J Biomol NMR (2013) 56:3–15
DOI 10.1007/s10858-012-9699-1

ARTICLE

NMR studies of a new family of DNA binding proteins: the THAP proteins

Virginie Gervais · Sébastien Campagne · Jade Durand · Isabelle Muller · Alain Milon

Received: 30 October 2012 / Accepted: 20 December 2012 / Published online: 11 January 2013

Abstract The THAP (THanatos-Associated Protein) domain is an evolutionary conserved C2CH zinc-coordinating domain shared with a large family of cellular factors (THAP proteins). Many members of the THAP family act as transcription factors that control cell proliferation, cell cycle progression, angiogenesis, apoptosis and epigenetic gene silencing. They recognize specific DNA sequences in the promoters of target genes and subsequently recruit effector proteins. Recent structural and functional studies have allowed getting better insight into the nuclear and cellular functions of some THAP members and the molecular mechanisms by which they recognize DNA. The present article reviews recent advances in the knowledge of the THAP domains structures and their interaction with DNA, with a particular focus on NMR. It provides the solution structure of the THAP domain of THAP11, a recently characterized human THAP protein with important functions in transcription and cell growth in colon cancer.

Keywords DNA binding domain · THAP · NMR · Protein–DNA complex · Loop dynamics · Transcription factor · Isothermal titration calorimetry ITC · Differential scanning fluorimetry DSF · DNA binding affinity

The THAP domain: a new class of zinc-finger DNA binding domain involved in human pathologies

At the beginning of the 2000's, a new family of nuclear proteins was discovered and called the THAP proteins (THanatos-Associated Protein) (Roussigne et al. 2003a, b). All of them contain in their sequence a ~80 residues module which defines an evolutionary conserved zinc-coordinating domain, the THAP domain. More than 200 THAP proteins have been identified across animal species and the human genome encodes for twelve THAP proteins (THAP0-11), which vary in length between 213 and 903 residues (Fig. 1a). The THAP signature, exclusively found at the N-terminus of the human proteins as a single copy, consists of a C2CH module providing ligands for zinc coordination, four invariant hydrophobic residues and a C-terminus AVPTIF motif relatively well conserved within the THAP family (Clouaire et al. 2005). The consensus THAP signature has also been identified in the sequence-specific DNA binding domain of *Drosophila* P element transposase, suggesting that the THAP domain constitutes a novel example of a DNA binding domain shared with cellular proteins and transposons from mobile genomic parasites (Roussigne et al. 2003b).

Therefore, it was demonstrated that the THAP domain of human THAP1, the founding member of the family,

Electronic supplementary material The online version of this article (doi:10.1007/s10858-012-9699-1) contains supplementary material, which is available to authorized users.

V. Gervais · J. Durand · I. Muller · A. Milon
CNRS, IPBS (Institut de Pharmacologie et de Biologie Structurale), 205 route de Narbonne, BP64182, 31077 Toulouse, France

V. Gervais (✉) · J. Durand · I. Muller · A. Milon (✉)
IPBS, Université de Toulouse-UPS, 31077 Toulouse, France
e-mail: virginie.gervais@ipbs.fr

A. Milon
e-mail: alain.milon@ipbs.fr

S. Campagne
ETH Zurich, HCI F437, Wofgang Pauli Strasse, 8093 Zurich, Switzerland

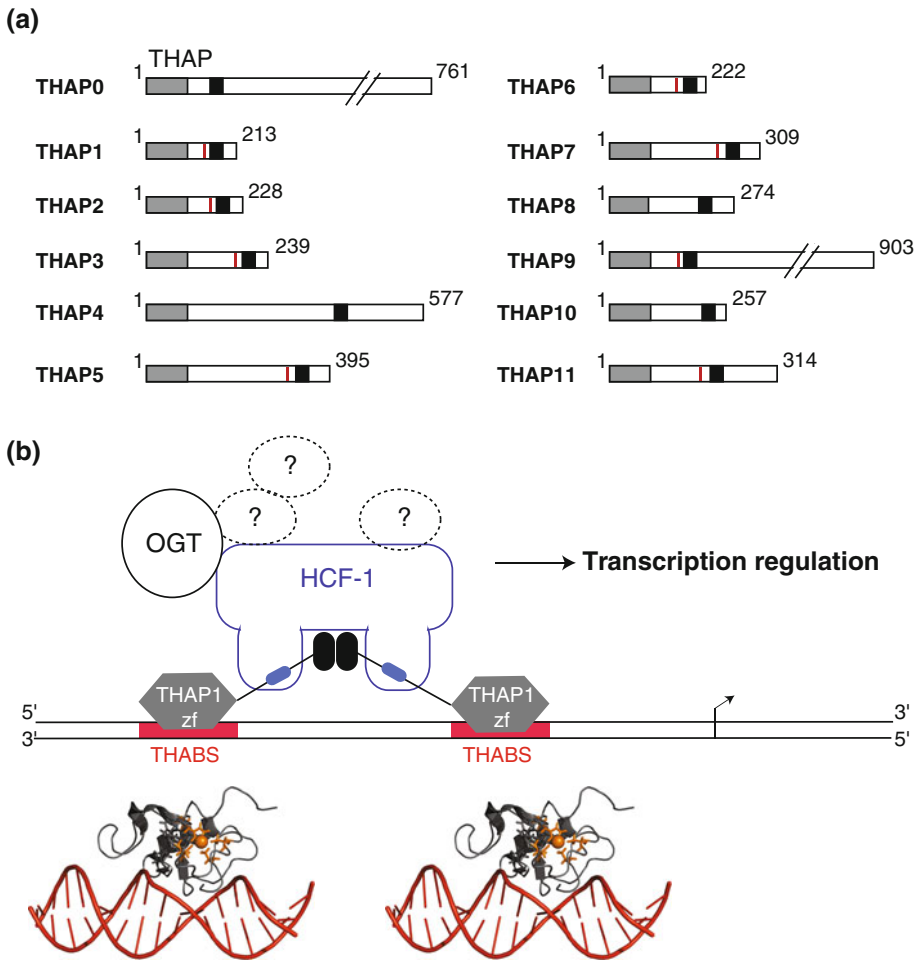

Fig. 1 The human THAP protein family and their transcription regulation mode. **a** Schematic representation of the human THAP protein family. Boundaries of each protein are labelled and a *grey box* represents the THAP domain, a *red box* represents the HCF-1 binding motif and a *black box* represents the coiled coil domain. **b** Schematic drawing of THAP1 action mode. The THABS sequences present in the *RRM1* promoter (Cayrol et al. 2007) are represented by *red boxes*, THAP1 is shown with the same colour code as in **a**. Human THAP1 recruits HCF-1 through interaction between HCF-1 and the HBM motif (*blue cylinder*) of human THAP1. Then, HCF-1 binds to several partners, including OGT, and this complex promotes transcription regulation of the targeted gene. The binding mode of the THAP domain onto the THABS motif is shown with the zinc binding coordination site in *orange*

exhibits specific DNA binding to an 11-nucleotides sequence coined THABS (THAP1 binding sequence) and that this DNA binding activiy is zinc-dependent (Clouaire et al. 2005). Other THAP domains from human THAP2 and THAP3 proteins or *Caenorhabditis elegans* GON-14 failed to exhibit sequence-specific DNA binding activity toward the THABS sequence suggesting that the various THAP domains possess their own specificity (Bessiere et al. 2008).

The THAP domains drive the THAP proteins to their DNA targets for the formation of multi-protein complexes and the transcription regulation of gene subsets (Fig. 1b). In their C-terminal parts, human THAP proteins contain a nuclear localization sequence, a coiled coil domain, that was shown to be implicated in protein dimerization (Sengel et al. 2011) and, just upstream, a small motif (Vxx(D/E)H(S/N)YxV), called HBM for HCF-1 Binding Motif. The HBM motif allows several THAP proteins to fish HCF-1, a potent transcriptional coactivator and cell cycle regulator which acts as a platform for the recruitment of enzymatic activities modifying the chromatin structure, and thus promoting transcription regulation (Mazars et al. 2010; Dejosez et al. 2008; Parker et al. 2012). As an example, the human THAP3 protein takes part into a 0.6 MDa multi-protein complex including HCF-1 and the O-linked N-acetylglucosamine transferase (OGT), and several peptides of HCF-1 were found to be O-glycosylated (Mazars et al. 2010). Because the THAP proteins were shown to activate or repress (or both) the transcription of their target genes, it has been proposed that the post-translational modifications of HCF-1 might be responsible for the transcription effect.

