DDT – a novel domain in different transcription and chromosome remodeling factors

Tobias Doerks, Richard Copley and Peer Bork

Homology-based sequence analyses have revealed the presence of a novel domain (DDT) in bromodomain PHD finger transcription factors (BPTFs), chromatin remodeling factors of the BAZ-family and other putative nuclear proteins. This domain is characterized by a number of conserved aromatic and charged residues and is predicted to consist of three α helices. Recent studies indicate a likely DNA-binding function for the DDT domain.

Chromatin remodeling complexes have been implicated in the disruption or reformation of nucleosomal arrays resulting in modulation of transcription[1]. Associated with these complexes are several bromodomain-containing proteins; the bromodomain is involved in protein–protein interactions and other regulation processes[2,3].

The characterized domains of BPTF, a bromodomain PHD finger (C4HC3 zinc-finger homology) transcription factor[4], are thought to perform protein-binding functions[2,5]. We analyzed different regions of this protein in order to identify possible DNA-binding domains. PSI-BLAST (Ref. 6) searches against NRDB (non-redundant database) with the N-terminal region (conserved residues 102–162, Fig. 1) of BPTF reveal significant similarity to proteins with N-terminal AT-Hooks[8] in *Caenorhabditis elegans* ($E = 7 \times 10^{-11}$) and *Drosophila melanogaster* ($E = 2 \times 10^{-15}$) and nearly 100% identity to a C-terminal truncated human transcription factor (FALZ or FAC1, 810 amino acids in length), which is believed to play a role in Alzheimer's disease[9].

A second round of PSI-BLAST searching revealed homology to a four-PHD-domain[5]-containing protein of unknown function, MOI20 ($E = 10^{-3}$). After further iterations the conserved region was also retrieved in homeobox[10]-containing, putative DNA-binding proteins ($E = 10^{-4}$) and in hypothetical proteins of *Arabidopsis thaliana* ($E = 4 \times 10^{-6}$) as well as (with less significance, $E \approx 0.04$) in hypothetical yeast proteins. Additional Hidden Markov Model searches[11] indicate similarity just above the default threshold to members of the BAZ-family of chromatin remodeling factors ($E = 0.15$); BAZ (bromodomain adjacent to zinc finger proteins)-family members[12] are proteins with a similar domain composition to BPTF-transcription factors. These proteins are implicated in Williams Syndrome, a complex developmental disorder with multisystemic defects[13–15]. The significance of these latter homologies is confirmed by MACAW alignment analysis[16] (*P*-values between 10^{-12} and 10^{-50}).

We named the newly discovered region the DDT domain (after the better-characterized DNA-binding homeobox-containing proteins and the different transcription and chromatin remodeling factors in which it is found). This domain is exclusively associated with nuclear domains (see Fig. 2) and is generally ~60 amino acids in length. Multiple sequence alignment reveals several regions of particular conservation, including conserved charged residues, N-terminal phenylalanines and C-terminal leucines (Fig. 1). As previously reported[12], the putative protein–protein interaction LXXLL motif[18] is not conserved, even within the BAZ subfamily members, which makes a fundamental functional role unlikely.

