STRUCTURE-FUNCTION ANALYSIS OF THE THAP-ZINC FINGER OF THAP1. A LARGE C2CH DNA-BINDING MODULE LINKED TO RB/E2F PATHWAYS

Damien Bessière\textsuperscript{1,2,6}, Chrystelle Lacroix\textsuperscript{1,3,6}, Sébastien Campagne\textsuperscript{1,2}, Vincent Ecochard\textsuperscript{1,3}, Valérie Guillet\textsuperscript{1,4}, Lionel Mourey\textsuperscript{1,4}, Frédéric Lopez\textsuperscript{1,4}, Jerzy Czaplicki\textsuperscript{1,2}, Pascal Demange\textsuperscript{1,2}, Alain Milon\textsuperscript{1,2}, Jean-Philippe Girard\textsuperscript{1,3,6} and Virginie Gervais\textsuperscript{1,2,6}

\textsuperscript{1}University of Toulouse, Institute of Pharmacology and Structural Biology, IPBS-CNRS-UPS, Toulouse, France
\textsuperscript{2}Laboratory of NMR and protein-membrane interactions, \textsuperscript{3}Laboratory of Vascular Biology and \textsuperscript{4}Laboratory of Structural Biophysics, IPBS-CNRS-UPS, Toulouse, France
\textsuperscript{5}INSERM-IFR31, CHU Rangueil, Toulouse, France
\textsuperscript{6}These authors contributed equally to the work

Running title: Structure-function analysis of the THAP-zinc finger of THAP1

Address correspondence to: J.P.G or: V.G. University of Toulouse, Institute of Pharmacology and Structural Biology, IPBS-CNRS-UPS, Toulouse, France; E-mail: jean-philippe.girard@ipbs.fr or: virginie.gervais@ipbs.fr

THAP1, the founding member of a previously uncharacterized large family of cellular proteins (THAP proteins), is a sequence-specific DNA-binding factor that has recently been shown to regulate cell proliferation through modulation of pRB/E2F cell-cycle target genes. THAP1 shares its DNA-binding THAP-zinc finger domain with drosophila P element transposase, zebrafish E2F6 and several nematode proteins interacting genetically with the retinoblastoma protein pRb. In this study, we report the three-dimensional structure and structure-function relationships of the THAP zinc finger of human THAP1. Deletion mutagenesis and multidimensional NMR spectroscopy revealed that the THAP domain of THAP1 is an atypical zinc finger of ~80 residues, distinguished by the presence between the C2CH zinc coordinating residues of a short anti-parallel β-sheet interspersed by a long loop-helix-loop insertion. Alanine scanning mutagenesis of this loop-helix-loop motif resulted in the identification of a number of critical residues for DNA recognition. NMR chemical shift perturbation analysis was used to further characterize the residues involved in DNA binding. The combination of the mutagenesis and NMR data, allowed the mapping of the DNA-binding interface of the THAP-zinc finger to a highly positively charged area harboring multiple lysine and arginine residues. Together, these data represent the first structure-function analysis of a functional THAP domain, with demonstrated sequence-specific DNA-binding activity. They also provide a structural framework for understanding DNA recognition by this atypical zinc finger, which defines a novel family of cellular factors linked to cell proliferation and pRb/E2F cell cycle pathways in humans, fishes and nematodes.

Zinc finger proteins represent the most abundant class of DNA-binding proteins in the human genome. Zinc fingers have been defined as small, functional, independently folded domains that require coordination of a zinc atom to stabilize their structure (1). The zinc finger super family includes the C2H2-type zinc finger, a compact ~30 amino-acid DNA-binding module repeated in multiple copies in the proteins structure (2,3), the C4-type zinc finger found in the GATA family of transcription factors (4) and the zinc coordinating DNA-binding domain of nuclear hormone receptors (5). We recently described an atypical zinc finger motif, characterized by a large C2CH module (CysX\textsubscript{2-4}CysX\textsubscript{35-53}CysX\textsubscript{2}His) with a spacing of up to 53 amino acids between the zinc coordinating C2 and CH residues (6). This motif, designated THAP domain or THAP-zinc finger, defines a previously uncharacterized large family of cellular factors with more than one hundred distinct members in the animal kingdom (6,7). We showed that the THAP domain of THAP1, the prototype of the THAP family (8), possesses zinc-dependent sequence-specific DNA-binding activity and recognizes a consensus DNA target sequence of 11 nucleotides (THABS, for THAP1 binding sequence) (7), considerably larger than the 3-4 nucleotides motif typically recognized by classical C2H2 zinc fingers (2,7). Interestingly, the consensus C2CH signature of the THAP domain was identified in the sequence-specific

\textsuperscript{1,2,6}Copyright 2007 by The American Society for Biochemistry and Molecular Biology, Inc.
DNA-binding domain of Drosophila P element transposase, suggesting that the THAP-zinc finger constitutes a novel example of a DNA-binding domain shared between cellular proteins and transposons from mobile genomic parasites (6,9).

Although the biological roles of cellular THAP proteins remain largely unknown, data supporting an important function in cell proliferation and cell cycle control have recently been provided. We found that human THAP1 is an endogenous physiological regulator of endothelial cell proliferation and G1/S cell cycle progression, which modulates expression of several pRb/E2F cell cycle target genes. In addition, we identified RMM1, a G1/S-regulated gene required for S-phase DNA synthesis, as a direct transcriptional target of endogenous THAP1 (10). These data provided the first links in mammals between THAP proteins, cell proliferation and pRb/E2F cell cycle pathways and complemented genetic data previously obtained in model animal organisms. Indeed, in zebra fish and other fish species, the ortholog of cell cycle transcription factor E2F6, a repressor of E2F-dependent transcription during S phase (11) was found to contain a THAP-zinc finger at its amino-terminus (7). In the nematode Caenorhabditis elegans, five distinct THAP-zinc finger proteins (LIN-36, LIN-15B, LIN-15A, HIM-17 and GON-14) (7) were shown to interact genetically with LIN-35/Rb, the sole C. elegans retinoblastoma homolog (12-16).

Among these, GON-14 appeared to function as a positive regulator of cell proliferation, since cell division defects were observed in the intestine, gonad and vulva of gon-14 null mutant (16). In contrast, LIN-36 and LIN-15B, initially characterized for their role in the specification of vulval cell fates (synthetic Multivulva class B genes, synMuvB) (12,13), were found to function as inhibitors of the G1/S cell cycle transition (14). LIN-36 behaved most similar to LIN-35/Rb and Efl-1/E2F, the ortholog of mammalian cell cycle transcription factors E2F4/5, and was therefore proposed to act in a transcriptional repressor complex with these factors to repress G1/S control genes (14,17,18). However, LIN-36, LIN-15B and THAP1 were not found in the evolutionary conserved pRb/E2F protein complexes (DREAM or DRM complexes) that have recently been described in Drosophila, C. elegans and human cells and that contain pRb/p130, E2F4/5, DP and five other synMuvB gene products LIN-9, LIN-37, LIN-52, LIN-53, LIN-54 (19-22). This suggests that THAP-zinc finger proteins may function in distinct transcriptional regulatory complexes to regulate E2F target genes. Although not associated with Rb complexes, THAP zinc finger proteins may still act at the level of chromatin regulation since several C. elegans THAP family members have been found to interact genetically with components of diverse chromatin-modifying and/or chromatin-remodeling complexes, including members of the Nucleosome Remodeling Deacetylase (NuRD) complexes and components of the Tip60/NuA4 histone acetyltransferase complex (12-16,23,24). In addition, the human THAP7 protein has also been shown to interact with chromatin-modifying enzymes (25). Together, these observations indicate that both in humans and model animal organisms, THAP-zinc finger proteins appear to be critical regulators of cell proliferation and cell cycle progression, likely to act at the level of chromatin regulation.