The THAP-containing proteins act as transcriptional regulators linked to cell proliferation, cell cycle progression, maintenance of pluripotency, angiogenesis, apoptosis and epigenetic gene silencing (Roussigne et al. 2003a; Macfarlan et al. 2005; Cayrol et al. 2007; Dejosez et al. 2008; Balakrishnan et al. 2009). Within the human THAP family, THAP1 and THAP11 are the most characterized proteins at the functional level. The THAP1 protein interacts with the pRb/E2F pathway and promotes transcription activation of several genes coding for essential proteins required for DNA synthesis during S phase of the cell cycle. It binds to two THABS sequences on the

promoter of the *RRM1* gene, an essential gene for DNA synthesis, and activates transcription via the recruitment of protein partners (Cayrol et al. 2007; Mazars et al. 2010). The THAP11 protein, the most recently identified member of the family has been first characterized as a transcription factor involved in regulation of embryogenesis and pluripotency in mouse embryonic stem (ES) cells (Dejosez et al. 2008; Dejosez et al. 2010) and controls the expression of a gene subset responsible for the unimpeded growth of ES cells (Dejosez et al. 2010). In addition, THAP11 was shown to directly control c-Myc gene expression and function as a cell growth suppressor by negatively regulating the expression of c-Myc (Zhu et al. 2009).

Altogether, these results support a major role for the human THAP proteins in cell cycle control and proliferation. Moreover, deregulations in functions performed by human THAP proteins have been associated with severe human diseases such as dystonia (Bragg et al. 2011), heart disease (Balakrishnan et al. 2009) and several types of cancer (Parker et al. 2012; De Souza et al. 2008; Lian et al. 2012). Notably, in 2009, a genetic link between the THAP1 gene and a hereditary disease called primary dystonia DYT6 has been discovered (Fuchs et al. 2009; Bressman et al. 2009; Djarmati et al. 2009). Primary torsion dystonias refer to a variety of movement disorders that are associated with dysfunction in central nervous system regions controlling movement [for review (Muller 2009)]. Many mutations, including frame-shift mutations and single-point mutations have been identified in the *THAP1* gene of DYT6 patients (Bonetti et al. 2009; Xiao 2010; Sohn et al. 2010; Song et al. 2011; Lohmann et al. 2012). Mutations are found in regions encoding for functional parts of the THAP1 protein and half of them target the THAP DNA binding domain, supporting that transcription deregulation mediated by the THAP1 mutations might be responsible for the DYT6 disease apparition. Moreover, THAP1 downregulates the expression of the *TOR1A* gene, a gene linked to another primary dystonia DYT1 (Gavarini et al. 2010; Kaiser et al. 2010). A detailed biophysical characterization of several DYT6 mutations has shown that the major cause for the loss of THAP1 function is not an abrogation of DNA binding activity but should rather be associated with a decrease of protein stability (Campagne et al. 2012). Other THAP members have been linked to several types of cancers. In particular, THAP11 acts as a tumour-promoting factor in animal model of teratocarcinoma (Dejosez et al. 2008). Moreover, upregulation and downregulation effects have been reported in various cell cancer types (Zhu et al. 2009; Nakamura et al. 2012; Parker et al. 2012), suggesting that the THAP11 function is context dependent and subtle modification of its activity can be detrimental. Following these lines, the importance of THAP11/c-Myc pathway in chronic myologenous leukemia progenitor cell proliferation has recently been illustrated (Nakamura et al. 2012).

Because of their implications in the development of neurological diseases and cancers, human THAP proteins appear as potential pharmacological targets, and insights into the structure/function relationships are necessary to provide a global view of the regulation processes they mediate. We, and others have contributed to a better understanding of the THAP domains structures, both alone and in complex with their DNA targets, as discussed below.

Three-dimensional structures of the THAP family members

Zinc fingers are small, functional and independently folded domains that require coordination of a zinc atom to stabilize their structure (Laity et al. 2001). The THAP domain defines an atypical zinc finger motif, characterized by a large C2CH module with a spacing of up to 53 amino acids between the zinc-coordinating C2 and CH residues (Clouaire et al. 2005). In addition to this feature, a C-terminal extension including a relatively conserved AVPTIF motif is crucial to maintain the fold of the THAP zinc finger. Indeed, expression of a truncated form of the THAP domain of human THAP1 deleted for the AVPTIF motif led to an unfolded protein (Bessiere et al. 2008). As shown in Fig. 2a, several amino acids are strictly conserved among human THAP proteins including the C2CH zinc coordinating residues, a proline residue in the beginning of loop L2, a tryptophan at the centre of helix H1, a phenylalanine following the last histidine zinc coordinating residue and the proline of the C-terminal AVPTIF motif. The integrity of those amino acids defining the THAP signature is crucial to maintain the specific DNA binding activity of THAP1 (Clouaire et al. 2005).

In order to get insights into the structural features of the THAP zinc finger, the solution structure of the THAP domain of human THAP1 (Met^{1}-Phe^{81}) was solved using multidimensional NMR, on the basis of 1,539 distance restraints and 104 angle restraints (Bessiere et al. 2008). The THAP zinc finger folds into a globular βαβ fold structure stabilized by the C2CH tetrahedral zinc coordination (Fig. 2b) and a network of hydrophobic interactions implying strictly conserved amino acids of the THAP signature. Furthermore, the conserved residues defining the THAP signature adopt equivalent positions in the known structures of other THAP domains and all the conserved hydrophobic residues interact together to ensure the packing of the edifice. Although the THAP signature was first described as a 90-residue long domain based on sequence conservation, our NMR data helped to redefine the minimal boundaries of the THAP domain structural motif (Bessiere

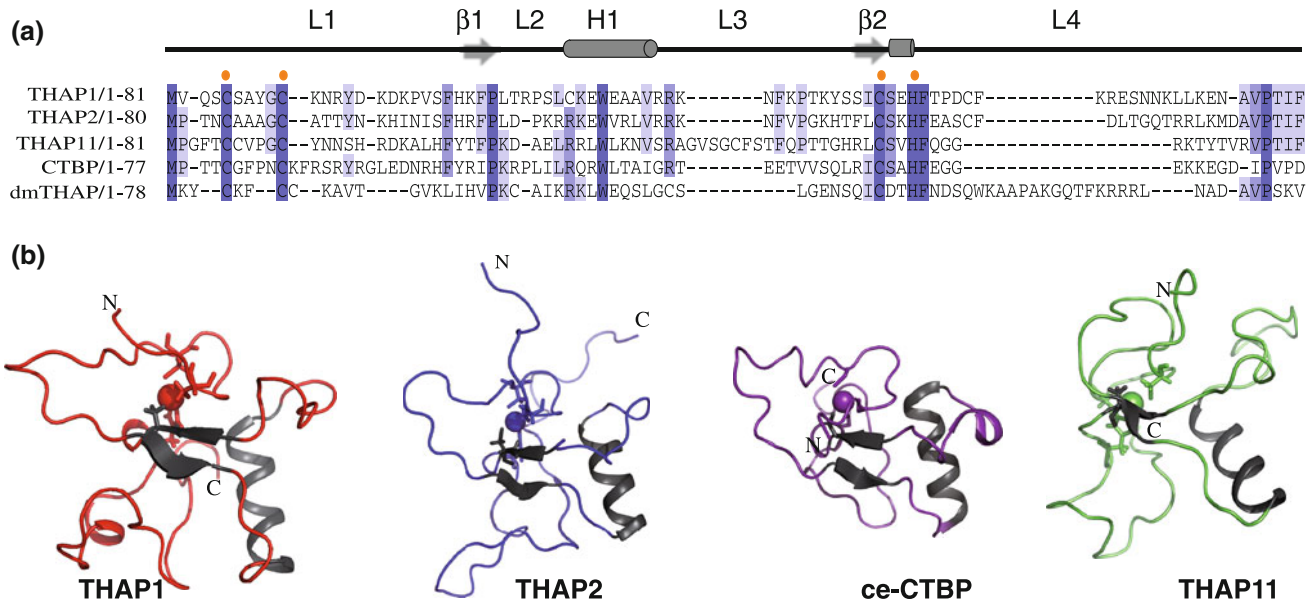

Fig. 2 Structure of the THAP signature. **a** Structure-guided sequence alignment of THAP members with known 3D structures. On the top of the alignment, secondary structures shared with the THAP domains are indicated. Conserved residues are highlighted. *Orange dots* indicate the C2CH zinc coordinating residues. **b** Ribbon representations of the structures of THAP1 in *red* (Bessiere et al. 2008) (PDB *id* 2jtg), THAP2 in *blue* (PDB *id* 2d8r), Ce-CtBP in magenta (Liew et al. 2007) (PDB *id* 2jm3) and THAP11 in *green* (PDB *id* 2lau)

et al. 2008). The dynamic behaviour of the THAP zinc finger domain was studied using ^{15}N relaxation experiments. The core of the protein appears to be rigid due to interactions between the structural elements, while some loop regions (such as the beginning of loop L4) expand rapid motions (Bessiere et al. 2008).

