Evolutionary divergence in the broad complex, tramtrack and bric à brac/poxviruses and zinc finger domain from the candidate tumor suppressor gene hypermethylated in cancer

Cateline Guerardel, Sophie Deltour, Dominique Leprince*

CNRS UMR 8526, Institut de Biologie de Lille, Institut Pasteur de Lille, 1 Rue Calmette, 59021 Lille Cedex, France

Received 7 April 1999; received in revised form 16 April 1999

Abstract Hypermethylated in cancer, a new candidate tumor suppressor gene located in 17p13.3, encodes a protein with five Krüppel-like C_2H_2 zinc finger motifs and a N-terminal protein/protein interaction domain called broad complex, tramtrack and bric à brac/poxviruses and zinc finger domain. Hypermethylated in cancer appears unique in the broad complex, tramtrack and bric à brac/poxviruses and zinc finger family since it contains a 13 amino acid insertion located in a loop between the conserved β-strand β5 and helix α5 which are involved in dimerization and scaffolding of the broad complex, tramtrack and bric à brac/poxviruses and zinc finger domain. Cloning and sequencing of a murine hypermethylated in cancer gene suggests that this insertion has been acquired late in the evolution since it is present in two mammalian hypermethylated in cancer genes but absent in its zebrafish and avian counterparts. This is a unique example of a high divergence of the same broad complex, tramtrack and bric à brac/poxviruses and zinc finger domain in different species.

Key words: Hypermethylated in cancer; γF1-binding protein isoform B; Broad complex; Tramtrack; bric à brac/poxviruses; Zinc finger domain; Zinc finger; Tumor suppressor gene; Chromosomal band 17p13.3

1. Introduction

The broad complex, tramtrack and bric à brac (BTB) or poxviruses and zinc finger (POZ) domain is a ca. 120 amino acid protein-protein interaction domain conserved from *Dictyostelium* to human and is found in developmentally regulated transcription factors and actin-binding proteins [1–3]. In most cases, the BTB/POZ domain is associated with C_2H_2 zinc finger motifs in proteins involved in transcriptional regulation through chromatin modelling. For example, the *Drosophila* GAGA factor, also known as trithorax-like, can counteract histone H1-mediated transcriptional repression, is a member of the trithorax group genes which maintain the transcriptional status of the initially activated homeotic selector genes and finally is an enhancer of position effect variegation (PEV) [2,4]. However, many BTB/POZ and zinc finger proteins are transcriptional repressors such as *Drosophila* tramtrack and vertebrates ZF5, γF1-binding protein isoform B (γFBP-B), PLZF and BCL6. In these two latter cases, the BTB/POZ domain has been shown to mediate homodimerization and to recruit a repressing complex containing SMRT/N-CoR, mSin3A/B and HDAC-1, an histone deacetylase [5,6]. Strong structural support for these observations has been recently brought by the crystal structure of the BTB/POZ domain from PLZF solved at a 1.9 ÅA resolution [3]. This structure has revealed a tightly intertwined dimer with an extensive hydrophobic region. A surface-exposed groove lined with conserved residues is formed at the dimer interface and has been proposed to represent the site of interaction with nuclear co-repressors and/or other nuclear proteins [3].

A candidate tumor suppressor gene, hypermethylated in cancer (HIC-1) was recently identified because of its association with a 'CpG island' at 17p13.3 that is aberrantly hypermethylated and transcriptionally inactivated in several common types of human cancers [7,8]. HIC-1 encodes a typical nuclear BTB/POZ protein with five Krüppel-like C_2H_2 zinc finger motifs in its C-terminal part and a BTB/POZ domain at its N-terminus [7]. However, the HIC-1 BTB/POZ domain appears unique among all the BTB/POZ domains known so far since it contains a 13 amino acids insertion, rich in alanine (eight out of 13 residues) [7]. Based on a sequence alignment of BTB/POZ proteins and on the PLZF structure, this insertion is located in a loop, possibly exposed to solvent, between β-strand β5 and helix α5 which are involved in scaffolding and dimerization of the domain [3] (Fig. 1). Strikingly, this specific insertion is only very partially conserved in its avian homologue, γFBP-B which has been isolated as a sequence-specific transcriptional repressor of the γF-crystallin gene during embryonic development [9] (Figs. 1 and 2). Such a sequence divergence is quite unusual as exemplified by ZF5 where the murine and avian BTB/POZ domains only differ by a conservative mutation (leucine/valine) [10]. In this report, we have cloned the murine HIC-1 gene and shown that the specific insertion is present in these two mammalian BTB/POZ domains whereas it is absent in its zebrafish and avian homologues.