Solution structures of the THAP domains from two previously uncharacterized proteins, human THAP2 and C. elegans CtBP, have recently been reported (THAP2, PDB code 2DR8; CtBP, PDB code 2JM3 (26)). However, sequence-specific DNA-binding properties have not yet been demonstrated for these two domains. Here, we report the first structure-function analysis of a functional THAP domain, the THAP-zinc finger of human THAP1. The three-dimensional structure of the domain was determined by multidimensional NMR spectroscopy and its DNA-binding interface was characterized by a combination of alanine scanning mutagenesis and NMR chemical shift perturbation analysis. Together, these data provide a better understanding of the structure-function relationships of this atypical zinc finger.

**Experimental procedures**

**Plasmid constructions** The THAP domain of human THAP1 (Met1-Phe81 or Met1-Lys90; GenBank # NP_060575) was amplified by PCR and cloned in frame with a carboxy terminal His-Tag into a modified pET-26 plasmid (Novagen). The pET-21c-THAP domain expression vectors for human THAP2 (residues 1-90; GenBank # NP_113623), human THAP3 (residues 1-92; GenBank # AA92427), C. elegans CtBP (residues 1-88; GenBank # NP_508983) and C. elegans GON-14 (residues
1-84; GenBank # NP_741558) were generated by PCR as previously described for the pET-21c-THAP1 (Met1-Lys90) plasmid (7). Construction of the pcDNA3-THAP1 eukaryotic expression vector has previously been described (7). The THAP1 alanine scanning single point and triple mutants were obtained by PCR using specific primers containing the corresponding mutations, and cloned as EcoRI-XbaI fragments in pcDNA3 expression vector.

**Protein expression and purification** - For NMR experiments, recombinant THAP domains of human THAP1 {Met1-Lys90} and {Met1-Phe81} were produced as His-tag fusion proteins in E. coli BL21(DE3). Cells were grown in LB medium at 37°C to an A600 of 0.8 before induction with 1 mM IPTG, to obtain an unlabeled sample. ZnCl₂ was added at this step (final concentration of 0.01 mM). Isotopically ¹⁵N/²⁵C-labeled THAP domain {Met1-Lys90} and ¹⁵N-labeled THAP domain {Met1-Phe81} were expressed in minimal (M9) medium containing ¹³NH₄Cl and ¹⁵N Cetolene and either ¹³C glucose or ¹²C glucose. Proteins were purified using a Ni-NTA column (HiTrap, Amersham) followed by gel-filtration chromatography on Sephadex G75 (Amersham). After digestion with thrombin (Novagen), proteins were further purified on a gel filtration column. NMR samples were concentrated to 0.4-1.7 mM in 50 mM deuterated Tris-HCl, pH 6.8, 1 mM DTT with either 10 mM NaCl (for protein structure determination) or 250 mM NaCl (for DNA binding studies).

For gel-shift assays, the recombinant THAP domain of human THAP1 was produced as previously described (7). The recombinant THAP domains of THAP2, THAP3, Ce-CtBP or GON-14 were produced in E. coli strain BL21trxB(DE3), transformed with the different pET-21c-THAP domain expression vectors. Protein expression and purification was performed according to manufacturer’s instructions (Novagen, Madison, WI), as previously described (7). The purity of the different THAP domains was assessed by SDS-PAGE and protein concentrations were determined using Bradford protein Assay (Bio-Rad). Full-length THAP1 wild type and mutants were synthesized in vitro in rabbit reticulocyte lysate (RLL). The corresponding pcDNA3 expression vectors were used with the TNT-T7 kit (Promega). Protein production was performed in the presence of ³⁵S-labeled methionine and verified by SDS-PAGE and autoradiography.

**Surface Plasmon Resonance Experiments** - DNA interaction kinetics was investigated by Surface Plasmon Resonance (SPR) assays using a four-channels BIAcore 3000 optical biosensor instrument (BIAcore AB, Uppsala, Sweden). Immobilization of biotinylated single-stranded DNA probes was performed on a streptavidin-coated sensorchip (BIAcore SA sensorchip) in HBS-EP buffer (10 mM Hepes pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.0005% Surfactant P20) (BIAcore AB, Uppsala, Sweden). All immobilization steps of biotinylated single-stranded DNA probes were performed at a final DNA concentration of 100 ng/ml and at a flow rate of 2 µl/min. Hybridization of complementary DNA strands was performed in HBS-EP buffer supplemented with 200 mM NaCl. Biotinylated oligonucleotide sequences and complementary DNA strands were purchased from MWG Biotech.

Binding analyzes were performed with multiple injections of THAP domain {Met1-Lys90} at different protein concentrations over the immobilized surfaces at 25°C. A second DNA probe with unrelated sequence was used as a control. All samples were diluted in the running buffer containing 50 mM Tris (pH 6.8), 250 mM NaCl, 1mM DTT and were injected over the sensor surface for 4 minutes at a flow rate of 20 µl/min. No binding was observed to the DNA probe with unrelated sequence. The SPR signal was therefore analyzed as difference sensograms between the two DNA sequences immobilized to separate channels of the sensor chip (the signal from the unrelated DNA used as control was subtracted).

**NMR spectroscopy** - NMR experiments were recorded at 296 K on Bruker Avance 800, 700 and 600 MHz spectrometers. Backbone and side-chain resonances (¹H, ¹⁵N and ¹³C) of the THAP domain {Met1-Lys90} were assigned using a set of heteronuclear experiments (27). This information was partially used to assign ¹H and ¹⁵N resonances of the THAP domain {Met1-Phe81} using a combination of homonuclear and ¹⁵N heteronuclear spectra. φ/ψ torsion angles were derived using TALOS (28) and the ³JHN-Hα coupling constants obtained from 3D-HNHA (29). Four X-Pro bonds were identified as being in the trans-configuration on the basis of strong NOEs between the Hδ proton of each Pro residue and the Hα protons of the preceding residues (30) and confirmed by ¹³C chemical
From characteristic NOEs, the residue Pro26 was identified as being in the cis form (30). 15 stereo-specific assignments of Hβ methylene were obtained using DQF-cosy spectrum (32). NMR data were processed using TOPSPIN software (Bruker) and NMRPipe (33) and analyzed using XEASY (34) and NMRView (35).

15N relaxation data were recorded at 296 K on a 0.9 mM protein sample at 10 mM and 250 mM NaCl using standard pulse sequences. The heteronuclear NOEs were determined from two 15N-HSQC spectra recorded in presence and absence of 2H presaturation period of 3 s and with a recycling delay of 5 s (36).

Structure calculations- To solve the structure of the THAP domain {Met1-Phe81}, a set of distances was extracted from integration of 2D 1H and 3D 15N heteronuclear NOESY spectra. The secondary structure elements were derived from analysis of coupling constants, from identification of slowly exchanging amide protons and from characteristic NOEs. To maintain well-defined secondary structure elements, hydrogen bonds were added with restraints of 1.8 to 2.4 imposed on the distance between hydrogen and acceptor oxygen and restraints of 2.3 to 3.2 imposed on the distance between the donor nitrogen and acceptor oxygen.

Preliminary structure calculations run either with Nδ1 or Ne2 of the His57 ring allowed us to identify Ne2 as the zinc-bound atom. Subsequent structural refinement including a zinc ion together with constraints defining tetrahedral coordination (3,37) was performed. The structures were calculated using a torsion-angle dynamics simulated annealing protocol using the CNS software suite (38). From 500 initial structures, a set of 20 structures were selected as accepted structures, based on the following criterions: low total energy, no distance violation larger than 0.2 Å and no torsion angle violation greater than 2°. Their structural quality was analyzed using PROCHECK (39).