Like most of the described transcription factors, the THAP proteins are modular; they display in their N-terminus a DNA binding domain (THAP domain) and their C-terminal region acts as transactivation domain involved in the recruitment of protein partners. However, little information is available so far concerning the atomic resolution structures of full-length THAP proteins. Attempts in our laboratory to produce and crystallise full length THAP1 proteins failed to provide high quality crystals (L. Mourey et al., personal communication), while attempts to express stable isotope labelled C-terminus domains for NMR studies failed to provide soluble proteins.

The THAP zinc finger domain promotes the specific double-stranded DNA binding activity; however, until recently nothing was known on the molecular mechanisms responsible for the specific recognition. The THABS motif recognized by THAP1 has been determined by the PCR-based SELEX approach (Bouvet 2001). It involves a GGCA core motif (position 7–10) and an AT base pair at position 6 (Clouaire et al. 2005), Fig. 3. Later, target sites of other THAP members, including THAP9 (Sabogal et al. 2010) and Ronin (the mouse homolog of THAP11) (Dejosez et al. 2008) were determined using the SELEX approach. The 20-bp Ronin-binding sequence identified by Dejosez et al. appears distinct in sequence composition and length from the common core of target DNA motif proposed for THAP1, THAP9 and *Drosophila* transposase (Sabogal et al. 2010). Then a second AT-rich Ronin-binding motif has been determined using chromatin-immuno-precipitation and DNA sequencing approaches (Dejosez et al. 2010). However, this 15-bp DNA sequence is poorly correlated with the previously SELEX-derived motif.

Structural and thermodynamic insights into the DNA recognition

So far, most of our knowledge (at atomic resolution) on the molecular mechanism by which the THAP domain recognizes DNA comes from the structural studies performed on two THAP members, human THAP1 and the *Drosophila* P-element transposase in complex with their DNA targets (Sabogal et al. 2010; Campagne et al. 2010). Previous mutagenesis and NMR experiments performed on the THAP domain of THAP1 with a THABS sequence have shown the importance of several residues in the DNA recognition (Bessiere et al. 2008). These data suggested that the β-sheet together with loops played crucial role in DNA recognition and have excluded the helix as a main recognition element, implying an original DNA recognition mode (Bessiere et al. 2008). The solution structure determination of the THAP1-DNA complex required a substantial improvement of the quality of the sample (Campagne et al. 2010), prior to engaging 3D NMR experiments. The impact of several

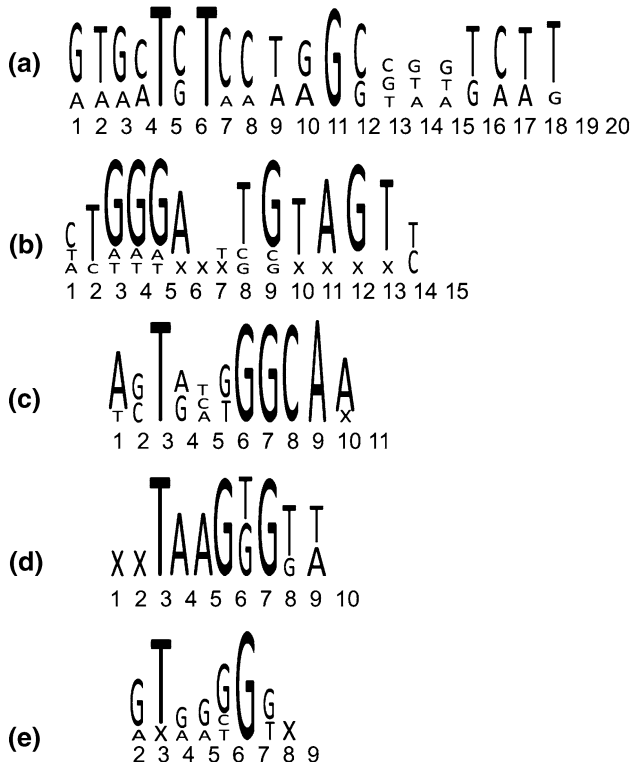

Fig. 3 Consensus binding sites by the THAP domains reported so far. **a** Ronin (mouse homolog of THAP11) -binding sequence derived from SELEX (Dejosez et al. 2008). **b** Ronin-binding motif determined using ChIP and ChIP-seq (Dejosez et al. 2010). **c** SELEX-derived THABS sequence recognized by human THAP1 (Clouaire et al. 2005). **d** Consensus sequence recognized by the *Drosophila* THAP domain (Sabogal et al. 2010) **e** SELEX-derived sequence recognized by human THAP9 (Sabogal et al. 2010)

parameters (free cysteine mutagenesis, salt concentration and length of the oligonucleotide target) onto the sample was addressed by recording 2D HSQC spectra (Campagne et al. 2010). The improvement of the spectra allowed assigning unambiguously the residues that exhibit chemical changes upon DNA binding. The solution structure of the complex formed between the THAP domain and its 16-bp DNA target, containing the THABS sequence, was determined on the basis of 3,011 experimental restraints for both protein and DNA and 39 intermolecular NOEs (Campagne et al. 2010). Binding to specific DNA is accompanied by several structural changes in the protein, affecting more particularly loops L3 and L4. As shown in Fig. 4a–b, the THAP domain uses its central β-sheet as well as several residues within loops to recognize the DNA major groove while a C-terminal basic loop targets the DNA minor groove (Campagne et al. 2010). The same year, Sabogal et al. have determined a closely related binding model for the association of the THAP domain from *Drosophila melanogaster* P element transposase (dmTHAP) with its 10-bp DNA target, Fig. 4c (Sabogal et al. 2010). These structural data point out that recognition specificity is partly insured thanks to a combination of unique protein side-chains that contact invariant DNA bases of each consensus sequence (Fig. 4d). The determinant role of some residues involved in the direct readout of the sequence was confirmed by mutagenesis studies. Dramatic decreases in binding affinity of dmTHAP (by factors of 12–20 times) towards its specific target were observed for several mutants (Sabogal et al. 2010). Similarly, in THAP1, the K24A mutant loses its ability to bind to its target DNA whereas it retains its ability to bind to a non-specific DNA sequence, as monitored by fluorescence anisotropy (Campagne et al. 2010). Moreover, a number of contacts between protein and nearby DNA sugar phosphates position the protein onto DNA and increase molecular recognition (Campagne et al. 2010). Our NMR studies together with the work performed by Sabogal et al. independently revealed that two THAP domains, despite belonging to distinct species adopt a common DNA-targeting mechanism to recognize specific DNA targets. More generally, these lines suggest that the sequence variability observed for some side-chain residues in the β-sheet could accommodate different DNA sequences in the major groove (Sabogal et al. 2010; Campagne et al. 2010). Remarkably, the two THAP domains use a C-terminal basic loop (L4) located between the second β-strand and the AVPTIF motif to target DNA minor groove (Sabogal et al. 2010; Campagne et al. 2010). The length of this loop varies among the THAP domains (for instance between THAP1 and THAP11, as will be discussed below). This might impact the binding affinity and/or the binding specificity, and could even change the recognition mode of the DNA target sequence.

