the structure of the repeats, it is unlikely that they have a catalytic role. Given their similarity to the Tudor and chromo domains, a binding role is more likely, in particular because the mutation of a residue in a putative binding site compromises the in vivo function. Post-translationally modified histones seem to be a probable binding target. The PRC1 complex is recruited to chromatin in which histone H3 has been methylated at lysine 27. This modification is thought to be recognized by the chromo domain of Polycomb. In Drosophila, Polycomb is involved in generating heterochromatin and re-
pressing homeotic genes. It is has been shown that the Polycomb chromo domain is essential for the binding of the PC protein to chromatin and is probably involved in compacting the chromosomal regions present in heterochromatin or heterochromatin-like complexes (34). Arginine residues in histones are also known to be methylated. For example, symmetric dimethylation of histone tails by the protein arginine methyltransferase PRMT5 is known to be associated with transcriptional repression (35). No protein or domain responsible for the recognition of this modification has been reported to date. Because the MBT repeat is similar in structure to the Tudor domain implicated in the binding of methylated arginine peptides, it is possible that the MBT domains of SCM provide a link between arginine methylation of histones and the PRC1 complex (36). Although it would appear that the binding function of SCM is not required for localization of the PRC1 complex, it may, under some circumstances, have a role in repression. The MBT domain is always found in a tandem of at least two repeats. It is possible that this arrangement is required to enable the protein to interact with multiple sites on chromatin that help to lock it into an inactive state. This scenario would fit with a model in which PRC1 function depends on chromatin modifications carried out by other complexes. Biochemical data suggest that SCM is not always part of the PRC1 complex. It is possible that arginine methylation and SCM are required for the repression of some targets in some tissues. Mammals have several SCM-like proteins, and it is likely that the regulation of PRC1 function is even more complex.

The data presented here strongly suggest that MBT repeats have a binding function, probably to modified histones. Identification of the binding partners is likely to provide important insights into the regulation of transcriptional repression in cell memory.

REFERENCES
1. Simon, J. (1995) Curr. Opin. Cell Biol. 7, 376–385
2. Pirrotta, V. (1998) Cell 93, 333–336
3. Satijn, D. P., and Otte, O. P. (1999) Mol. Cell. Biol. 19, 57–68
4. Shao, Z., Raible, F., Mollasghababa, R., Guyon, J. R., Wu, C. T., Bender, W., and Kingston, R. E. (1999) Cell 98, 37–46
5. King, I. F. G., Francis, N. J., and Kingston, R. E. (2002) Mol. Cell. Biol. 22, 7919–7928
6. Francis, N. J., Saurin, A. J., Shao, Z., and Kingston, R. E. (2001) Mol. Cell 8, 554–556
7. Sewalt, R. G., van der Vlag, J., Gunster, M. J., and O’Connor, M. (1993) Mol. Cell. Biol. 13, 3586–3595
8. van Lohuizen, M., Tijmes, M., Venecken, J. W., Schumacher, A., Magnuson, T., and Wientjens, E. (1998) Mol. Cell. Biol. 18, 3572–3579
9. Franke, A., DeCamillis, M. A., Zink, D., Cheng, N., Brock, H. W., and Paro, R. (1992) EMBO J. 11, 2841–2850
10. Jones, C. A., Ng, J., Peterson, A. J., Morgan, K., Simon, J., and Jones, R. S. (1998) Mol. Cell. Biol. 18, 7815–7820
11. Jones, R. S., and Gelbart, W. M. (1990) Genetics 126, 185–199
12. Bornemann, D., Miller, E., and Simon, J. (1996) Development 122, 1621–1630
13. Koga, H., Matsu, S., Hirota, T., Takebayashi, S., Okumura, K., and Saya, H. (1999) Oncogene 18, 3799–3809
14. Peterson, A. J., Kyba, M., Bornemann, D., Morgan, K., Brock, H. W., and Simon, J. (1997) Mol. Cell. Biol. 17, 6683–6692
15. Bornemann, D., Miller, E., and Simon, J. (1998) Genetics 150, 675–686
16. Bocconi, P., MacGregor, D., Scandura, J. M., and Nimer, S. D. (2003) J. Biol. Chem. 278, 15412–15420
17. Leslie, A. (1993) Recent Changes to the MOSFLM Package for Processing Film and Image Plate Data, SERC Laboratory, Daresbury, Warrington, United Kingdom
18. Kahch, K. (1988) J. Appl. Crystallogr. 21, 916–924
19. Project, C. C. (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
20. Terwilliger, T. C., and Berendzen, J. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 849–861
21. Terwilliger, T. C. (2000) Acta Crystallogr. Sect. D Biol. Crystallogr. 56, 965–972
22. Turk, D. (1992) Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektronende-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklearungen. Ph.D. thesis, Technische Universität München, Munich, Germany
23. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J-S, Kuzevski, J., Nilge, M., Pannsu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
24. Liu, Y., and Eisenberg, D. (2000) Protein Sci. 11, 1255–1263
25. Rousseau, F., Schymkowitz, J. W., and Itzhaki, L. S. (2003) Structure (Camb.) 11, 243–251
26. Wuyts, W. R., Tereshko, V., Bocconi, P., MacGregor, D., Nimer, S. D., and Patel, D. J. (2003) Structure (Camb.) 11, 775–789
27. Holm, L., and Sander, C. (1996) Nucleic Acids Res. 26, 316–319
28. Selenko, P., Sprangres, R., Stier, G., Behler, D., Fischer, U., and Sattler, M. (2001) Nat. Struct. Biol. 8, 27–31
29. Qu, C., Sawada, K., Zhang, X., and Cheng, X. (2002) Nat. Struct. Biol. 9, 217–224
30. Slater, L. M., Allen, M. D., and Bycroft, M. (2003) J. Mol. Biol.
31. Maurer-Stroh, S., Dickens, J., Huges-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C. P. (2003) Trends Biochem. Sci. 28, 69–74
32. Nielsen, P., Nielssch, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2002) Nature 416, 103–107
33. Jacobs, S. A., and Khorasanizadeh, S. (2002) Science 295, 2080–2083
34. Messmer, S., Frankel, A., and Paro, R. (1992) Genes Dev. 6, 1241–1254
35. Branscombe, T. L., Frankel, A., Lee, J. H., Cook, J. R., Yang, Z., Pestka, S., and Clarke, S. (2001) J. Biol. Chem. 276, 32971–32976
36. Sprangres, R., Groves, M. R., Srinivas, S., and Sattler, M. (2003) J. Mol. Biol. 327, 507–520
37. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950

Downloaded from http://www.ficat.org by guest on July 24, 2018