NMR chemical shift perturbation analysis-The 14-bp duplex DNA containing THABS was reconstituted by hybridizing the following oligonucleotides, 5'-CAAGTATGGCGAG-3' and 5'-CTTGGCCATCTTG-3' in a 1:1 ratio. For NMR titration, 2D 15N HSQC spectra of the THAP domain at the concentration of 0.4 mM were collected at 296K and 250 mM NaCl after each incremental addition of lyophilized DNA. 1D 1H spectra were recorded after each DNA addition and the DNA/protein ratio was followed from integration of 1H protein and DNA signals on 1H spectra. A DNA fragment with an unrelated sequence was reconstituted by hybridizing the oligonucleotides, 5'-GATTTCATTTTA-3' and 5'-TTAAAAATGCACATC-3', and added to the THAP domain following the same procedure as described above. Normalized chemical shift changes were calculated as: \( \Delta \delta = \sqrt{(\Delta \delta_{\text{HN}})^2 + (\Delta \delta_{\text{N}} \times 0.154)^2} \). (32)

Electrophoretic mobility shift assays (EMSA)-EMSA were performed with purified recombinant THAP domains produced in E. coli or with full length THAP1 wild type or THAP1 mutants synthesized in vitro in RRL, using the following THABS probes, 25-bp (5'-AGCAAGTAAGGGCAAACTACTCAT-3') and 36-bp (5'-TATCAACTGTTGGCAACACTACGGGCAACAGGTAATG-3'), as previously described (7). Increasing amounts of purified recombinant THAP domains were incubated for 20 minutes at room temperature in 20 μl binding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 0.1% NP-40, 100 μg/ml BSA, 2.5 mM DTT, 5% glycerol, 10 μg/ml poly(dl/dC)). For in vitro translated proteins, 3μl of RRLs expressing full length THAP proteins or THAP1 mutants were incubated in 20 μl binding buffer supplemented with 50 μg/ml of poly(dl/dC) and 50 μg/ml salmon sperm DNA. Electrophoresis was performed and gels were exposed as previously described (7). Supershift experiments were performed using 1 μg of anti-THAP1 affinity-purified rabbit polyclonal antibodies (10).

Model building of the complex between the THAP-zinc finger and the THABS sequence-Computational docking of the THABS DNA target onto the THAP-zinc finger {Met1-Phe81} of THAP1 was performed using HADDOCK1.3 (High Ambiguity Driven DOCKing) (40), in conjunction with CNS (38). The docking was performed using an ensemble of THAP-zinc finger structures of THAP1 and the 14 bp double-stranded DNA containing THABS built as a B-DNA template using Insight II (Accelrys). The active residues used to define the Ambiguous Interaction restraints (AIR) included residues showing relative solvent accessibility higher than 40% (as calculated by NACCESS, Hubbard and Thornton, University College London) and either displaying a
chemical shift perturbation higher than 0.2 ppm
upon DNA binding or giving rise to loss of DNA
binding from site-directed mutagenesis. Briefly,
avtive residues were Lys24, Glu37, Arg42, Lys
46 and Thr48. Solvent accessible residues that
were surface neighbors of the active residues
were defined as passive residues including
Lys34, Glu35, Ala38, Arg41, Lys43, Asn44,
Lys49 and Tyr50.

For DNA, bases corresponding to the
core GGCA motif of the THABS sequence
together with the thymine upstream were defined
as active residues, as based on already reported
data obtained from scanning mutagenesis (7).

The docking was performed on a SGI cluster
equipped with 40 processors. Starting from 15
THAP domain structures of THAP1 and a model
of the THABS target in B-DNA conformation,
1000 rigid-body solutions were generated. The
best 200 solutions according to the HADDOCK
rigid-body score were selected for semi-flexible
refinement in torsion angle space; the top 200
structures were finally refined in explicit water.
The final ensemble of 200 solutions was
analyzed and clustered based on a pair-wise
RMSD matrix calculated over the backbone
atoms.

RESULTS

Biophysical characterization of the
THAP domain of THAP1 and identification of a
shorter functional fragment. The THAP domain
was originally assigned to the first 90 N-terminal
residues of THAP1 (6). A corresponding
fragment was initially expressed and purified for
Surface Plasmon Resonance (SPR) and NMR
studies. Inductively Coupled Plasma-Mass
Spectrometry (ICP-MS) experiments indicated
that the domain includes a zinc ion (data not
shown). The SPR experiments (Fig. 1A-B)
showed that the THAP-zinc finger of THAP1
binds to a 14 bp THAP domain binding site
(THABS DNA probe) that includes the GGCA
core motif at sequence positions 7-10,
previously found to be critical for recognition by
the THAP domain of THAP1 (7).

A series of triple-resonance NMR
experiments allowed us to assign unambiguously
residues 3-63 and 74-77 of the domain.
However, only partial assignment for residues
64-73 and 78-90 in the C-terminal tail could be
performed due to severe lack of connectivity for
these residues. These data together with
characteristic negative heteronuclear NOE
values (data not shown) indicated that the C-
terminus tail of the domain is unstructured.

These results led us to search for the minimal
size of the functional THAP-zinc finger. The
cysteine residues at the N-terminus of the THAP
domain (THAP1 Cys5 and Cys10) have
previously been shown to be required for the
functional activity of the domain, for both
human THAP1 (7) and drosophila P element
transposase (41), thus defining the N-terminal
boundary of the domain. In contrast, nothing
was known about the requirements at the C-
terminus downstream of the conserved AVPTIF
motif (6) containing the essential Pro78 residue
(7). Alanine scanning mutagenesis was therefore
performed and revealed that residues 82 to 90
are not required for DNA-binding activity of the
THAP domain of THAP1 (data not shown). This
was confirmed using a THAP1 deletion mutant,
THAP1Δ82-90, that exhibited a similar activity
in EMSA experiments than wild type THAP1
(data not shown). In addition, recombinant
THAP domains {Met1-Phe81} and {Met1-
Lys90} exhibited similar activities in EMSA
experiments indicating that residues 1-81 are
sufficient for sequence-specific DNA binding to
the THABS motif (Fig. 1C). A deletion mutant
referred to fragment {Met1-Phe63} was also
tested both by NMR and EMSA experiments,
but the HSQC spectrum corresponded to that of
an unfolded protein and the fragment did not
possess any DNA-binding activity (data not
shown). Therefore, the recombinant THAP
domain {Met1-Phe81} was selected for all
structural studies.

NMR solution structure of the THAP-zinc finger of human THAP1. We solved the
three-dimensional solution structure of the zinc-containing form of the THAP domain
{Met1-Phe81} using a set of distances extracted from
2D and 3D NOESY spectra. For the THAP
domain in its DNA-free state, spectra were
recorded in a buffer containing 10 mM NaCl. In
these conditions, 1539 distance restraints
obtained from 2D 1H NOESY and 3D 15N
HSQC-NOESY spectra and 104 angle restraints
were used for calculations (Table 1).

The core of the THAP-zinc finger of
THAP1 adopts a ββββ fold (Fig. 2A-D). Residues
Cys5, Cys10, Cys54, His57 form a single zinc-
binding site that begins with a long loop L1
(residues Gln3-Ser21), which precedes the first
β-strand (β1; residues Phe22-Lys24). This
portion is followed by a second loop L2 (Phe25-
Lys32), which continues into a α-helix H1
(residues Cys33-Val40). An additional loop L3 (residues Arg41-Ser51) is followed by the second β-strand (β2; residues Ser52-Cys54) running anti-parallel to β1. The zinc-binding site is completed by a short 3₁₀ helix (H2; residues Ser55-His57). A second 3₁₀ helix (H3; residues Pro60-Phe63) is followed by a flexible loop L4 (Lys64-Asn68) that continues into an extended region (Asn69-Pro78) followed by an additional 3₁₀ helix (H4) including residues Thr79-Phe81 that form part of the AVPTIF motif (6). Some regions of the loops are relatively well defined (L1: 6-16; L2: 26-32; L3: 46-50; L4: 69-78) with restricted mobility as judged by the heteronuclear NOEs, whereas the other parts of the loops are less ordered and display a mobility on ns-ps time scale (Fig. 2C and Supplementary Fig. 1). In particular, beginning of loop L4 displays a high degree of structural disorder although its heteronuclear NOEs are not much lower than those of loop L1. This is due to the scarcity of NOEs involving this region despite extensive search in the NOESY maps.