Specific DNA recognition by the THAP domain of THAP1 is thermodynamically favoured and both enthalpy ($\Delta H < 0$) and entropy ($-T\Delta S < 0$) driven (Campagne et al. 2012). The large negative binding enthalpy is representative of THAP1 binding in the major-groove and the formation of specific hydrogen bonds between side-chain residues and DNA bases. Upon DNA binding, the burial of hydrophobic residues and the release of water molecules from DNA major and minor-grooves binding lead to an increase of entropy (Privalov et al. 2007, 2011). Using several biophysical techniques, it was shown that the THAP domain binds to its specific DNA target with at least a 12 times better affinity than a non-specific DNA (Campagne et al. 2012). The binding mode is exclusively entropic when THAP1 domain associates with a non-specific DNA sequence, indicating that specificity is provided by hydrogen bonds with tightly associated residues (Campagne et al. 2012; Privalov et al. 2011). The analysis of the chemical shift changes of the backbone amide nitrogen resonances of THAP1 upon DNA binding has shown that the same regions are affected when the protein is in complex with a specific or a non-specific DNA

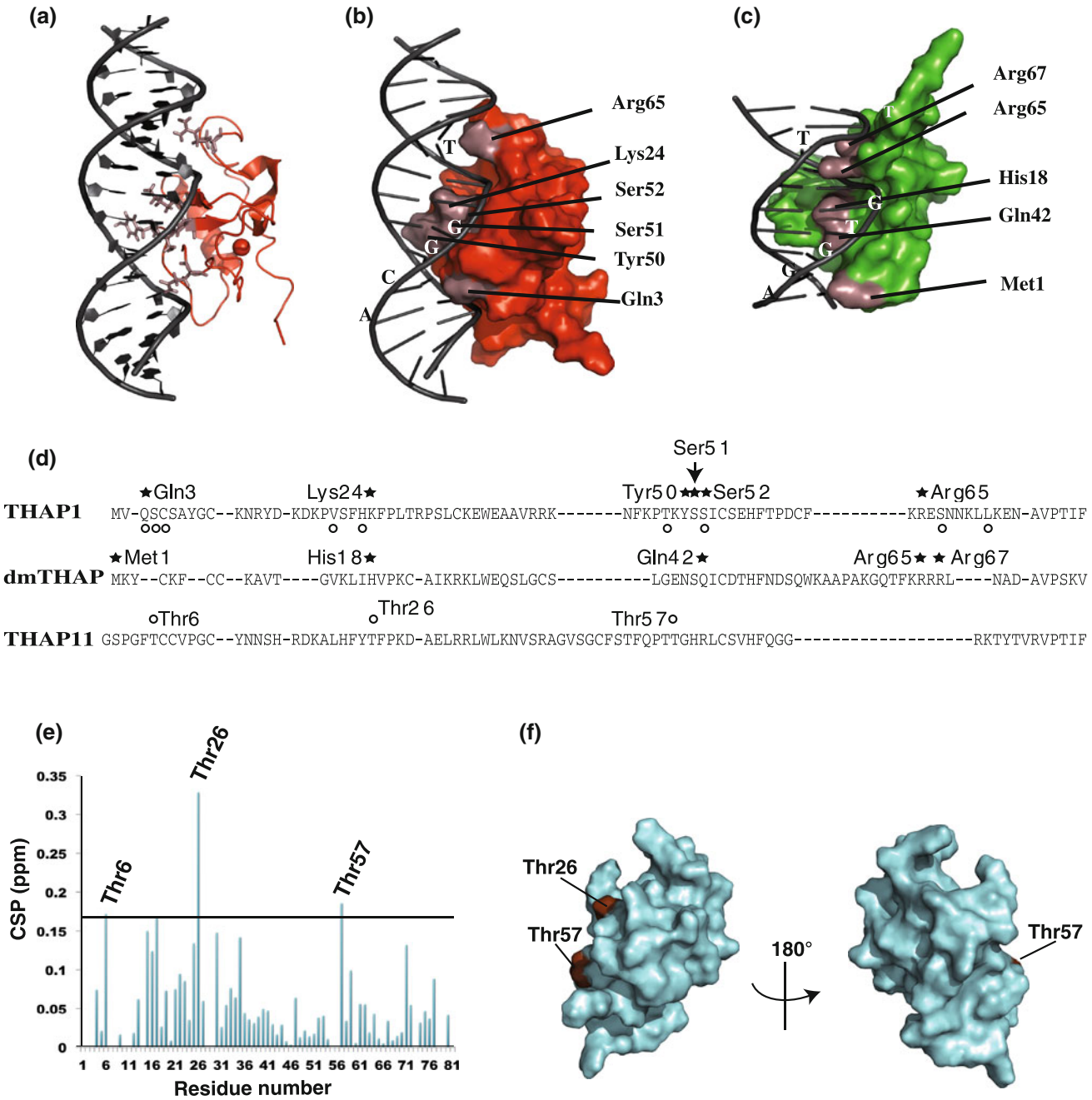

Fig. 4 DNA recognition by the THAP domains. **a** Protein-DNA interface illustrating the base-specific contacts in the major and minor grooves between THAP1 and its 16-bp DNA target (Campagne et al. 2010). The THAP domain is shown as a ribbon diagram *in red*; residues that make contacts with DNA are shown as *brown sticks*. The zinc is shown as a *red sphere*. **b** Surface representation of the THAP domain of THAP1 (*in red*) with the same orientation as in **a**. Consensus bases of the THABS motif are indicated on the DNA backbone. The DNA interacting residues are indicated on the surface and coloured *in brown*. **c** Surface representation of the dmTHAP (PDB *id* 3kde) *in green* (Sabogal et al. 2010). The orientation of the protein surface is identical to that shown in **a**–**b**. The numbering refers to that used for each THAP domain. **d** Sequence alignment for THAP1, dm-THAP and THAP11. *Filled asterisks* indicate the DNA-binding residues of THAP1 and dm-THAP, according to 3D structures. The residues of THAP1 (Campagne et al. 2010) and THAP11 with the largest chemical shift changes are indicated with open circles. **e** Chemical shift changes observed for the THAP domain of human THAP11 in the presence of a 20-bp DNA sequence, corresponding to the Ronin-binding sequence present in the c-Myc promoter and assumed to be the specific target (Zhu et al. 2009). **f** Surface representation of the THAP domain of THAP11 *in cyan*. The residues that display the strongest chemical shift changes are indicated on the surface (Thr6 is buried in the protein core). The orientation is the same as in **a**–**c**

sequence (Campagne et al. 2010). This result suggests that multiple non-specific contacts, such as favourable electrostatic interactions, drive the efficient positioning of the protein onto the DNA and allow side chains to bases specific contacts to take place (Campagne et al. 2010). A scaffold would thus be in place for further efficient specific binding, supporting previous results for the *lac* repressor binding to specific operators (Kalodimos et al. 2004).

As discussed above, mutations on the THAP domain have been correlated with the occurrence of primary dystonia DYT6 (28), providing an excellent case study for structure-activity relationships. Numerous mutated residues in DYT6 were studied by a combination of biophysical approaches (Campagne et al. 2012). In particular, NMR (chemical shift mapping) and differential scanning fluorimetry [DSF, for measuring denaturation temperatures of proteins, Tm, (Layton and Hellinga 2011)] were used to address the effect of the mutations onto the structure and stability of the variant domains, compared with the wild-type THAP domain. Among the most stable mutations, S21T (Tm of 45 °C ± 0.2) has a minor effect onto the structure of the domain, since only the position 21 shows chemical shift changes. By contrast, the S6F variant is strongly destabilized (more than 9 °C compared with the wild-type) and important chemical shift changes are observed for hydrophobic residues located in the helix and for several residues at the end of loop L3 or in the β-sheet (Campagne et al. 2012). These structural consequences induced by the mutation and observed by NMR explain the decrease in the stability followed by DSF. Using ITC and fluorescence, none of the mutations was found to disrupt completely DNA binding. The mutant proteins retain their ability to bind the cognate DNA sequence, with Kd_{mutant}/Kd_{wt} ratios varying from 0.5 to 7.5 depending on the mutation position. On the basis of thermodynamic analysis, more favourable enthalpy change is observed in many DYT6 variants and compensated with weaker or even unfavourable entropy in some DYT6 mutants, resulting in small Gibbs free energy changes. As enthalpic contribution accounts for formation of specific hydrogen bonds and van der Waals interactions upon DNA binding, new/original contacts appear to take place between the mutated THAP1 domain and DNA target. Selectivity towards the DNA sequence could however be affected; for example, the N12K mutant binds to its target DNA with an affinity similar to that of the THAP1 domain whereas it recognizes a non-specific DNA with only a 3 times lower affinity (Campagne et al. 2012).

As discussed above, the THAP domains of THAP1 and *Drosophila* transposase adopt an analogous DNA recognition strategy, where the C-terminal loop L4 contacts DNA minor groove, presumably playing a significant role in DNA binding affinity. Moreover, this loop is the most variable region of THAP domains, being one of the longest in dmTHAP. To further investigate the role of loop L4 in DNA binding, we have undertaken the structural analysis of a cancer-promoting factor THAP11, which, of the 12 human proteins possesses the shortest loop L4.