2. Materials and methods

2.1. Cloning of a murine HIC-1 gene

Genomic DNA prepared from the liver of a male mouse (strain 129 SV) was partially digested with *Sau*3A and cloned into the *Bam*HI-digested λDASHII vector. This library was screened at a high stringency with a human HIC-1 BTB/POZ probe obtained by PCR using a previously described human recombinant phage as a template [11]. One candidate recombinant (among the two obtained) was purified and used for the studies described here. In particular, a 3.5 kbp *Not*I fragment shown by Southern analyses to contain the entire BTB/POZ domain was subcloned into the pSK plasmid (Stratagene) and subjected to nucleotide sequencing.

*Corresponding author. Fax: (33) (3) 20 87 1019.
E-mail: dleprinc@infobiogen.fr

Abbreviations: HIC-1, hypermethylated in cancer; γFBP-B, γF1-binding protein isoform B; BTB, broad complex, tramtrack and bric à brac domain; POZ, poxviruses and zinc finger domain

0014-5793/99/$20.00
PII: S0014-5793(99)00583-9

2.2. Nucleotide sequencing

Searches in the expressed sequence tag (EST) database identified a partial murine sequence corresponding to 151 nucleotides in the 5′ part of the BTB/POZ domain (GenBank accession number AA103371). This sequence was used to design oligonucleotides to start the sequencing project. DNA sequencing on both strands of the complete BTB/POZ domain was carried out by using an ABI PRISM 377 automated DNA sequencer (Perkin Elmer Applied Biosystems).

3. Results

3.1. Nucleotide sequence of the murine HIC-1 BTB/POZ domain

To address in more detail the phylogenetic conservation of the HIC-1-specific insertion absent in its avian homologue γFBP-B and located between the β5-strand and the α5 helix conserved among BTB/POZ domains and shown in PLZF to be involved in dimerization, we have cloned and sequenced the BTB/POZ domain of a murine HIC-1 gene. Restriction mapping and Southern blotting analyses of a recombinant phage isolated from a male 129 genomic library [11] showed that the murine HIC-1 encoding sequence is included in a ca. 3.5 kbp *Not*I junction fragment in sharp contrast with the clustering of five *Not*I sites in the human gene [7]. Complete nucleotide sequencing of the murine HIC-1 BTB/POZ domain and comparison with its human homologue revealed that the C-terminal part of this domain is less well-conserved than its N-terminal part. Among the 34 nucleotide changes observed between the human and murine sequences, 26 are clustered in the C-terminal half of the BTB/POZ domain (data not shown). A striking example is the region corresponding to the HIC-1 insertion where 10 mutations (out of 39 nucleotides) are found. Most of them, however, affect the third nucleotide of the codon, indicating a strong selection pressure at the amino acid level. In fact, the murine and human BTB/POZ domains differ only by three amino acid changes localized in the 5′ part of the specific insertion (Fig. 2). The partial murine HIC-1 EST sequence (GenBank accession number AA103371) corresponds in fact to amino acids 2–51 of our sequence. In the database, we have also found a partial zebrafish HIC-1 sequence (GenBank accession number AA497316) which is highly homologous to the mammal and avian BTB/POZ domains. However, the specific insertion found in the two mammalian HIC-1 BTB/POZ domains is also very poorly conserved in zebrafish (Fig. 2).