Only a few amide protons are protected as observed from H₂O/D₂O experiments. They mainly correspond to some residues located within the helix H1 (Glu35, Trp36, Glu37, Ala39, Val40) and the two-stranded β-sheet (Ile53, Cys54) as well as few additional residues (Ser6, Ala7, Cys10, Asn12) in the vicinity of the zinc-binding site. Apart from these residues, the THAP domain amide protons exchange rapidly with the solvent (data not shown).

A structure-based sequence alignment of the THAP zinc finger of THAP1 with representative THAP domains is shown in Fig. 2E. Besides the strictly conserved C₂CH motif that provides ligands for the zinc ion, the THAP domain is defined by a C-terminal AVPTIF motif and four residues (Pro26, Trp36, Phe58 and Pro78, numbering refers to THAP1) that are invariant in more than hundred THAP domain sequences and that are absolutely required for DNA-binding activity (6,7). The unique tryptophane (Trp36) located in the α-helix H1 is a key element of the THAP-zinc finger structure and constitutes the anchoring residue that makes hydrophobic contacts with the conserved Phe58 residue (Fig. 2D) and the surrounding aromatic residues, namely Phe25 and Phe63. In addition, NOEs are detected between Trp36 at the centre of the hydrophobic core and the two invariant prolines (Pro26 in the loop L2 and Pro78 in the AVPTIF motif). Both of the prolines display strongly upfield-shifted resonances due to the proximity of Trp36. NOEs are also observed between Phe81 in the AVPTIF motif and Ala39-Val40 in the helix H1 (data not shown). Therefore, the AVPTIF motif appears to play an essential role in the folding of the THAP-zinc finger by bringing together the C-terminus and the α-helix H1 (Fig. 2D).

The THAP-zinc fingers share the same three-dimensional fold but not the same DNA binding site. Comparison of the structure of the THAP zinc fingers from human THAP1, THAP2 and C. elegans CtBP revealed that the overall fold and the packing around the tetrahedrically zinc-coordinating site are similar for the three THAP domains (Fig. 3A). The structural homology is higher between the THAP domains of THAP1 and THAP2, as expected for closely related sequences (48%). Indeed, the solution structure of the THAP domain of THAP1 can be superimposed onto that of THAP2 for 80 Cα equivalent residues with an RMSD value of 2.8 Å. A weaker score is found for the superimposition of the THAP domain of THAP1 onto that of CtBP with 66 Cα equivalent atoms that could be superimposed with an RMSD value of 3.1 Å. It is noteworthy that the sequence identity between these two domains is only of 27%. The core fold consisting of the anti-parallel β-sheet with the two strands separated by a loop-helix-loop insertion is conserved among the three THAP domains. Nevertheless, the C-terminus displays structural variability. Indeed, the THAP domain of THAP1 shows two additional short 3₁₀ helices H3 and H4, encompassing residues 60-63 and 79-81 respectively whereas the THAP domains of THAP2 and CtBP display additional two-stranded anti-parallel β-sheets. In the structure of the THAP domain of THAP2, the second anti-parallel β-sheet is formed by residues that would correspond to residues F63-K64 and L71-L72 in THAP1. In the CtBP structure instead, the second anti-parallel β-sheet involves residues that would correspond to THAP1 residues A76-V77 in the AVPTIF motif and the two residues L82-C83 that follow the AVPTIF motif. Since the M1-F81 fragment of THAP1 retains its capacity to bind DNA (Fig. 1C), the second β-sheet observed for the THAP domain of CtBP is unlikely to be important in the molecular scaffold of the THAP-zinc finger.

It is noteworthy that the flexible loop L4 (residues 64-68) between H3 and H4 in THAP1
is also observed in THAP2 but is absent in CtBP since it corresponds to an 8-residues sequence insertion compared to the THAP domain of CtBP. Despite these discrepancies in the secondary structure elements, residues in the C-terminal region keep mostly equivalent positions in the three structures and positions of residues A76-F81 that form part of the AVPTIF motif in THAP1 are identical when compared to that of THAP2.

Since different THAP-zinc fingers appear to exhibit a similar three-dimensional fold, we then studied the possibility they may recognize the same DNA target sequence. Among the 12 human THAP proteins, THAP2 and THAP3 are the most closely related to THAP1 and for instance, the THAP domains of THAP1 and THAP3 exhibit up to 50 % identity (6). Therefore, we tested the ability of these two proteins to bind to the THABS motif specifically recognized by THAP1 (7). As shown in Fig. 3B, we found that the recombinant THAP domains of THAP2 and THAP3 did not bind to the THABS probe in gel shift-assays. In contrast, the recombinant THAP domain of THAP1, used in the same conditions, exhibited strong binding to the THABS sequence (Fig. 3B). These results were confirmed using in vitro translated full-length THAP proteins, which provided an independent source of THAP domains. In contrast to THAP1, the full-length THAP2 and THAP3 proteins did not bind to the THABS probe in gel-shift assays (data not shown).

Liew et al. recently reported binding of the THAP domain of C. elegans CtBP (CeCtBP) to the THABS sequence recognized by human THAP1 (26). However, their gel-shift assays were performed in the absence of competitor DNA and we considered the possibility that their observations may correspond to a non-sequence specific DNA-binding activity of the THAP domain of CeCtBP. We therefore performed gel shift assays with the recombinant CeCtBP THAP domain in the presence or absence of the synthetic poly(dI/dC) non-specific competitor DNA. In agreement with the results reported by Liew et al. (26), we observed binding of the THAP domain of CeCtBP to the THABS motif in the absence of competitor (Fig. 3C). However, no specific protein-DNA complex was observed in the presence of the poly(dI/dC) competitor. Similar results were obtained with the recombinant THAP domain from another C. elegans THAP protein, the cell proliferation and developmental regular GON-14 (Fig. 3C). In contrast, strong binding of the THAP domain of THAP1 to the THABS sequence was observed in the presence of competitor (Fig. 3C). We concluded that CeCtBP and GON-14 THAP-zinc fingers do not bind specifically to the THABS motif recognized by THAP1.

Together with the findings on human THAP2 and THAP3, these results indicate that the different THAP-zinc fingers share the same three-dimensional fold but not the same DNA target sequence.

**Structure-function analysis of the THAP-zinc finger of THAP1 by site-directed mutagenesis.** We have previously shown that the eight invariant residues that define the THAP domain are absolutely required for DNA-binding activity (7). To get further insights into the role of other residues in DNA recognition, thirty additional residues were individually mutated to alanine and the resulting mutants were tested in gel-shift assays (Fig. 4A-D and Table 2). These included twenty-four consecutive residues (Leu27 to Ser52) from the long loop-helix-loop motif (L2-H1-L3) inserted into the anti-parallel β-sheet, one of the most distinctive features of the THAP-zinc finger (Fig. 2). Single-point mutation of the invariant Trp36 in the centre of the α-helix was used as a control and, as expected, this mutation completely abolished the interaction. Similarly to mutation of Trp36, mutation of residues Lys24, Arg29, Arg42, Phe45 and Thr48 led to a complete loss of DNA-binding activity whereas mutation of residues Lys11, Leu27, Glu37, Val40 and Tyr50 decreased but did not abrogate the interaction of the THAP-zinc finger with its THABS DNA target sequence (Fig. 4A-D and Table 2). Triple mutations of residues Thr28-Arg29-Pro30, Arg41-Arg42-Lys43 and Pro47-Thr48-Lys49 to alanines were also performed and confirmed the importance of these regions for DNA-binding activity of the THAP domain of THAP1 (Fig. 4C-D and Table 2). Interestingly, triple alanine mutation of residues Tyr50, Ser51 and Ser52 revealed a critical role for these residues that was less apparent in the single point mutants. Mapping of the essential residues on the THAP domain structure of THAP1 revealed that residues Arg29 and Phe45 make contacts with the hydrophobic core of the domain. Indeed, the Arg29 residue in the loop L2 is shown to make several NOE contacts with protons of residues Leu32 in the loop L2, Trp36 in the helix H1 and Ala76 that is part of the AVPTIF motif (Fig.
The residue Phe45 that is part of the loop L3 gives NOEs to Val40 in the helix H1 and to Phe25 in the loop L2 (Fig. 4E). Therefore the loss in DNA binding after mutation of the two residues Arg29 and Phe45 could be due to a disruption of local structure.

