Correspondence

Molecular characterization of *h-l(3)mbt-like*: a new member of the human mbt family

Jasmine Wismar*

First published online 5 October 2001

Several genes from human, mouse and other species have been found to contain so-called mbt-repeats, a protein domain of approximately 100 amino acids [1]. In *Drosophila melanogaster* three genes are known to contain mbt-repeats. The first one is the Polycomb group gene *Sex comb on midleg* (*Scm*) which exhibits two mbt-repeats [2], the second is the tumor suppressor gene *lethal(3)malignant brain tumor* (*l(3)mbt*) which contains three mbt-repeats [1] and the third gene with so far unknown function was recently identified and localized to the region 2L34A8-9 [3]. Its protein structure is similar to *l(3)mbt* in respect of the known domains but it contains four mbt-repeats. The function of the mbt-repeats is not yet known. Nevertheless these repeats are highly conserved between the *Drosophila* genes and their human homologs pointing to an important function. The human mbt family consists so far of three published genes containing mbt-repeats. H-l(3)mbt with three mbt-repeats was the first reported homolog of *l(3)mbt* [4]. Two genes with only two mbt-repeats, SCMH1 [5] and SCML2 [6], seem to be the human homologs of *Scm*. The third homolog of *Scm*, SCML1 [7] does not contain any mbt-repeat.

To identify human homologs of the *l(3)mbt* gene or at least novel members of the human mbt family of genes we used a computer-based strategy. The mbt-repeats of *l(3)mbt* were used for screening the human EST database. Several EST clones showing striking homologies have been identified. The sequence of these ESTs was further used to search the human EST and the NR division of the GenBank database. The EST clones have been sequenced using the vector primers first and thereafter gene-specific internal primers ordered from Eurogentec and MWG Biotech.

Some of the identified EST clones (H23073, R17498, R61020 and BE265013) correspond to the already published *h-l(3)mbt* gene [4]. Besides these EST clones we identified a number of EST clones belonging to a second gene (accession numbers. AW452738, T99539, R06448, H22704, BE894488, AW264224, AW408651, AA984407, etc.), which we call *h-l(3)mbt-like*. Most of these EST clones which were obtained from RZPD Berlin and Research Genetics Huntsville have been completely sequenced. They represent two different splice variants (Fig. 1A). The smaller transcript of 3194 bp defines an ORF encoding a protein of 706 aa (H-l(3)mbt-like a) whereas the larger transcript of 3299 bp encodes a protein of only 615 aa (H-l(3)mbt-like b) due to a 105 bp insertion (exon 15) that leads to a frame shift and subsequently to an earlier stop codon. Interestingly the potential nuclear localization signal present C-terminal of the mbt-repeats in H-l(3)mbt-like a at position 620 to 632 is missing in H-l(3)mbt-like b. Thus the two proteins might function in different compartments. H-l(3)mbt a in the nucleus and H-l(3)mbt b in the cytosol. Both proteins contain the C2–C2 zinc finger and four mbt-repeats but lack the SAM domain [8] present in H-l(3)mbt (Fig. 1B,C). This domain which most probably mediates protein–protein interactions is also called SPM domain [2]. Interestingly *h-l(3)mbt-like* is the only one of the human mbt family of proteins known so far to contain the C2–C2 zinc finger N-terminal of the mbt-repeats (Fig. 1C) like it is the case for *l(3)mbt* [1]. As both putative H-l(3)mbt-like proteins, the potentially cytosolic a and the nuclear b isoform contain the C2–C2 zinc finger, this motif seems likely to function in protein–protein interaction rather than in direct DNA binding.

Comparing the *h-l(3)mbt-like* cDNA sequence with the genomic sequence given by clones HS85F18 and HS756G23 (22q13.3) revealed a gene consisting of 18 exons (Fig. 1A). The exon 15 present only in transcript b is found just in a few EST clones obtained from fetal liver and spleen. Tissue expression analysis using a human Poly(A)$^+$ RNA multiple tissue Northern blot membrane (Clontech) hybridized with the ^{32}P-labeled 1.7 kb *Not*I–*Hin*dIII *h-l(3)mbt-like* cDNA fragment contained in EST clone H22704 showed that the gene is expressed in all tissues analyzed (data not shown). We identified a prominent 3.6 kb transcript ubiquitiously present in all human adult tissues in varying amounts. This transcript most probably corresponds to the splice variant *h-l(3)mbt-like a*. The transcript level was found to be highest in the brain and pancreas and is also abundant in the heart, placenta, liver, skeleton muscle and kidney while it is barely expressed in the lung. Besides the 3.6 kb transcript an additional rarely expressed 3.8 kb transcript can be detected which most likely represents the splice variant *h-l(3)mbt-like b*.