Solution structure of the THAP-zinc finger domain (1–81) from the cell growth suppressor human THAP11 protein

A minimal 81 residue-long THAP domain of THAP11 including a non-native N-terminal Gly and Ser residue (positions 1 and 2) was purified, as described in the Methods section. The solution structure of the Zn-containing THAP domain was solved at 20 °C in a deuterated NMR buffer (50 mM Tris, pH 6.8, NaCl 30 mM, DTT 5 mM) using nearly complete ^{15}N, ^{13}C and ^{1}H resonance assignments obtained with conventional backbone and side-chain experiments (Sattler et al. 1999). In the NMR buffer, the protein displays a Tm value of 39 °C (data not shown). An ensemble of 19 lowest energy structures was calculated using CNS1.2.1 (Brunger et al. 1998) with a total of 1,766 distance restraints, 66 dihedral angle restraints, 36 hydrogen bonds and 14 zinc-coordination restraints, as described in Table I. As expected, the THAP domain of THAP11 (PDB accession number 2lau) adopts the βαβ fold typical of THAP zinc-finger domains, including a short two-stranded antiparallel beta-sheet (strand β1, residues Phe24-Tyr25 and strand β2, residues Leu61-Cys62) interspersed by a two-turn α helix (residues Ala31-Val41) that contains Trp37, one of the invariant residues. Two short 3–10 helices involving residues Pro55-Thr57 and Ser63-His65 are found (Fig. 5a–b). The domain is stabilized by the presence of a zinc-binding site formed by Cys7, Cys12, Cys62 and His65 that coordinate a single zinc ion. As previously observed in other THAP domains, the Nε2 atom of the His65 ring is the zinc bound atom (Bessiere et al. 2008; Liew et al. 2007). Together with the zinc-binding site, the four invariant residues, namely Pro28 in loop L2 (Pro 26 in THAP1), Trp37 in helix1 (Trp36 in THAP1), Phe66 in loop L3 (Phe58 in THAP1) and Pro78 in loop L4 (Pro78 in THAP1) form a hydrophobic domain with residues Val9, Val41, leu61 and Phe81. The THAP domain possesses in solution only 12 % of secondary structure. Many residues are not involved in hydrogen bonds and most of the amide protons of the THAP domain of THAP11 exchange almost instantaneously with the solvent. After 20 min of NMR H-D exchange, only the backbone HN protons corresponding to Val9 close to the zinc coordinating residues, Trp37 and Leu38 in the helix and His65 of the C2CH zinc-binding motif remain protected. The stability of the protein thus appears slightly lower than

that previously reported for THAP1, its human ortholog (Bessiere et al. 2008), where about 11 amide protons remained protected for over 20 mn: Ser6, Ala7, Cys10, Asn12 in the vicinity of the zinc-binding site; Glu35, Trp36, Glu37, Ala39, Val40 in the helix H1 and Ile53, Cys54 in the β-sheet.

The r.m.s deviation after superimposing the secondary structures and the zinc-coordination site (residues 7, 12, 24–26, 31–41, 55–57, 60–62, 65) in the final 19 structures is 0.42 Å (Table 1). When the superimposition is done on the full-length backbone, the r.m.s value drops to 1.43 Å, reflecting the presence of several unstructured regions of the THAP domain. Indeed, the domain includes four loops that link the secondary structure elements and that display mobility and/or conformational exchange. Loop L1 in the N-terminus that contains the first pair of zinc ligands is highly mobile from Gly4 to His23, consistent with the lack of NOEs in this region (Fig. 5c) and the ^{15}N relaxation parameters (Fig. 5d). Two loop regions display high mobility on the ps-ns time scale, namely the region that encompasses residues Gly45-Gly48 in loop L3 and the region (Gly69-Thr74) at the beginning of loop L4. ^{15}N longitudinal and transverse relation times were used to determine a rotational correlation time (τ_c) of 5.9 ns consistent with a monomeric state of the THAP domain (data not shown).

Structure comparison with other THAP domains

As shown on the schematic drawing (Figs. 2b, 5a), THAP11 displays the βαβ topology previously observed in the structures of THAP domains for THAP1 [PDB 2jtg, (Bessiere et al. 2008)], THAP2 (PDB 2d8r), Ce-CtBP [PDB 2jm3, (Liew et al. 2007)] and transposase [PDB 3kde, (Sabogal et al. 2010)]. The β-sheet and the α-helix correspond to regions of high sequence similarity within the THAP domain. In particular, the helix contains the invariant Trp residue (Trp37 in THAP11) and two residues (leu38 and val41) relatively well conserved within the THAP family. The four invariant side-chain residues of the domain occupy structural equivalent positions in the THAP domains and constitute a molecular scaffold with the zinc-binding site, around which regions of sequence diversity are incorporated. Although the THAP domain of THAP11 displays the highest sequence identity with the THAP domain of THAP2 (23 % against 20 % with Ce-CtBP), the best structural homology was found with the THAP domain of Ce-CtBP (3.26 Å for 61 aligned positions *vs*. 3.04 Å for 54 aligned positions with THAP2). Besides, the domain includes a conserved short 3–10 helix at residues Ser63-His65. The sequence of the THAP domain of THAP11 incorporates original features, including loop L3 (Ser42-Arg60), the longest one among the human THAP members (18 residues instead of 11 in THAP1) that incorporates a short 3–10 helix (Pro55-Thr57) in THAP11. The insertion of 7 residues within loop L3 in the sequence of THAP11 increases mobility on the ps-ns time scale for these residues Gly45-Gly48, as judged by the heteronuclear NOEs (Fig. 5d). By contrast, of the 12 human THAP proteins, THAP11 has the shortest C-terminal loop L4. This loop L4 follows the second β-strand and stands between two invariant residues important for the protein folding (Phe66 and Pro78). The sequence alignment of Fig. 2a shows that 9 residues are missing between Gly69 and Arg70, which causes the loop shortening. This loop is longer in most THAP domains, apart from the THAP domain Ce-CtBP from *C. elegans,* which displays a loop L4 of similar size with THAP11. Interestingly, the residues that are missing correspond to highly mobile region in THAP1, (Bessiere et al. 2008). The shortened loop L4 displays low mobility on the ps-ns time scale, except in a short region centred at Lys71, which displays higher mobility and a poor number of NOE restraints (Fig. 5c–d).

The shortened loop L4 in THAP11 may have negative effect on DNA binding specificity

Noteworthy, loop L4 has been shown previously to insert into the minor groove and contact specific bases (Fig. 4a–d). In particular, an arginine (Arg65) found at equivalent positions in both THAP1 and dmTHAP domains contacts the thymine of the consensus DNA target TxxxGGCA (Campagne et al. 2010; Sabogal et al. 2010). In these two examples, the length of the loop is sufficient to bring the arginine closer to the DNA minor groove for specific recognition. In THAP11, the length of the loop is significantly shortened, the arginine residue (Arg70, numbering corresponding to THAP11) present in the loop is way too far from the consensus thymine (more than 10 Å), precluding DNA minor groove recognition, at least using this thymine as a specific recognition site and a canonical, straight, B-DNA conformation.

Nevertheless, THAP11 and its murine homolog have been claimed to recognize specific DNA targets and subsequently collaborate with interaction partners to achieve a specific regulatory response (Parker et al. 2012; Zhu et al. 2009). Two DNA consensus sequences (Ronin-binding sequence and Ronin-binding motif) distinct from the THABS motif were identified (Fig. 3a–b). The SELEX-derived Ronin-binding sequence is found in the c-Myc promoter (Zhu et al. 2009). Besides, as discussed above, this sequence (Dejosez et al. 2008) correlates poorly with the Ronin-binding motif determined later using chromatin-immuno-precipitation and DNA sequencing approaches (Dejosez et al. 2010).