We propose a DNA-binding function for the DDT domain. Several of the proteins we have identified are known to bind DNA, but lack a known DNA-binding domain [e.g. ATP-dependant chromatin assembly factor 1 (ACF1)[19] and BPTF (Ref. 12)]. Furthermore, recent experimental analysis of FAC1 (a truncated version of BPTF) has shown that the N-terminal region (1–398), which includes the DDT domain, is essential for DNA-binding[20]. A secondary-structure prediction using the PHD program[7] suggests that the domain consists of three α helices, and such α-helical composition is typical of many DNA-binding domains of known structure. However, fold-recognition procedures do not produce convincing evidence of homology to known DNA/RNA binding α-helical bundles (data not shown). The delination of the widespread DDT domain and its boundaries should allow directed structural studies. Our hypothesis of DNA-binding can be tested experimentally.

```
BPTF(FAC1) hs 102 NEHIMNVIAIYEVLRNFGTVLR-LSPFR------FEDFCAALVSQ-EQCTLMAEMHVVLLKAVLREEDT Q9UIG2
CG17135    dm 189 NTHVLRALSIYEVLRRFRHMVR-LSPFR------FEDLCAALACE-EQSALLTEVHIMLLKAILREEDA Q9W0T0
F26H11     ce 253 TASIMDAVEIYEILRSYHRTLR-ITPFT------FEDFCAALISH-NNSCIMAEVHMALLRNCLKSDDE O45409
BAZ1A      hs 573 PEIFGDALMVLEFLNAFGELFD-LQDEFPDG-VTLEVLEEALVGN-DSEGPLCELLFFFLTAIFQAIAE Q9UIG1
BAZ1B      hs 605 NTLFGDVAMVVEFLSCYSGLL--LPDAQYP--ITAVSLMEALSADKGGFLYLNRVLVILLQTLLQDEIA Q9UIG0
ACF1       dm 347 EHLLGDAFMMREFMHTYTGLLSGIEVFRQN--LSFYEMTRALTAR-EIAGPLSDILLVLLGTVFDLQKE Q9Y0W1
ZK783      ce 525 SQGFADALMVHEFVQNFGHVLG-IDLEIA---PKLESLCAGLDGDANHAEQTLQLTRQLLRLALEFPGM Q23590
H20J04     ce 473 NAEFEDYLFIFQFFNSFKQLLP-LKEIRGSDEVQFSDIIIAIKCNDPQNSSFADLLRVLLSIRTDIADE AAF39888
F3M18      at 658 DETVGNLLMVWRFLISFSDVLD-LWPFT------LDEFIQAFHDY-DSR-LLGEIHVTLLRSIIRDVED Q9SGP0
MLN1       at 515 DENVANLLMVWRFLITFADVLG-LWPFT------LDEFAQAFHDY-DPR-LMGEIHIVLLKTIIKDIEG BAB10985
MOI20      at 193 EEAVAHLLSVYGFLRSFSFQLY-ICPFE------LNDFVGALYFS-GPNSLLDAVHVALLRALKGHLER BAA98208
MAH20      at 298 MDCVGDLLMVWDFCTSFGRQLH-LWRFS------LEDFENAVCHKESNLVLIMEVHASLFRFLINERGD BAB10012
Ypl216     sc 376 QFPTERLLVVYQFLSFFGRFIG-LSHFN------FDQFLTTIKCT-SPEALVDEYVKINFLKTYNSKGS Q08964
F1N21      at 217 TEEAGNVCQLFEFCSAFGKALA-LKEGHAET-IVRELFICGRNTRRQQYCSITQMMIQLLDLISKDREM O49273
YGN3       sc 424 FDSFGKLLQAYQFLNTFGSKIC-LSHFS------LDQFITSLKCT-DPYELKGEVVLVNIRTQTSKEQE P53125
hypAT1     at 187 EEAVVYLLSVYGFLRSFSVQLY-ICPFG------LDDFVGALNFL-GPNSLLDAVHVALMRALKGHLER BAB11682
hypAT2     at 258 PEDAGNVFQFLEFCSAFGKALD-LRKGQAE--CVIREMLSGRSKRRQQYSTLTQMIIQLLTVILEDRGE BAB10159
Consensus   (80%) ...h.phh.lhpFhpsFuphL..lp.hp......hpph..ul.sp.p....h.plhhhLLphhhpp...
Sec.struc.pred.   ...hhHHHHHHHHHHHHHHh.............HHHHHHHH.......hHHHHHHHHHHHHHHh....
                                                                                  TiBS
```