In contrast, residues Lys24, Arg42 and Thr48 are exposed at the surface of the THAP domain. Interestingly, they map onto the area of the domain that is highly positively charged due to the presence of several exposed basic side chains of lysines and arginines consistent with DNA interaction (Fig. 4F). These data strongly suggest that the three residues Lys24, Arg42 and Thr48 may be directly involved in DNA binding.

Identification of the DNA-binding interface of the THAP-zinc finger by NMR chemical shift perturbation analysis. To further characterize the DNA-binding interface, binding of the THAP domain of THAP1 to the THABS DNA target sequence was probed by NMR chemical shift perturbation analysis. The 15N-labeled THAP domain dissolved in NMR buffer containing 250 mM NaCl was titrated with 14-bp duplex DNA containing THABS. In the absence of DNA, the spectrum recorded at 250 mM NaCl was similar to the one recorded at 10 mM NaCl, except a few peaks that slightly shifted (data not shown). In addition, based on 15N longitudinal and transverse relaxation times, rotational correlation times (τc) were determined to be 6.03 ± 0.1 ns and 6.89 ± 0.4 ns at 10 mM and 250 mM NaCl, respectively, indicating that the protein is monomeric in both salt conditions (data not shown). In the presence of increasing concentrations of the THABS oligonucleotide, several cross-peaks were significantly affected during titration (Fig. 5A). A similar 2D 15N-HSQC spectrum recorded in the presence of an unrelated 14-bp duplex DNA did not reveal any significant shift (data not shown). The chemical shift perturbations observed in the presence of the specific THABS sequence were not further affected when DNA:protein ratio was above 1:1 suggesting a 1:1 binding stoichiometry, in agreement with Surface Plasmon Resonance experiments (Fig. 1A-B).

During titration, the majority of the affected signals could be followed as the fast-exchange manner, i.e., a single cross peak with intermediate chemical shift between that of the free and bound forms (Fig. 5A). Signals with the largest chemical shift changes could be followed as the slow-exchange manner with two peaks corresponding to the free and bound forms with intensities proportional to the free/bound ratio. Finally, the coalescence could be observed for a couple of residues (Arg29, Leu32) with a single signal (larger line-width at intermediate DNA/protein ratio) at a chemical shift variation of about 50 Hz. The average chemical shift changes (Δδ = [(ΔδHN)2 + (ΔδN x 0.154)2] ½) between the free and the DNA-bound state of the THAP domain were plotted versus the THAP1 residue numbers (Fig. 5B). Several protein residues experienced significant chemical shift perturbation (Δδ higher than 0.2 ppm) and these were mainly organized into three different patches. The first one includes residues Ala7, Tyr8, Lys11, Arg13, Val20, Ser52, Ile53, Cys54 and Glu56 located in the region encompassing the zinc atom. The second patch reveals residues Arg29, Trp36, Ala38 and Ala39 within or nearby the α-helix H1. Residues in the third patch (Asn68, Asn69, Lys70, Leu72, Ala76) are located in the loop L4. Two additional residues undergoing large chemical shift changes are located in the loop L3 following the α-helix H1 and correspond to residues Lys46 and Thr48. Notably, most of these residues are located either in structured regions of the domain (H1, H2 and β2), or in ordered regions of the loops L1 (residues 7-11), L2 (Arg29), L3 (Thr48) and L4 (residues 69-76) that exhibit restricted motions (Fig. 2C and Supplementary Fig. 1).

Interestingly, the clusters of large chemical shift changes upon DNA binding map to the highly positively charged area of the THAP domain (Fig. 4F), further supporting the potential role of this region as the DNA-binding interface of the THAP-zinc finger.

DISCUSSION

We report here the three-dimensional structure and structure-function analysis of the sequence-specific DNA-binding THAP-zinc finger of human THAP1, the prototype of a novel family of cellular factors involved in pRb/E2F cell cycle pathways. We recently demonstrated that THAP1 is a physiological regulator of cell proliferation. Silencing of THAP1 by RNA interference in human primary endothelial cells resulted in inhibition of G1/S cell cycle progression and down-modulation of several pRb/E2F cell cycle target genes, including RRMI, a gene activated at the G1/S transition and essential for S-phase DNA synthesis (10). We showed that the THAP-zinc finger of THAP1 recognizes a consensus
THAP1-binding site in the RRM1 promoter and that endogenous THAP1 associates in vivo with this site, indicating that RRM1 is a direct target gene of THAP1. The solution structure of the THAP domain of THAP1 is therefore the first structure of a THAP-zinc finger with demonstrated biochemical activity as a sequence specific DNA-binding domain (7) and associated with a known biological function, i.e. recruitment of THAP1 on the pRb/E2F target gene RRM1 (10). In contrast, although the structure of the THAP domains from human THAP2 (PDB code 2D8R) and C. elegans CtBP have been determined (26), these two proteins have not been functionally characterized and it is not yet known whether their THAP domains possess sequence-specific DNA-binding properties.

The structure of the THAP-zinc finger differs from that of other DNA-binding modules belonging to the zinc finger superfamily. For instance, the βαβ topology, the long spacing between the two pairs of zinc ligands (up to 53 residues in some THAP domains) distinguish the THAP-zinc finger from the classical DNA-binding C2H2-zinc finger, which exhibits a ββα topology with a shorter spacing (10-12 residues) between the two pairs of zinc-coordinating residues (3). The position of the zinc is also an interesting feature; in the classical zinc finger, the zinc atom plays a central role in the structure by coordinating four ligands that anchor one end of the helix to one end of the β sheet, whereas in the C2CH THAP motif, the zinc is not buried in the interior of the protein and it links the N terminus of the domain to the second β-strand, without involving the α-helix that is distal to the zinc ion. The presence of the long loop-helix-loop insertion in the two-stranded anti-parallel β-sheet is one of the most distinctive features of the THAP-zinc finger. It explains the atypical spacing of the C2 and CH residues in the C2CH zinc coordinating module and the relatively large size of the THAP domain (~80 residues) compared to the C2H2-zinc finger (~30 residues). The above features are very unique and are not found in other classes of zinc-coordinating DNA-binding modules.

Surprisingly, however, the THAP-zinc finger exhibits structural similarities with a protein-protein interaction module, the Zinc Finger-Associated Domain (ZAD, PDB code 1PZW) of the drosophila transcription factor Grauzone (26,42). These structural homologies include the presence between the zinc-coordinating residues of a similar loop-helix-loop insertion into the two-stranded anti-parallel β-sheet. However, despite these similarities in their molecular scaffolds, the THAP-zinc finger and the ZAD domain are linked to different functions. The ZAD domain mediates protein-protein interactions and exhibits a highly negative electrostatic potential inconsistent with DNA-binding properties (42). In contrast, the THAP-zinc finger of THAP1 functions as a sequence specific DNA-binding module with a highly positively charged surface (Fig. 4F).