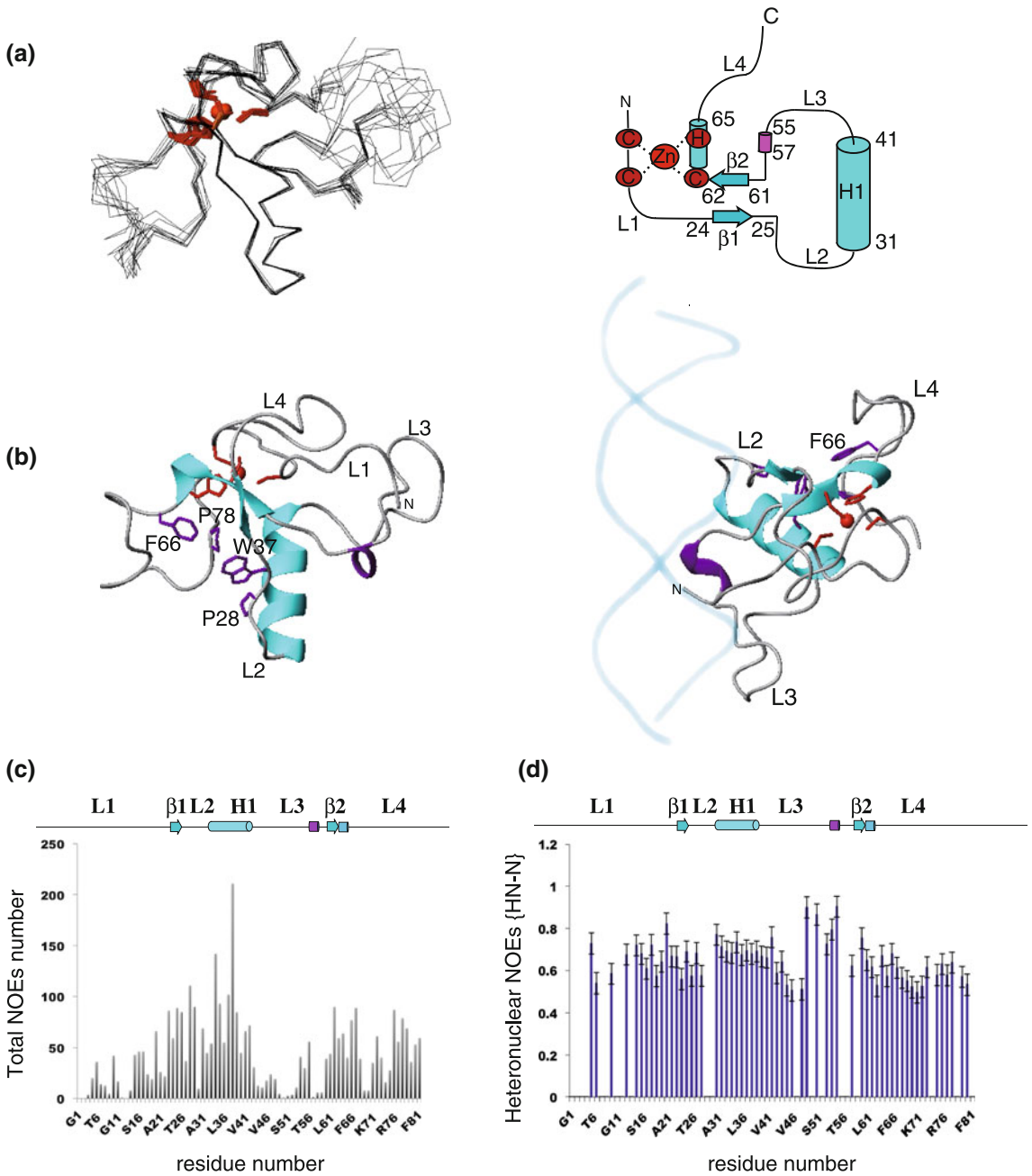

Fig. 5 Three-dimensional structure of the THAP domain of human THAP11 (Gly1-Phe81). **a** Cα trace for the 19 lowest energy superimposed NMR structures and topologic view of the secondary structure elements. **b** Ribbon diagram of the THAP domain of THAP11 showing the secondary structure elements in cyan and purple (*left*). The zinc and the four coordinating side-chain residues are coloured in *red*. Side-chains of the four invariant residues are indicated. (*right*) The orientation is identical to that of Fig. 4, illustrating the putative side of DNA recognition. **c** Number of NOE distances as a function of residue number. **d** Histogram showing the heteronuclear NOE values for the THAP domain as a function of residue number

In order to address the DNA binding properties of THAP11, 2D ^{15}N–^{1}H HSQC spectra were recorded to follow the residues that exhibit chemical shift changes of their backbone amide nitrogen resonances upon DNA binding (Fig. 4e). In the presence of a 20-bp DNA fragment containing the consensus Ronin-binding sequence present in the c-Myc promoter at 20 mM NaCl, a large number of residues were affected, displaying however very slight chemical shift perturbations (<0.15 ppm). The largest change was observed for Thr26 in the β-sheet. Looking closely into the sequence-structural alignment indicates that several of these residues occupy positions in THAP1 and in dmTHAP that are involved in direct DNA contacts, suggesting an orientation similar to that previously

Table 1 Structural Statistics for the THAP domain of human THAP11

Restraints for final structure calculations	
NOE distance restraints	
Total NOE	1,766
Intraresidue	0
Sequential ($\|i - j\| = 1$)	739
Medium-range ($1 < \|i - j\| < 4$)	356
Long-range ($\|i - j\| > 4$)	671
Dihedral angle restraints	66
Angle φ (°)	35
Angle ψ (°)	31
Hydrogen bonds	36
Zinc coordination	14
Structure statistics	
R.m.s deviations from idealized covalent geometry	
Bond length (Å)	0.009 ± 0.00025
Bond angles (°)	1.150 ± 0.037
Improper angle (°)	0.982 ± 0.038
NOE restraints (Å)	0.062 ± 0.001
Final energies (kcal mol^{-1})	1,543.9 ± 32.9
Ramachandran plot[b]	
Residues in most favored region (%)	70.2
Residues in additional allowed regions (%)	23.5
Residues in generously allowed regions (%)	6.3
Residues in disallowed regions (%)	0
R.m.s deviation (Å)[a]	
Backbone atoms in secondary structures	0.423
Backbone atoms (4–80)	1.432
Heavy atoms in secondary structures	0.792
Heavy atoms (4–80)	1.865

[a] Average pairwise r.m.s deviation between the ensemble of the 19 structures. Secondary structures refer to residues 24–26, 31–41, 60–62, 55–57 and 7, 12, 62, 65 (zinc site)

[b] Residues 7–41, 58–67, 75–81

described for the complexes THAP1 and dmTHAP. The comparison between the chemical shift changes observed in THAP11 and THAP1 (Campagne et al. 2010) upon DNA binding reveals a striking similarity, in particular with major shifts on Thr6 (Ser4 in THAP1), Thr26 (Lys24 in THAP1), Thr57 (Thr48 in THAP1) and Lys71 (Leu72 in THAP1), (Fig. 4d–f). These shifts however, are of weaker intensity in the case of THAP11 compared with THAP1 (two to three fold weaker). Furthermore, NMR experiments recorded for THAP11 in the presence of the THABS sequence or in the presence of a non-specific target exhibited a similar CSP profile. Similarly DNA binding affinity measurements, using ITC and fluorescence spectroscopy, have revealed little difference between the so far assumed DNA targets of THAP11 (c-Myc and Ronin-binding motif) and random sequences of identical length (data not shown). Based on these observations, our current hypothesis, which remains to be tested before going further into precise structural characterisation of THAP11-DNA complexes, is that the actual DNA target of THAP11 may still be ill defined. Interestingly, the THAP domain of Ce-CtBP, which displays similar features to THAP1 such as a short loop L4, has also been proposed to bind to the THABS motif (Liew et al. 2007). However, we have shown that this binding is non-specific in vitro, again casting doubt on its actual DNA target (Bessiere et al. 2008).

Conclusion

Since the discovery of the THAP domain, more than 10 years ago, major breakthroughs have been obtained, improving the knowledge on the structure and biological functions of the THAP domains. They possess an overall βαβ conserved fold around the zinc-coordination site that acts as a molecular scaffold to recognize specific DNA sequences, provided that an optimal combination of side chains to base contacts is satisfied. The solution structure of the THAP domain of human THAP11 is the fifth structure to be solved, after human THAP1 and THAP2, *C. elegans* CtBP and *Drosophila* transposase. It contains the structural scaffold with the βαβ topology typical of the THAP domain; it lacks however several residues in the C-terminal loop L4 that provides DNA minor groove contacts in other THAP domains. Therefore, specific DNA binding capacities are likely to be affected. Thus, beyond the current model derived from THAP1-DNA and dm-THAP-DNA structures, THAP11 and Ce-CtBP-DNA complexes may reveal new mode of specific recognition.