Fig. 1. Multiple sequence alignment of DDT domains of BPTF (bromodomain, PHD finger transcription factors), transcription-factor-like proteins (BPTF, CG17135, F26H11), BAZ (bromodomain adjacent to zinc finger)-family proteins [BAZ1A, BAZ1B, ATP-dependant chromatin assembly factor 1 (ACF1), ZK783, H20J04], Hox-domain-containing proteins (F3M18, MLN1), PHD (C4HC3 zinc-finger homology)-domain-containing protein (MOI20) and other hypothetical proteins (MAH20, Yp1216, F1N21, YGN3, hyp AT1, hyp AT2). First column, protein names; second column, species names; third column is the start of the domain in each sequence; rightmost column, database accession numbers. Conserved charged residues are shown in red; conserved hydrophobic residues are shown in blue; other conserved residues are shown in purple. The consensus sequence (conserved in 80% of the sequences) is shown below the sequences; h, p, u, s and l indicate hydrophobic, polar, tiny, small, aliphatic and negatively charged residues, respectively. The predicted secondary structure (sec.struc.pred.) taken from the consensus of the alignment is shown (H, helix predicted with expected average accuracy >82%; h, helix predicted with expected average accuracy <82%)[7]. Abbreviations: at, *Arabidopsis thaliana*; ce, *Caenorhabditis elegans*; DDT, after DNA-binding homeobox-containing proteins and the different transcription and chromatin remodeling factors in which it is found; dm, *Drosophila melanogaster*; hs, *Homo sapiens*; sc, *Saccharomyces cerevisiae*.

 0968-0004/01/$ – see front matter PII: S0968-0004(00)01769-2

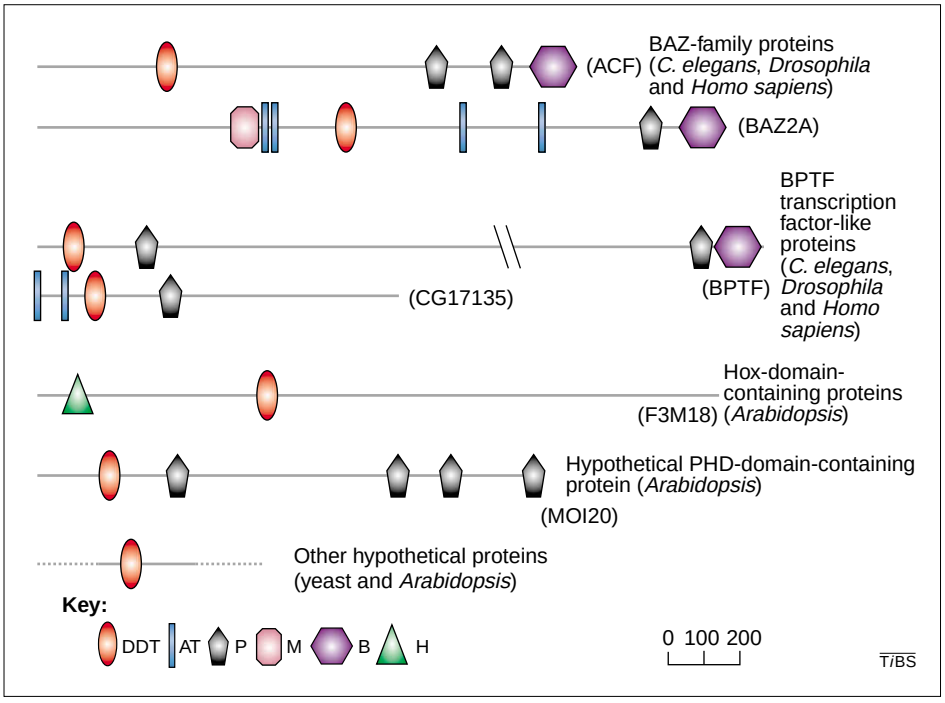

Fig. 2. Domain architecture of proteins that contain the DDT domain. Only proteins with distinct modular organizations are shown. The domain names are according to the Simple Modular Architecture Research Tool[17] (http://smart.embl-heidelberg.de). Abbreviations: AT, AT-Hook (a DNA-binding domain with preference for A/T regions); B, bromodomain (might be involved in transcriptional regulation and protein–protein interactions); DDT, after DNA-binding homeobox-containing proteins and the different transcription and chromatin remodeling factors in which it is found; H, DNA-binding homeodomain; M, methyl-CpG-binding domain; P, PHD C4HC3 zinc finger. The scale bar represents length of protein in amino acids.