3.2. Definition of the boundaries of the murine HIC-1 BTB/POZ-containing exon

By sequence alignment of various BTB/POZ proteins, the 3′ limit of this domain in HIC-1 and γFBP-B would be the motif CHLR [1–3] (Fig. 2). A second region just 3′ of this motif also deserves special attention since it has been proposed to correspond to an exon boundary [7]. Indeed, Makos-Wales et al. reported that the human HIC-1 BTB/POZ domain is included in a single encoding exon fused by splicing to a second encoding exon and preceded by a unique non-coding exon [7]. The 3′ boundary of this first encoding exon is the motif CHLRGGGGGGGGYAPY whereas in the chicken γFBP-B cDNA, the motif CHLRGGYAPY is found [7,9] (Figs. 2 and 3). These eight glycine residues are all encoded by a GGC codon yielding a 24 bp sequence reminiscent of the RRY(26+) cryptic repeats which are enriched in encoding regions of DNA-binding proteins/transcription factors [12]. Cryptic RRY(i) repeats may be a candidate for triplet repeat genetic instability and, when mutated in somatic cells, may contribute to carcinogenesis [12]. Sequencing of the murine HIC-1 BTB/POZ domain revealed a very similar CHLRGGGSGGGGYAPY motif, with the fourth GGC codon mutated in another RRY type codon, **A**GC.

Strikingly, we noticed that the proposed human splice donor site at the end of the first encoding exon [7] is not functional in mouse. Indeed, the canonical dinucleotide gt found in the human sequence and thus proposed to be the splice donor site of this exon is mutated into GG in the mouse (Fig. 3, bottom). These two codons encode a glycine residue. Furthermore, the open reading frame of the human and murine BTB/POZ domains extends within these 'intronic' sequences and are homologous (data not shown). Finally, the partial nucleotide sequence of a chicken γFBP genomic clone that we have isolated is strictly co-linear in this region with the published chicken cDNA sequence [9], arguing against the presence of an intron (data not shown). By contrast, we observed that the proposed splice acceptor site in the human sequence located just upstream of the ATG initiation codon in the BTB/POZ domain is strictly conserved in mouse [7] (Fig. 3,

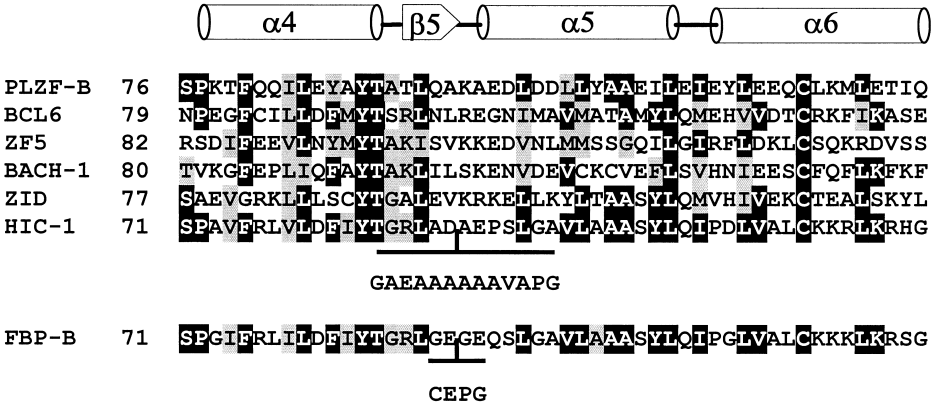

Fig. 1. Secondary structure of the C-terminal part from the PLZF BTB/POZ domain and sequence alignment of selected BTB/POZ domains. With the exception of chicken γFBP-B, the proteins are of human origin. Identical and similar residues are respectively printed in reverse type and shaded. The sequences aligned correspond roughly to the C-terminal end of the BTB/POZ domain and amino acid numbering is according to the original publications. The partial PLZF structure is derived from [3] where its complete structure and a sequence alignment of the complete BTB/POZ domains can be found.