Materials and methods

Protein expression and purification

The THAP zinc-finger domain of human THAP11 (GenBank™ NP_065190) was amplified by PCR and sub-cloned in-frame into a pGEX-4T2 expression vector (GE Healthcare), using the BamH1 and XhoI restriction sites. The recombinant THAP domain was produced in *Escherichia coli* BL21(DE3). Cells were transformed with the expression vector and grown at 20 °C in auto-inducing minimal medium (AIMM) for 60 h to obtain an unlabeled THAP domain sample (Studier 2005). Uniformly ^{15}N and $^{15}N/^{13}C$–labeled THAP domains were expressed in the same AIMM containing respectively ^{15}N NH_4Cl or either ^{15}N NH_4Cl and (^{13}C) glycerol. The THAP domain was produced as a GST fusion protein and purified from the cell lysate using glutathione Sepharose beads (GE Healthcare)

in a PBS buffer at pH 7.4 and 2 mM DTT. The sample was eluted in 100 mM Tris–HCl buffer, pH 7.8, 150 mM NaCl, 2 mM DTT and 10 mM reduced glutathione (Sigma). The THAP domain was released from the GST tag using thrombin (Novagen), prior to gel filtration chromatography on Superdex 75 (Amersham Biosciences) using 50 mM Tris–HCl as buffer, pH 7.4, 150 mM NaCl and 2 mM DTT. Sample purity was verified by SDS-PAGE with a single strong band at 10 kDa (estimated $M_W = 9{,}112$ da). Sample concentration was determined by measuring absorption of the recombinant human THAP11 domain at 280 nm (estimated ε of 9,970 M^{-1} cm^{-1}). The purified THAP domain was concentrated to 0.5–0.8 μM and buffer-exchanged by successive dilution-concentration steps into NMR buffer (50 mM deuterated Tris–HCl pH 6.8, 30 mM NaCl, 5 mM deuterated DTT and 0.01 % (w /v) NaN_3).

NMR experiments

NMR experiments were performed in a solvent mixture of 90 % H_2O/10 % D_2O at 293 K using Bruker Avance 950 MHz and 600 MHz spectrometers equipped with a cryo-probe system and on a Bruker Avance 700 MHz spectrometer. NMR spectra were recorded on unlabeled, ^{15}N single labeled and ^{15}N/ ^{13}C double labeled THAP domains. ^{1}H, ^{15}N and ^{13}C backbone and side-chain resonances were assigned by a standard procedure using a combination of homonuclear and heteronuclear experiments (Sattler et al. 1999). Stereospecific assignments of methylene Hβ protons were obtained on the basis of the patterns of HN–Hβ and Hα–Hβ NOEs and $^{3}J_{HAHB}$ couplings (Wagner et al. 1987). All X-Pro cis-trans isomers were identified on the basis of strong NOEs between the H^{δ} proton of each Pro residue and the H^{α} protons of the preceding residues and confirmed by $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ chemical shifts (Wuthrich 1986). $^{2}H_2O$-exchanged samples were prepared from overnight freeze-drying and resuspension into a 99.9 % $^{2}H_2O$ buffer. Spectra were processed with Topsin software (Bruker BioSpin) and analysed with CARA (Keller 2004). For backbone amide relaxation measurements, ^{15}N heteronuclear relaxation parameters (T1, T2, NOE) were recorded at 600 MHz, using standard pulse sequences. Longitudinal T1 relaxation delays were set to 10, 50, 100, 200, 400, 700, 1,000, 1,500, 2,000 and 3,000 ms. T2 relaxation delays were set to 20, 36, 60, 92, 120, 140, 170, 202 and 282 ms. The heteronuclear NOEs were determined from two 2D ^{1}H–^{15}N HSQC spectra recorded in the presence or the absence of ^{1}H presaturation period of 3 s and with a recycling delay of 5 s (Auguin et al. 2004). In both cases, spectra were processed with NMR Pipe and peak volumes were fitted to a single-exponential decay function using the program NMR View (Johnson 2004). The overall and internal mobility parameters were determined using the Tensorv 2.0 software (Dosset et al. 2000).

NMR structure determination

Structures of the THAP domain of human THAP11 were solved using torsion angle dynamics simulated annealing protocol, followed by water refinement using the CNSv1.21 software suite (Brunger et al. 1998). Inter-proton distances were extracted from 3D ^{15}N HSQC-NOESY (Tm 130 ms) and 3D $^{13}C_{ali}$ HSQC-NOESY (Tm 80 ms) spectra recorded at 950 MHz and a homonuclear ^{1}H NOESY (Tm 150 ms) spectrum recorded at 700 MHz. Torsion angle restraints were obtained from TALOS (Cornilescu et al. 1999) as well as $^{3}J_{HN\text{-}H\alpha}$ coupling constants determined from a 3D HNHA experiment (Ponstingl and Otting 1998). Hydrogen bonds were introduced based on the presence of secondary structure elements, identified by slowly exchanging amide protons, $^{3}J_{HAHB}$ coupling constants and characteristic NOEs. Zinc tetrahedral coordination constraints including S^{γ} of the C^{7}, C^{12}, C^{62} and $N^{\varepsilon 2}$ of the His^{65} were added, as described previously (Bessiere et al. 2008). After the first few rounds of calculation, preliminary structures were used to correct misidentified NOEs and to end up with the right final NOE restraints list. Ultimately, from 100 structures, the 19 structures with the lowest energy were selected to represent the ensemble of the THAP domain structures. Its structural quality was analyzed with PROCHECK-NMR (Laskowski et al. 1996).

DNA-binding titration experiments

The 20-bp duplex containing the Ronin-binding sequence present in the c-Myc promoter (Zhu et al. 2009) was reconstituted by hybridizing the oligonucleotides, 5′CCAC TCTCCCTGGGACTCTT3′ and 5′AAGAGTCCCAGGGA GAGTGG3′ (Eurofins MWG) in a 1:1 ratio. The oligonucleotide containing the Ronin binding motif (Dejosez et al. 2010) was reconstituted using the oligonucleotides, 5′CTG GGAATTGTAGTT3′ and 5′AACTACAATTCCCAG3′. A non-specific 20-bp poly(dA/dT) oligonucleotide was also used for binding experiments.

For NMR titration, 2D ^{15}N HSQC spectra of the THAP domain of THAP11 (concentration of 150 μM) were recorded at 293 K after each incremental addition of lyophilized DNA (Ronin-binding sequence), as described previously (Bessiere et al. 2008). ITC experiments were conducted at 293 K on a Microcal ITC200 instrument following the protocol described previously (Campagne et al. 2012). Fluorescence spectra were recorded following the tryptophan (Trp37) emission at 324 nm as described previously (Campagne et al. 2012).

Accession codes

The atomic coordinates for the THAP domain of human THAP11 structures have been deposited at the Protein Data Bank (PDB) under accession number 2LAU, and experimental data are available from the BioMagRes Data Bank, under the Accession code 17456.

Acknowledgments EU structural funds and the Région Midi-Pyrénées are acknowledged for funding the NMR equipment. Microcalorimetry and thermal shift assays equipments have been acquired by the IBiSA Integrated Screening Platform of Toulouse (PICT, IPBS, CNRS—Université de Toulouse). Access via TGIR national facility for high field NMR was provided in ICSN, CNRS, Gif/Yvette, where 950 MHz NMR spectra were recorded. J.P. Girard is acknowledged for initiating this work at the IPBS and for continuous support and advices.

References

Auguin D, Barthe P, Auge-Senegas MT, Stern MH, Noguchi M, Roumestand C (2004) Solution structure and backbone dynamics of the pleckstrin homology domain of the human protein kinase B (PKB/Akt). Interaction with inositol phosphates. J Biomol NMR 28(2):137–155

Balakrishnan MP, Cilenti L, Mashak Z, Popat P, Alnemri ES, Zervos AS (2009) THAP5 is a human cardiac-specific inhibitor of cell cycle that is cleaved by the proapoptotic Omi/HtrA2 protease during cell death. Am J Physiol Heart Circ Physiol 297(2): H643–H653

Bessiere D, Lacroix C, Campagne S, Ecochard V, Guillet V, Mourey L, Lopez F, Czaplicki J, Demange P, Milon A, Girard JP, Gervais V (2008) Structure-function analysis of the THAP zinc finger of THAP1, a large C2CH DNA-binding module linked to Rb/E2F pathways. J Biol Chem 283(7):4352–4363

Bonetti M, Barzaghi C, Brancati F, Ferraris A, Bellacchio E, Giovanetti A, Ialongo T, Zorzi G, Piano C, Petracca M, Albanese A, Nardocci N, Dallapiccola B, Bentivoglio AR, Garavaglia B, Valente EM (2009) Mutation screening of the DYT6/THAP1 gene in Italy. Mov Disord Off J Mov Disord Soc 24(16):2424–2427