Our previous mutagenesis studies have revealed that mutation of any of the eight residues that define the THAP-zinc finger motif (including the C2CH residues) abrogate DNA-binding activity of the domain (7). Mapping of these residues on the THAP-zinc finger structure indicates that these amino acids play an essential role in the folding of the domain (Fig. 2D). In the present study, we identified five additional residues that are essential for DNA-binding activity. Two of these residues (Arg29 and Phe45) could play a structural role by anchoring the loops of the loop-helix-loop motif to the hydrophobic core of the domain, potentially limiting the motions of these loops (Fig. 4E). The three other essential residues (Lys24, Arg42, Thr48) are exposed on the positively charged surface of the THAP domain of THAP1 (Fig. 4F) and are therefore less likely to contribute to the folding or structure of the domain. Rather, they may play a direct role in DNA binding. Although Lys24 and Arg42 do not display significant chemical shift changes upon DNA binding, the residue Thr48 is clearly affected and undergoes the largest change in chemical shift, consistent with it being directly involved in DNA interactions (Fig. 5B). Furthermore, the residue Thr48 is poorly conserved among the THAP domains (Fig. 2E) suggesting a key role in binding specificity.

Surprisingly, with the exception of Trp36, mutation of residues located in the helix of the loop-helix-loop motif did not abrogate DNA binding. Therefore, despite the fact the α-helix is the most common protein structural element used for DNA recognition in zinc fingers and other types of DNA-binding domains (2,43,44), our results strongly suggest that the α-helix of the THAP-zinc finger may not be the main DNA recognition element. In contrast, the positively charged region is likely to play a key role (electrostatic contacts) together with more
specific contacts involving residues in the loops, namely the loop L3 of the loop-helix-loop motif, (Arg42, Thr48) and the loop L4 at the C terminus. Based on the NMR and mutagenesis data, a model of the complex between the THAP-zinc finger of THAP1 and its DNA target was built using HADDOCK1.3 (40) (Fig. 6). The proposed model shows a good shape complementarity between the loop-helix-loop motif (L2-H1-L3) and DNA. The helix does not appear as the major recognition element but is located along the DNA chain so that the two loops on its sides fit into the DNA grooves. Remarkably, the loop L3 enters into the major groove to contact DNA and in particular, side chain of Thr48 gives a polar contact with the GCGA core. Although the EMSA assays allowed us to define critical residues in this loop, these qualitative assays may be too insensitive to detect the role of other residues in DNA binding. Additional experiments may reveal, for instance, a role for Lys46, a residue that undergoes significant chemical shift changes upon DNA binding (Fig. 5), and is predicted to be in close proximity to DNA in the model of the protein-DNA complex (Fig. 6).

Interestingly, genetic data obtained in C. elegans for THAP family members LIN-36 and HIM-17 have revealed several single-point mutations which affect the functional activity of the THAP-zinc finger. Most of these mutations concern residues that are critical for the folding of the domain. For instance, mutation of the second Cys of the C2CH motif was found in one of the THAP domains of HIM-17 (15), while two independent mutations were found in the last Pro residue of the THAP motif in the LIN-36 protein (13). However, other mutations have been found to affect residues that do not appear to be part of the hydrophobic core of the domain and these may correspond to residues exposed on the surface and directly involved in DNA binding. Finally, a double alanine mutation introduced into the THAP–zinc finger of drosophila P element transposase at the level of residues His18 and Cys22 (corresponding to THAP1 residues Lys24 and Arg29) has previously been shown to abrogate sequence-specific DNA-binding activity of the protein (41). This suggests that the essential residues we have identified in the present study are likely to be also critical for the functional activity of other THAP-zinc fingers.

In this study, we show that the different THAP-zinc fingers, despite sharing some structural homologies, do not recognize the same DNA target sequence. We found that recombinant THAP domains from human THAP2 and THAP3, and C. elegans CtBP and GON-14 do not exhibit sequence-specific DNA-binding activity towards the DNA sequence motif recognized by human THAP1 (Fig. 3). Although Ce-CtBP and GON-14 were able to bind the THAP1 target sequence, this DNA-binding activity was completely eliminated in the presence of non-specific competitor DNA. Together with the observation that distinct THAP domains sequences within a single species exhibit less than 50 % identity between each other (7), this suggests that each THAP-zinc finger may possess its own specific DNA-binding site. This possibility is further supported by the observation that the DNA target sequence of the THAP-zinc finger of THAP1 does not share homology with the AT-rich motif recognized by the THAP-zinc finger of P element transposase (7,45). However, we cannot exclude at this stage that some THAP-zinc fingers may lack sequence specificity or even DNA-binding activity, and may rather function as protein-protein interaction modules.

Finally, protein-protein interactions mediated by other domains of the THAP proteins may be critical to increase the DNA-binding activity of the THAP-zinc finger, which appears to be relatively weak. In this respect, the C-terminal coiled-coil domain found in THAP1, as well as several other human THAP proteins, may enhance the affinity of the full-length protein for DNA by allowing dimerization or multimerization. Future studies will help to resolve these issues and will provide important new insights about the structure and functions of THAP-zinc finger proteins both in humans and model animal organisms.

REFERENCES

1. Klug, A. (1999) J Mol Biol 293(2), 215-218
2. Pavletich, N. P., and Pabo, C. O. (1991) Science 252(5007), 809-817
3. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., and Wright, P. E. (1989) Science 245(4918), 635-637
4. Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., and Gronenborn, A. M. (1993) Science 261(5120), 438-446
5. Schwabe, J. W., Neuhaus, D., and Rhodes, D. (1990) Nature 348(6300), 458-461
6. Roussigne, M., Kossida, S., Lavigne, A. C., Clouaire, T., Ecochard, V., Glories, A., Amalric, F., and Girard, J. P. (2003) Trends Biochem Sci 28(2), 66-69
7. Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F., and Girard, J. P. (2005) Proc Natl Acad Sci U S A 102(19), 6907-6912
8. Roussigne, M., Cayrol, C., Clouaire, T., Amalric, F., and Girard, J. P. (2003) Oncogene 22(16), 2432-2442
9. Hammer, S. E., Strehl, S., and Hagemann, S. (2005) Mol Biol Evol 22(4), 833-844
10. Cayrol, C., Lacroix, C., Mathe, C., Ecochard, V., Ceribelli, M., Loreau, E., Lazar, V., Dessen, P., Mantovani, R., Aguilar, L., and Girard, J. P. (2007) Blood 109(2), 584-594
11. Giangrande, P. H., Zhu, W., Schlisio, S., Sun, X., Mori, S., Gaubatz, S., and Nevins, J. R. (2004) Genes Dev 18(23), 2941-2951
12. Ferguson, E. L., and Horvitz, H. R. (1989) Genetics 123(1), 109-121
13. Thomas, J. H., and Horvitz, H. R. (1999) Development 126(15), 3449-3459
14. Boxem, M., and van den Heuvel, S. (2002) Curr Biol 12(11), 906-911
15. Reddy, K. C., and Villeneuve, A. M. (2004) Cell 118(4), 439-452
16. Chesney, M. A., Kidd, A. R., 3rd, and Kimble, J. (2006) Genetics 172(2), 915-928
17. Lay, D. S., Keenan, S., and Han, M. (2002) Genes Dev 16(4), 503-517
18. Koreth, J., and van den Heuvel, S. (2005) Oncogene 24(17), 2756-2764
19. Lewis, P. W., Beall, E. L., Fleischer, T. C., Georlette, D., Link, A. J., and Botchan, M. R. (2004) Genes Dev 18(23), 2929-2940
20. Korenjanak, M., Taylor-Harding, B., Binne, U. K., Satterlee, J. S., Stevaux, O., Aasland, R., White-Cooper, H., Dyson, N., and Brehm, A. (2004) Cell 119(2), 181-193
21. Litovchick, L., Sadasivam, S., Florens, L., Zhu, X., Swanson, S. K., Velumurugan, S., Chen, R., Washburn, M. P., Liu, X. S., and DeCaprio, J. A. (2007) Mol Cell 26(4), 539-551
22. Harrison, M. M., Ceol, C. J., Lu, X., and Horvitz, H. R. (2006) Proc Natl Acad Sci U S A 103(45), 16782-16787
23. Ceol, C. J., and Horvitz, H. R. (2004) Dev Cell 6(4), 563-576
24. Poulin, G., Dong, Y., Fraser, A. G., Hopper, N. A., and Ahringer, J. (2005) Embo J 24(14), 2613-2623
25. Macfarlan, T., Kutney, S., Altman, B., Montross, R., Yu, J., and Chakravarti, D. (2005) J Biol Chem 280(8), 7346-7358
26. Liew, C. K., Crossley, M., Mackay, J. P., and Nicholas, H. R. (2007) J Mol Biol 366(2), 382-390
27. Sattler, M., Schleucher, J., and Griedinger, C. (1999) Progr. Nucl. Magn. Reson. Spectr. 34, 93-158
28. Cornilescu, G., delaglio, F., and Bax, A. (1999) J. Biomol. NMR 13, 289-302
29. Ponstingl, H., and Otting, G. (1998) J Biomol NMR 12(2), 319-324
30. Wüthrich, K. (1986) NMR of proteins and nucleic acids, Wiley Press
31. Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002) J Biomol NMR 24(2), 149-154
32. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) Biochem. Biophys. res. Commun. 117, 479-485
33. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) J Biomol NMR 6(3), 277-293
34. Bartels, C., Xia, T.-H., Billeter, M., Güntert, M., and Wüthrich, K. (1995) J. Biomol. NMR 5, 1-10
35. Johnson, B. A. (2004) Methods Mol Biol 278, 313-352
36. Auguin, D., Barthe, P., Auge-Senegro's, M. T., Stern, M. H., Noguchi, M., and Roumestand, C. (2004) J Biomol NMR 28(2), 137-155
37. Daikun, G. P., Fairall, L., and Klug, A. (1986) Nature 324, 688-699
38. Brünger, 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 D Biol Crystallogr 54(Pt 5), 905-921
39. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) J. Biomol. NMR. 8, 477-486
40. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) Journal of the American Chemical Society 125(7), 1731-1737
41. Lee, C. C., Beall, E. L., and Rio, D. C. (1998) The EMBO J. 17, 4166-4174
42. Jauch, R., Bourenkov, G. P., Chung, H. R., Urlaub, H., Reidt, U., Jackle, H., and Wahl, M. C. (2003) Structure 11(11), 1393-1402
43. Laity, J. H., Lee, B. M., and Wright, P. E. (2001) Curr Opin Struct Biol 11(1), 39-46
44. Garvie, C. W., and Wolberger, C. (2001) Mol Cell 8(5), 937-946
45. Kaufman, P. D., Doll, R. F., and Rio, D. C. (1989) Cell 59(2), 359-371