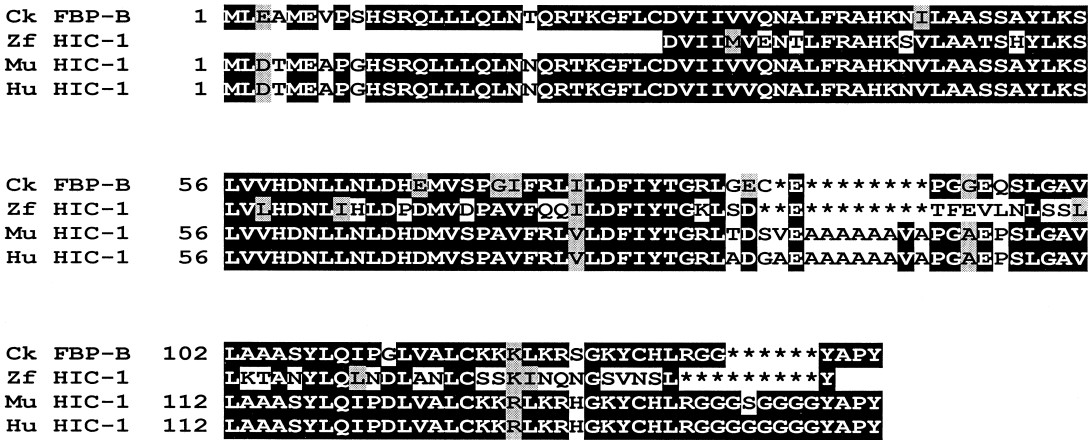

Fig. 2. Comparison of the amino acid sequences of human HIC-1, murine HIC-1, zebrafish HIC-1 and chicken γFBP-B BTB/POZ domains. The four sequences starting with the first methionine residue in the HIC-1 and γFBP-B BTB/POZ domains are aligned. In the mouse sequence, this methionine is preceded by another one (see Fig. 3). The beginning of the zebrafish sequence is not known. Identical and similar residues in at least 50% of the sequences are respectively printed in reverse type and shaded. Gaps (*) have been introduced to maximize the alignment. Amino acid numbering is according to the original publications [7,9]. The murine sequence has been deposited in the EMBL/GenBank database (accession number MMU132691).

top). The human HIC-1 genomic structure and nucleotide sequence have been recently corrected in GenBank (Carter and Baylin, HUMHIC1G, L41919, direct submission). The human HIC-1 is actually a 714 amino acid protein encoded by a unique encoding exon preceded by a first untranslated exon in good agreement with our results.

4. Discussion

HIC-1 and γFBP-B are two BTB/POZ zinc finger proteins of respectively 714 and 641 amino acids. Except for the 99% homology between residues 490–594 of HIC-1 and 441–545 of γFBP-B which correspond to the four canonical Krüppel-like zinc fingers likely to be involved in sequence-specific DNA-binding, the overall homology between the two proteins is relatively low (Fig. 4). In contrast, the murine and the chicken ZF5 proteins are more than 95% homologous [10]. Notably, the other identified functional domain, the BTB/POZ domain, is only 80% conserved between HIC-1 and its avian homologue which is again unusual in this protein family. This is mainly due to encoding sequence acquisition which occurs lately during the evolution of the HIC-1 gene since it seems to be restricted to the mammalian HIC-1 genes as shown here. Although several nucleotides changes are found in this region between the murine and human sequences, this insertion and specially the alanine stretch is subject to a strong selection pressure at the amino acid level, arguing for an important functional role. What would be the advantage to acquire in a functional domain a new sequence? This specific insertion acquired late in the evolution of the HIC-1 BTB/POZ domain could modify the two main properties ascribed to this domain in other proteins, namely the dimerization and transcriptional repression. From its location between the β5 strand and the α5 helix conserved in the BTB/POZ domain and shown in PLZF to be involved in dimerization, we could speculate that this insertion may somehow affect the dimerization potential of the HIC-1 BTB/POZ domain. Extrapolating again from the PLZF structure, this insertion could also be exposed at the external part of the dimer. In addition, the stretch of eight glycine residues found just after the human HIC-1 BTB/POZ domain is almost perfectly conserved in the murine gene (seven glycine and a serine residue) whereas only two glycines are found in the chicken γFBP-B domain. These specific features could create a hinge between the BTB/POZ domain and

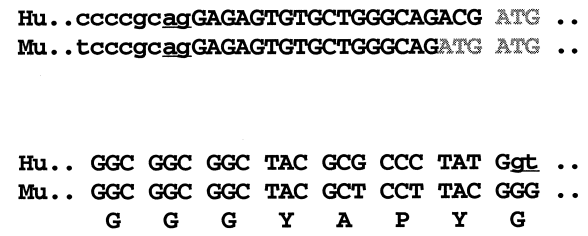

Fig. 3. A proposed splice donor site in the human HIC-1 gene is mutated in the mouse. The human [7] and murine (this study) sequences encompassing the boundaries of the BTB/POZ-containing exon are aligned. The canonical dinucleotides **ag** (splice acceptor) and **gt** (splice donor site) are underlined. The human and murine ATG initiation codon are printed in gray. We noticed the presence of a second upstream ATG in the murine sequence due to a single point mutation when compared to its human homologue. In the 3′ part of the sequence, the corresponding amino acid is indicated below each codon using the single letter code.