The screen for mbt-repeat containing ESTs revealed many *h-l(3)mbt* ESTs including some representing a so far unpublished alternatively spliced transcript. Thus the *h-l(3)mbt* gene structure was analyzed in detail using the identified EST clones as well as the available sequences of the DKFZP586P1522 (gi:7661701) and KIAA0681 [9] proteins. Comparing the cDNA sequence with the genomic sequence represented by clones HS862K6 and HS138B7 it was found that the gene consists of 19 exons (Fig. 1B). Three splice variants containing mbt-repeats have been identified. Two with incomplete SAM domain have been already published [4] but the source of the first 139 bp included in the published sequence was found to be from chromosome 15 instead of region 20q13.12 where the gene is localized. Thus this part of the sequence is most probably the result of a cDNA cloning artifact. The third splice variant is represented by the KIAA0681 sequence. The alternative usage of exon 19b instead of 19a leads to a complete SAM domain (Fig. 1C).

In addition to the EST clones belonging to *h-l(3)mbt*, *h-l(3)mbt-like* and the two already known Scm homologs SCMH1 and SCML2 we found through the intensive database searches a number of EST clones representing three further members of the human mbt family each containing four mbt-repeats, the KIAA1617 [10], FLJ20055 (gi:7019904) and RU1 (gi:6635352) genes. The genes could be localized to chro-