FOOTNOTES

Acknowledgements- We are grateful to E. Guittet at the ICSN Institute, Gif/Yvette for providing access to the 800-MHz national facility. We wish to acknowledge O. Saurel and P. Ramos at the IPBS, Toulouse for NMR technical assistance and J.P ESTÈVe at the CHU Rangueil, Toulouse for help with SPR experiments. The 600 MHz cryoprobe and 700 MHz IPBS spectrometers were acquired within the CPER 2000-2006 involving CNRS, University of Toulouse, Région Midi-Pyrénées and European Structural Funds (FEDER). This work was supported by the French Research Ministry (ACI BCMS), Ligue Nationale Contre le Cancer (Equipe labellisée, to J.P.G) and by FRM (to D.B.) and ARC (to C.L.).

Data Deposition - The atomic coordinates for the structures of the THAP domain (Met1-Phe81) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (code 2jtg). Chemical Shift informations for the THAP domains {Met1-Lys90} and {Met1-Phe81} are available from the BioMagRes Data Bank (http://www.bmrb.wisc.edu), under the accession numbers 15300 and 15289, respectively.

Supplementary Information is available at The Journal of Biological Chemistry Online (http://www.jbc.org).

FIGURE LEGENDS

Fig. 1. Specific THABS-DNA binding of the THAP domain of THAP1 as observed by SPR and EMSA. A. SPR difference sensograms with increasing concentrations of protein from 50 nM to 45 µM as indicated on the sensogram lines (responses with the unrelated DNA sequence were subtracted). B. Kd determination. Data points represent the equilibrium responses values as function of the protein concentration for each of the experiments shown in A. The global binding constant obtained by fitting the SPR data to a 1:1 stoichiometry-binding model was found to be 8 µM. The solid line represents the fitting theoretical curve calculated for the 1:1 binding model. C. The first 81 residues of the THAP-zinc finger of THAP1 are sufficient for sequence-specific DNA-binding activity. EMSA experiments were performed using a consensus 25-bp THABS probe and increasing amounts (5, 10 and 100 nM) of recombinant THAP-zinc finger Met1-Lys90 or Met1-Phe81 of human THAP1. Black arrows, THAP1-THABS DNA complex.

Fig. 2. Solution structure of the THAP-zinc finger of human THAP1. A. Backbone traces of the NMR ensemble for the 20 lowest energy structures in stereo view. The zinc and the four ligands are shown in orange. B. Topology diagram showing the secondary structure elements. The α-helix, the 3₁₀ helices and the β-sheet are depicted in cyan, blue and purple, respectively. The zinc and the four...
ligands are shown in orange. C. A spine representation of the lowest energy structure (residues 4-81) with variable radius, the radius representing the mobility as judged by heteronuclear NOEs (Supplementary Fig. 1). The orientation is the same as in A. D. Ribbon diagram of the THAP-zinc finger of THAP1 showing side chains of the four invariant residues in red. The left side is in the same orientation as in A. The right side is rotated 90° around horizontal axis. E. Structure based sequence alignment of THAP-zinc fingers from human (hTHAP1-3), zebrafish (ze2F6), Drosophila melanogaster (dmTRP) and C. elegans (Ce-CtBP) members of the THAP family. The secondary structure elements are represented with the same color as in 1B. The zinc ion together with the four zinc ligands is depicted in orange. Highly conserved hydrophobic residues are shown in green. The four invariant residues (Pro26, Trp36, Phe58 and Pro78) are shown in red. The conserved basic residues are shown in blue.

**Fig. 3.** The THAP-zinc fingers share a similar fold but do not recognize the same DNA target sequence. A. Ribbon diagrams of the THAP domains of THAP1, THAP2 and CtBP. B. Recombinant THAP-zinc fingers from human THAP2 and THAP3 do not bind to the THABS. EMSA experiments were performed by incubating the 25-bp THABS probe in the presence of poly(dI/dC) with increased amounts (0.01, 0.1, 1 and 5 µM) of recombinant THAP domains from THAP2 and THAP3 produced in *E. coli*. As control, EMSA was performed with THABS probe and the recombinant THAP-zinc finger of THAP1 (0.1 µM) in the presence of poly(dI/dC). C. The THAP-zinc fingers of Ce-CtBP and GON-14 bind DNA but do not interact specifically with the THABS motif recognized by human THAP1. Increasing amounts (0.1, 1, 5 and 10 µM) of recombinant THAP-zinc fingers of Ce-CtBP or GON-14 were incubated with the THABS probe in the presence or absence of the non-specific competitor poly(dI/dC) and analyzed by EMSA. As a control, EMSA was performed with the recombinant THAP-zinc finger of THAP1 (0.1 µM) in the presence of poly(dI/dC).