References

1 Kornberg, R.D. and Lorch, Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryotic chromosome. *Cell* 98, 285–294

2 Barlev, N.A. *et al.* (1998) Repression of GCN5 histone acetyltransferase activity via bromodomain-mediated binding and phosphorylation by KU-DNA-dependent protein kinase complex. *Mol. Cell. Biol.* 18, 1349–1358

3 Dhalliun, C. *et al.* (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491–496

4 Jones, H.J. *et al.* (2000) Identification and characterization of BPTF, a novel bromodomain transcription factor. *Genomics* 63, 35–39

5 Aasland, R. *et al.* (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* 20, 56–59

6 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389–3402

7 Rost, B. *et al.* (1994) PHD an automatic mail server for protein secondary structure prediction. *CABIOS* 10, 53–60

8 Aravind, L. and Landsman, D. (1998) AT-Hook motifs identified in wide variety of DNA-binding proteins. *Nucl. Acids Res.* 26, 4413–4421

9 Bowser, R. *et al.* (1995) FAC1, a novel gene identified with the monoclonal antibody Alz50, is developmentally regulated in human brain. *Dev. Neurosci.* 17, 20–37

10 Gehring, W.J. *et al.* (1994) Homeodomain proteins. *Annu. Rev. Biochem.* 63, 487–526

11 Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* 14, 755–763

12 Jones, H.J. *et al.* (2000) A novel family of bromodomains. *Genomics* 63, 40–45

13 Morris, C.A. *et al.* (1988) Natural history of William's syndrome: physical characteristics. *J. Pediatr.* 113, 318–326

14 Lu, X. *et al.* (1998) A novel human gene, WSTF, is deleted in Williams Syndrome. *Genomics* 54, 241–249

15 Bochar, D.A. *et al.* (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 1038–1043

16 Schuler, G.D. *et al.* (1991) A workbench for multiple alignment construction and analysis. *Proteins* 9, 180–190

17 Schultz, J. *et al.* (2000) SMART: a Web-based tool for the study of genetically mobile domains. *Nucl. Acids Res.* 28, 231–234

18 Heery, D.M. *et al.* (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387, 733–736

19 Ito, T. *et al.* (1999) ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev.* 13, 1529–1539

20 Jordan-Sciutto, K.L. *et al.* (1999) Fetal Alz-50 clone 1, a novel zinc finger protein, binds a specific DNA sequence and acts as a transcriptional regulator. *J. Biol. Chem.* 274, 35262–35268

Tobias Doerks*
Richard Copley
Peer Bork
European Molecular Biology Laboratory, 69012 Heidelberg, Meyerhofstr. 1 and Max-Delbrueck-Centrum, PO Box 740238, D-13092 Berlin, Germany.
*e-mail: doerks@EMBL-HEIDELBERG.DE

Protein sequence motifs

Protein Sequence Motif is a regular column for brief reports of new motifs or sequence homologies that have been recognized in published sequences. Contributions to this column should be short (less than 500 words plus one figure) and will be subject to peer review. Preference will be given to reports of motifs or sequence homologies with profound biological significance. All sequences should have been published elsewhere in full and/or be freely available in the appropriate databases (e.g. GenBank, SWISS-PROT, etc.), but the particular motif or sequence alignment noted should not have been described before. Adequate statistical evaluation and, where appropriate, structural correlations should be given.