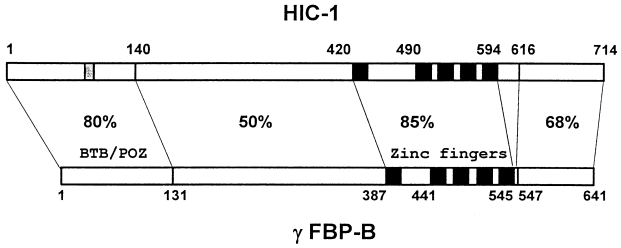

Fig. 4. Schematic representation of the HIC-1 and γFBP-B proteins. The structure of the HIC-1 protein is derived from a recent correction in GenBank (HUMHIC1G) (Carter and Baylin, direct submission) of reference [7] whereas γFBP-B is derived from the original publication [9]. The five Krüppel-type zinc finger motifs are represented as black boxes whereas the gray box refers to the 13 amino acids specific insertion in the HIC-1 BTB/POZ domain [3,7]. Degrees of conservation (% of identity and similarity) are indicated below protein representations. The amino acid positions of the homologous regions are shown.

the remainder of the molecule in the mammalian HIC-1 proteins, thus possibly creating and exposing, in addition to the central groove, an interface for another set of interactions with nuclear proteins. Clearly, the crystal structure from the HIC-1 BTB/POZ domain and its comparison with PLZF would be highly informative.

Despite this high degree of evolutionary divergence, the BTB/POZ domains of HIC-1 and γFBP-B appear as similar autonomous transcriptional repression domains (S. Deltour et al., unpublished results). Functional studies are in progress to determine the effect of these sequence particularities on the known properties of the BTB/POZ domains in transcriptional repressors, their homodimerization and their interaction with the nuclear co-repressors.

Acknowledgements: We are indebted to Prof. D. Stehelin for constant interest and support. We thank Agnes Begue for the gift of the male 129 genomic library and Dr B. van den Bunder for critical reading of the manuscript. This work was supported by the Centre National de la Recherche Scientifique, the Institut Pasteur de Lille and the Association pour la Recherche contre le Cancer. S. Deltour holds a fellowship from the Ligue Nationale contre le Cancer.

References

[1] Bardwell, V.J. and Treisman, R. (1994) Genes Dev. 8, 1664–1677.
[2] Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G. and Leprince, D. (1995) Cell Growth Differ. 6, 1193–11981.
[3] Ahmad, K.F., Engel, C.K. and Prive, G.G. (1998) Proc. Natl. Acad. Sci. USA 95, 12123–12128.
[4] Wilkins, R.C. and Lis, T.J. (1997) Nucleic Acids Res. 25, 3963–3968.
[5] Dhordain, P., Albagli, O., Lin, R.J., Ansieau, S., Quief, S., Leutz, A., Kerckaert, J.P., Evans, R.M. and Leprince, D. (1997) Proc. Natl. Acad. Sci. USA 94, 10762–10767.
[6] Lin, R.J., Nagy, L., Inoue, S., Shao, W., Miller Jr., W.H. and Evans, R.M. (1998) Nature 391, 811–8148.
[7] Makos-Wales, M., Biel, M., El Deiry, W., Nelkin, B.D., Issa, J.P., Cavenee, W.K., Kuerbitz, S.J. and Baylin, S.B. (1995) Nat. Med. 1, 570–577.
[8] Fujii, H., Biel, M.A., Zhou, W., Weitzman, S.A., Baylin, S.B. and Gabrielson, E. (1998) Oncogene 16, 2159–2164.
[9] Liu, Q., Shalaby, F., Puri, M.C., Tang, S. and Breitman, M.L. (1994) Dev. Biol. 165, 165–177.
[10] Bhathal, H.S. and Stumph, W.E. (1996) Biochem. Biophys. Acta. 1308, 114.
[11] Deltour, S., Guerardel, C., Stehelin, D. and Leprince, D. (1998) Biochem. Biophys. Acta 1443, 230–232.
[12] Ricke, D.O., Liu, Q., Gostout, B. and Sommer, S.S. (1995) Genomics 26, 510–520.