**Fig. 4.** Site-directed mutagenesis of the THAP-zinc finger of THAP1 reveals critical residues for DNA-binding activity. A-D. Alanine scanning mutagenesis. A,C. Wild-type THAP1 (wt) and single point or triple mutants in the THAP-zinc finger were *in vitro* translated in rabbit reticulocyte lysate in the presence of 35S labeled methionine and analyzed by SDS-PAGE and autoradiography. Molecular weight markers are shown on the left (kDa). B,D. Single point mutation of residues Lys24, Arg29, Arg42, Phe45 or Thr48 abrogates DNA-binding activity of the THAP-zinc finger of THAP1. EMSA were performed with the 36-bp THABS probe and THAP1 wild-type (wt) or the indicated mutant proteins, in the presence of poly(dI/dC) and salmon sperm DNA competitors. The previously described Trp36A mutant (7) was included as a control for loss of DNA-binding activity. Wt + Ab, super shift experiment with anti-THAP1 antibody to demonstrate the specificity of the protein-DNA complexes. RRL, unprogrammed rabbit reticulocyte lysate; black arrowhead, THAP1-THABS DNA complex; white arrowhead, antibody-THAP1-THABS DNA complex; asterisks, non-specific complexes. E. A view illustrating the interactions of Arg29 and Phe45 (in red) with Leu32, Trp36, Ala76 and Phe25, Val40, respectively (in green). F. Representations of the electrostatic surface potential of the THAP domain {Met1-Phe81} showing the exposed residues that are found to be critical for DNA binding from site-directed mutagenesis experiments (underlined) or that undergo more than 0.2 ppm chemical shift change (marked with an asterisk). Exposed residues that are positively charged are indicated on the surface and colored yellow, otherwise black. Representations of the two corresponding ribbons are shown for clarity (in gray).

**Fig. 5.** Chemical shift perturbation analysis of the THAP-zinc finger of THAP1 upon DNA binding. A. Overlay of selected regions of 1H-15N HSQC spectra of the THAP domain in the absence (green) and in the presence (red contour levels) of an equimolar concentration of the 14-mer oligonucleotide (THABS). Two examples of residues affected (Thr28) and not (Asp15) are represented in A. and B., respectively with additional blue contour levels for half an equimolar concentration of DNA (blue contour levels). Residue Thr48 that displays the largest chemical shift change is also indicated. B. Histogram of chemical shift changes upon DNA binding as a function of
the THAP-zinc finger residue number. Reported chemical shifts Δδ represent combined $^{15}$N and $^1$H chemical shifts ($\Delta \delta = [(\Delta \delta_{HN})^2 + (\Delta \delta_N \times 0.154)^2]^{1/2}$). Secondary structure elements are depicted with the same color as in Fig. 2.

Fig. 6. A proposed model of the complex between the THAP-zinc finger of THAP1 (gray, pink, yellow) and its DNA target (green, blue) in stereo view. Amino acids with backbone chemical shifts that undergo more than 0.2 ppm chemical shift change upon DNA binding are colored pink. Side-chains of residues that are found critical from site-directed mutagenesis experiments are depicted in yellow. The side-chain of K46 is colored in pink. The THABS DNA molecule is colored green except bases at the GGCA core motif shown in blue. The THABS molecule was docked onto the THAP-zinc finger of THAP1 using HADDOCK1.3 (High Ambiguity Driven Protein Protein Docking) (40).
Table 1. Structural Statistics of the THAP-zinc finger {Met1-Phe81} of THAP1

| Restraints for calculation                  |     |
|--------------------------------------------|-----|
| Intraresidue                               | 801 |
| Sequential                                 | 329 |
| Medium-range (2≤ | i-j | ≤4)    | 103 |
| Long range (| i-| j | >4)    | 306 |
| Dihedral angle restraints (TALOS)          |     |
| angle φ                                    | 52  |
| angle ψ                                    | 52  |
| Hydrogen bond restraints                   | 10  |
| Zinc coordination                          | 14  |

Characteristics

| R.m.s deviation from constraints            |     |
|--------------------------------------------|-----|
| NOE restraints (Å)                          | 0.0348 ± 0.0004 |
| Dihedral angle restraints (°)               | 0.27 ± 0.02  |

| R.m.s deviation from idealized geometry (± s.d.) |     |
|------------------------------------------------|-----|
| Bond length (Å)                                | 0.0048 ± 0.0005 |
| Bond angle (°)                                 | 0.60 ± 0.05    |
| Improper angle (°)                             | 0.39 ± 0.07    |

| Final energies (kcal mol⁻¹ ± s.d.)             |     |
|------------------------------------------------|-----|
| Overall                                       | 477.6 ± 6.7 |
| van der Waals                                 | 96.44 ± 3.3 |
| Bonds                                         | 33.3 ± 0.7  |
| Angles                                        | 144.2 ± 2.5  |
| NOE                                           | 185.33 ± 4.2 |
| Dihedrals                                     | 0.61 ± 0.1  |

Coordinate precision

| R.m.s deviation of backbone atoms (Å)          | 0.530 |
| R.m.s deviation of all heavy atoms (Å)        | 1.297 |

Ramachandran plot

| Residues in most favored region (%)           | 64.0 |
| Residues in additional allowed regions (%)   | 32.2 |
| Residues in generously allowed regions (%)   | 3.8  |
| Residues in disallowed regions (%)           | 0.0  |

*a* Average r.m.s deviation from the mean structure (residues 4-16, 21-40, 46-63, 69-81)

*b* The φ and ψ dihedral angles were analyzed using the PROCHECK program (39)
Table 2. Alanine scanning mutagenesis of the THAP-zinc finger of human THAP1.
DNA-binding activity of the different mutants was tested using gel-shift assays. The eight evolutionary conserved residues, which define the THAP-zinc finger and have previously been shown to be essential for DNA binding (7) are indicated in bold. *Within the limits of detection of the present assay; this qualitative classification does not imply that the corresponding residues play no role at all in the binding affinity and selectivity.

| DNA binding to THABS probe | THAP1 mutants |
|---------------------------|---------------|
| Severely affected          | C5A, C10A, C54A, H57A, P26A, W36A, F58A, P78A, K24A, R29A, R42A, F45A, T48A |
|                           | *Triple mutants: T28A-R29A-P30A, R41A-R42A-K43A, P47A-T48A-K49A, Y50A-S51A-S52A |
| Partially affected         | K11A, L27A, E37A, V40A, Y50A |
| Not affected*              | S4A, S6A, Y8A, K16A, T28A, P30A, S31A, L32A, C33A, K34A, E35A, R41A, K43A, N44A, K46A, P47A, K49A, S51A, S52A, S55A |
Fig. 1.

A

![Graph showing response (RU) over time (s) for different protein concentrations.](image)

B

![Graph showing equilibrium response (RU) against protein concentration (µM).](image)

C

![Image showing gel with bands for THAP1 Met1-Lys90 and THAP1 Met1-Phe81.](image)
Fig. 3.

A

THAP1

THAP2

Ce-CtBP

B

|       | THAP2 THAP domain | THAP1 THAP domain | THAP3 THAP domain | THAP1 THAP domain |
|-------|-------------------|-------------------|-------------------|-------------------|
|       | + polydl/dC       | + polydl/dC       | + polydl/dC       | + polydl/dC       |

C

|       | Ce-CtBP THAP domain | Ce-CtBP THAP domain | GON-14 THAP domain | GON-14 THAP domain |
|-------|---------------------|---------------------|--------------------|--------------------|
|       | 0 polydl/dC         | + polydl/dC         | 0 polydl/dC        | + polydl/dC        |

Free THABS

THABS
Structure-function analysis of the thap-zinc finger of thap1, a large C2CH DNA-binding module linked to RB/E2F pathways

Damien Bessiere, Chrystelle Lacroix, Sébastien Campagne, Vincent Ecochard, Valérie Guillet, Lionel Mourey, Frédéric Lopez, Jerzy Czaplicki, Pascal Demange, Alain Milon, Jean-Philippe Girard and Virginie Gervais

J. Biol. Chem. published online December 11, 2007

Access the most updated version of this article at doi: 10.1074/jbc.M707537200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2007/12/12/M707537200.DC1