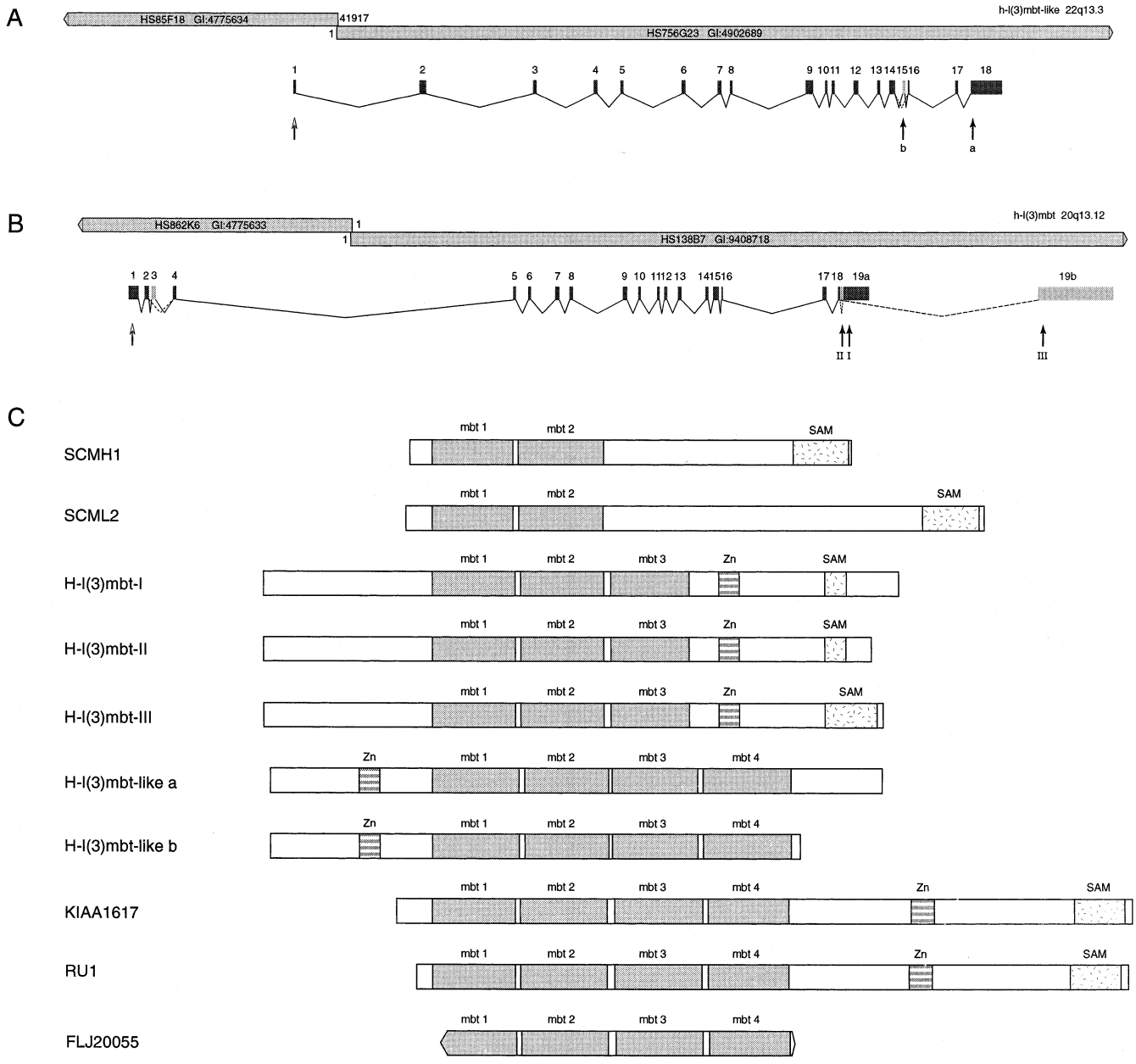

Fig. 1. A: Exon–intron structure of *h-l(3)mbt-like*. 18 different exons have been identified. The two alternatively used stop codons (a, b) are indicated by black arrows. The gray exons and the dotted lines represent alternative splicing. The start codon is indicated with an open arrow. The gray bars represent the corresponding genomic clones. B: Exon–intron structure of *h-l(3)mbt*. Usage of exon 19b instead of 19a leads to a complete SAM domain (compare with C). Three alternatively used stop codons (I, II, III) are indicated with black arrows. C: Protein structure of the human mbt family. Three domains are shown. The mbt-repeat [mbt], the SAM domain [SAM] and the C2–C2 zinc-finger (Zn). For the *h-l(3)mbt* gene three isoforms and for the *h-l(3)mbt-like* gene two isoforms are presented. The amino acid sequence known for FLJ20055 seems to be incomplete as indicated by the arrow-shaped boxes.

mosomes 10, 17 and 3 respectively using the BLASTN algorithm to screen the human genome and the nt databases. Comparing the genomic sequence with the cDNA and EST sequences it was found that both of the KIAA1617 and the RU1 genes consist of 21 exons while the given FLJ20055 sequence is composed of 11 exons. Besides the four mbt-repeats both of the KIAA1617 and the RU1 proteins contain a C2–C2 zinc finger and a C-terminal SAM domain while the FLJ20055 protein encodes only three and a half mbt-repeats which indicates that the sequence might be incomplete in the 5′ direction. A summary of the identified members of the human mbt family is presented in Fig. 1C.

Recently identified human EST clones (BF797869, BF377516, BF376727 and BF755043) showing strong homologies to the *l(3)mbt* mbt-repeats point to at least two more members of the mbt family which will be analyzed in the near future.

Sequence data from this article have been deposited with the EMBL/GenBank data Libraries under accession numbers. AJ305226 and AJ305227.

Acknowledgements: This work was supported by Grant Wi1540 from the Deutsche Forschungsgemeinschaft.

References

[1] Wismar, J., Löffler, T., Habtemichael, N., Vef, O., Geissen, M., Zirwes, R., Altmeyer, W., Sass, H. and Gateff, E. (1995) Mech. Dev. 53, 141–154.
[2] Bornemann, D., Miller, E. and Simon, J. (1996) Development 122, 1621–1630.
[3] Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wie, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C. (2000) Science 287, 2185–2195.
[4] Koga, H., Matsui, S., Hirota, T., Takebayashi, S., Okumura, K. and Saya, H. (1999) Oncogene 18, 3799–3809.
[5] Berger, J., Kurahashi, H., Takihara, Y., Shimada, K., Brock, H.W. and Randazzo, F. (1999) Gene 237, 185–191.
[6] Montini, E., Buchner, G., Spalluto, C., Andolfi, G., Caruso, A., den Dunnen, J.T., Trump, D., Rocchi, M., Ballabio, A. and Franco, B. (1999) Genomics 58, 65–72.
[7] Van de Vosse, E., Walpole, S.M., Nicolaou, A., van der Bent, P., Cahn, A., Vaudin, M., Ross, M.T., Durham, J., Pavitt, R., Wilkinson, J., Grafham, D., Bergen, A.A., van Ommen, G.J., Yates, J.R., den Dunnen, J.T. and Trump, D. (1998) Genomics 49, 96–102.
[8] Ponting, C.P. (1995) Protein Sci. 4, 1928–1930.
[9] Ishikawa, K., Nagase, T., Suyama, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. (1998) DNA Res. 5, 169–176.
[10] Nagase, T., Kikuno, R., Nakayama, M., Hirosawa, M. and Ohara, O. (2000) DNA Res. 7, 273–281.

*Fax: (49)-6131-3925845.
E-mail: wismar@mail.uni-mainz.de

Institute of Genetics, Becherweg 32, 55099 Mainz, Germany

PII: S0014-5793(01)02959-3