Bouvet P (2001) Determination of nucleic acid recognition sequences by SELEX. Methods Mol Biol 148:603–610

Bragg DC, Armata IA, Nery FC, Breakefield XO, Sharma N (2011) Molecular pathways in dystonia. Neurobiol Dis 42(2):136–147

Bressman SB, Raymond D, Fuchs T, Heiman GA, Ozelius LJ, Saunders-Pullman R (2009) Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study. Lancet Neurol 8(5):441–446

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905–921

Campagne S, Saurel O, Gervais V, Milon A (2010) Structural determinants of specific DNA-recognition by the THAP zinc finger. Nucleic Acids Res 38(10):3466–3476

Campagne S, Muller I, Milon A, Gervais V (2012) Towards the classification of DYT6 dystonia mutants in the DNA-binding domain of THAP1. Nucleic Acids Res 40(19):9927–9940

Cayrol C, Lacroix C, Mathe C, Ecochard V, Ceribelli M, Loreau E, Lazar V, Dessen P, Mantovani R, Aguilar L, Girard JP (2007) The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. Blood 109(2):584–594

Clouaire T, Roussigne M, Ecochard V, Mathe C, Amalric F, Girard JP (2005) The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proc Natl Acad Sci USA 102(19):6907–6912

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13(3):289–302

De Souza SantosE, De Bessa SA, Netto MM, Nagai MA (2008) Silencing of LRRC49 and THAP10 genes by bidirectional promoter hypermethylation is a frequent event in breast cancer. Int J Oncol 33(1):25–31

Dejosez M, Krumenacker JS, Zitur LJ, Passeri M, Chu LF, Songyang Z, Thomson JA, Zwaka TP (2008) Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell 133(7):1162–1174

Dejosez M, Levine SS, Frampton GM, Whyte WA, Stratton SA, Barton MC, Gunaratne PH, Young RA, Zwaka TP (2010) Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes Dev 24(14):1479–1484

Djarmati A, Schneider SA, Lohmann K, Winkler S, Pawlack H, Hagenah J, Bruggemann N, Zittel S, Fuchs T, Rakovic A, Schmidt A, Jabusch HC, Wilcox R, Kostic VS, Siebner H, Altenmuller E, Munchau A, Ozelius LJ, Klein C (2009) Mutations in THAP1 (DYT6) and generalised dystonia with prominent spasmodic dysphonia: a genetic screening study. Lancet Neurol 8(5):447–452

Dosset P, Hus JC, Blackledge M, Marion D (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J Biomol NMR 16(1):23–28

Fuchs T, Gavarini S, Saunders-Pullman R, Raymond D, Ehrlich ME, Bressman SB, Ozelius LJ (2009) Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nat Genet 41(3):286–288

Gavarini S, Cayrol C, Fuchs T, Lyons N, Ehrlich ME, Girard JP, Ozelius LJ (2010) Direct interaction between causative genes of DYT1 and DYT6 primary dystonia. Ann Neurol 68(4):549–553

Johnson BA (2004) Using NMR View to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278:313–352

Kaiser FJ, Osmanoric A, Rakovic A, Erogullari A, Uflacker N, Braunholz D, Lohnau T, Orolicki S, Albrecht M, Gillessen-Kaesbach G, Klein C, Lohmann K (2010) The dystonia gene DYT1 is repressed by the transcription factor THAP1 (DYT6). Ann Neurol 68(4):554–559

Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305(5682):386–389

Keller RLJ (2004) The computer aided resonance assignment tutorial, 1st edn. Cantina Verlag, Switzerland

Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol 11(1):39–46

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8(4):477–486

Layton CJ, Hellinga HW (2011) Quantitation of protein-protein interactions by thermal stability shift analysis. Protein Sci. doi: 10.1002/pro.674

Lian WX, Yin RH, Kong XZ, Zhang T, Huang XH, Zheng WW, Yang Y, Zhan YQ, Xu WX, Yu M, Ge CH, Guo JT, Li CY, Yang XM (2012) THAP11, a novel binding protein of PCBP1, negatively regulates CD44 alternative splicing and cell invasion in a human hepatoma cell line. FEBS Lett 586(10):1431–1438

Liew CK, Crossley M, Mackay JP, Nicholas HR (2007) Solution structure of the THAP domain from *Caenorhabditis elegans* C-terminal binding protein (CtBP). J Mol Biol 366(2):382–390

Lohmann K, Uflacker N, Erogullari A, Lohnau T, Winkler S, Dendorfer A, Schneider SA, Osmanovic A, Svetel M, Ferbert A, Zittel S, Kuhn AA, Schmidt A, Altenmuller E, Munchau A, Kamm C, Wittstock M, Kupsch A, Moro E, Volkmann J, Kostic V, Kaiser FJ, Klein C, Bruggemann N (2012) Identification and functional analysis of novel THAP1 mutations. Eur J Hum Genet EJHG 20(2):171–175

Macfarlan T, Kutney S, Altman B, Montross R, Yu J, Chakravarti D (2005) Human THAP7 is a chromatin-associated, histone tail-binding protein that represses transcription via recruitment of HDAC3 and nuclear hormone receptor corepressor. J Biol Chem 280(8):7346–7358

Mazars R, Gonzalez-de-Peredo A, Cayrol C, Lavigne AC, Vogel JL, Ortega N, Lacroix C, Gautier V, Huet G, Ray A, Monsarrat B, Kristie TM, Girard JP (2010) The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias. J Biol Chem 285(18):13364–13371

Muller U (2009) The monogenic primary dystonias. Brain 132(Pt 8):2005–2025

Nakamura S, Yokota D, Tan L, Nagata Y, Takemura T, Hirano I, Shigeno K, Shibata K, Fujisawa S, Ohnishi K (2012) Down-regulation of Thanatos-associated protein 11 by BCR-ABL promotes CML cell proliferation through c-Myc expression. Int J Cancer J Int Du Cancer 130(5):1046–1059

Parker JB, Palchaudhuri S, Yin H, Wei J, Chakravarti D (2012) A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function. Mol Cell Biol 32(9):1654–1670

Ponstingl H, Otting G (1998) Rapid measurement of scalar three-bond 1HN–1H alpha spin coupling constants in 15 N-labelled proteins. J Biomol NMR 12(2):319–324

Privalov PL, Dragan AI, Crane-Robinson C, Breslauer KJ, Remeta DP, Minetti CA (2007) What drives proteins into the major or minor grooves of DNA? J Mol Biol 365(1):1–9

Privalov PL, Dragan AI, Crane-Robinson C (2011) Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Res 39(7):2483–2491

Roussigne M, Cayrol C, Clouaire T, Amalric F, Girard JP (2003a) THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies. Oncogene 22(16):2432–2442

Roussigne M, Kossida S, Lavigne AC, Clouaire T, Ecochard V, Glories A, Amalric F, Girard JP (2003b) The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends Biochem Sci 28(2):66–69

Sabogal A, Lyubimov AY, Corn JE, Berger JM, Rio DC (2010) THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves. Nat Struct Mol Biol 17(1):117–123

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34(2):93–158

Sengel C, Gavarini S, Sharma N, Ozelius LJ, Bragg DC (2011) Dimerization of the DYT6 dystonia protein, THAP1, requires residues within the coiled-coil domain. J Neurochem 118(6):1087–1100

Sohn AS, Glockle N, Doetzer AD, Deuschl G, Felbor U, Topka HR, Schols L, Riess O, Bauer P, Muller U, Grundmann K (2010) Prevalence of THAP1 sequence variants in German patients with primary dystonia. Mov Disord Off J Mov Disord Soc 25(12): 1982–1986

Song W, Chen Y, Huang R, Chen K, Pan P, Yang Y, Shang HF (2011) Novel THAP1 gene mutations in patients with primary dystonia from Southwest China. J Neurol Sci 309(1–2):63–67

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr Purif 41(1):207–234

Wagner G, Braun W, Havel TF, Schaumann T, Go N, Wuthrich K (1987) Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. J Mol Biol 196(3):611–639

Wuthrich K (1986) NMR of Proteins and Nucleic Acids. Wiley, Hoboken

Xiao J (2010) Novel human pathological mutations. Gene symbol: THAP1. Disease: dystonia 6. Hum Genet 127(4):469

Zhu CY, Li CY, Li Y, Zhan YQ, Li YH, Xu CW, Xu WX, Sun HB, Yang XM (2009) Cell growth suppression by thanatos-associated protein 11(THAP11) is mediated by transcriptional downregulation of c-Myc. Cell Death Differ 16(3):395–405