Molecular BioSystems

CrossMark
← click for updates

Shared structural features of the 9aaTAD family in complex with CBP†

Martin Piskacek,‡*[a] Anna Vasku,[a] Roman Hajek[b] and Andrea Knight‡[a]

A number of transactivation domains for transcription factors including p53, E2A/HEB, MLL, cMyb, CREB, FOXO3, Gcn4, Oaf1 and Pdr1 have been reported to interact with the KIX domain of general transcriptional mediators CBP, p300 or MED15. Most of those factors belong to the already established Nine amino acid Transactivation Domain (9aaTAD) family. By using available structural data, we found binding analogy for the 9aaTAD in the MLL–KIX and also E2A/HEB–KIX complexes. We recognized two distinct TAD formations in the KIX complex. In the E2A/HEB–KIX complex, the leucine position is determined by the prolonged helical structure including the 9aaTAD and the leucine (long-helical TAD). However in the MLL–KIX complex, the equal position of 9aaTAD and proximal leucine is achieved differently by leucine-turn-helix structural architecture. Furthermore, the FOXO3–KIX complex shares structural analogy with the E2A–KIX complex in respect of both 9aaTAD and proximal leucine. Next, from (i) sequence alignment of the identified 9aaTADs in p53, E2A/HEB and MLL proteins and (ii) the resolved structure of the MLL–KIX and E2A/HEB–KIX complexes, we generated a plausible structural model for p53 that could be used also for other members of the 9aaTAD family. The position of 9aaTADs in Oaf1-, Pdr1- and Gcn4-MED15 KIX complexes and 9aaTAD composition are in good agreement with E2A, MLL, FOXO3 and p53. Analyses of structural data in this study define fundamental structural requirements and shed more light on the ambiguous 9aaTAD domain.

Introduction

Characterization of transactivation domains (TAD) has been a subject of many studies and their initial description has been derived from prototypical transcription factors Gal4 and Gcn4. So called acidic transactivation domains have been defined with essential acidic and clustered hydrophobic residues.[1–3] More recently, a number of transcription factors including E2A/HEB, MLL, p53, cMyb, CREB, Gcn4, Oaf1 and Pdr1 were reported to interact with the KIX domain of either the general transcriptional co-activator CBP or MED15 (NMR data).[4–17] Structural studies by Hinnebusch, Verdine and Wright labs (Gcn4, p53 and MLL), which reform the dogma of acidic transactivation domains from the early 1990s,[2] revealed true amphipathic clusters of hydrophobic residues in these transactivation domains.[1,5,6,12,18,19–23] Nevertheless, defined amphipathic motifs (ξ–ϕ–x–ϕ–ϕ, ϕ–x–x–ϕ–ϕ or ϕ–ϕ–x–x–ϕ motifs with acidic (ξ) and hydrophobic (ϕ) residues)[12,16] are neither specific nor well conserved in the binding of the TADs with the KIX domain, *e.g.* p53 or FOXO3.[12,24,25]

Previously, we have experimentally determined short transactivation domains in several transcription factors including Oaf1, Pip2 and Gal4.[26,27] Shared features of paralogs and orthologs enabled us to define a family of Nine amino acid Transactivation Domain (9aaTAD), which has been also identified in E2A, MLL, p53, NF-kB, NFAT, CEBP, Sox18, Oaf1, Pip2, Pdr1, Gal4, Gcn4, and VP16 transcription factors (www.uniprot.org/uniprot/?query=9aatad&sort=score).[26,27] These findings have encouraged subsequent structural studies with 9aaTAD peptides. Our 9aaTAD prediction, description & annotations service (online since 2007)[26] was used and shown by Näär lab that Oaf1 and Pdr1 9aaTADs bind correspondingly to the same region of the KIX domain of Med15.[9,10]

Furthermore, artificial peptides and Gcn4 TAD-I interactions with the Med15 were reported.[15,18,28,29] Unlike CBP, Med15 is evolutionary conserved from yeast to man and may explain conservation of the transactivation motif in eukaryotes. Importantly, Med15 has multiple binding sites for the TADs similar to CBP.[18] The KIX domain itself has two binding sites: the 9aaTAD of MLL and both p53 9aaTAD-I and II domains bind to the same region of the KIX domain, called the MLL site of the KIX domain. In addition, both 9aaTAD-I and II domains of the p53 and Myb TAD also bind to the Myb site.[6,8,12,19,30,31]

[a] *Department of Pathological Physiology, Faculty of Medicine, Masaryk University Brno, Czech Republic. E-mail: Piskacek@med.muni.cz*
[b] *Faculty Hospital Ostrava, Czech Republic*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4mb00672k
‡ These authors contributed equally to this study and should be considered as joint senior authors.

Recently, the structure of the transactivation domain E2A/HEB complex with the KIX domain was reported and some hydrophobic similarity with the MLL–KIX complex has been discussed.[16] The transactivation domains of E2A and its close relative HEB belong to the TCF family and their pattern is conserved in higher eukaryotes with T/S–DKEL–S/N–DLLD–F/L–S–M/A/K–M. Further motifs were reported within this sequence, such as the "LxxLL motif" as a part of an inhibitory PCET motif (p300/CBP and ETO target in E proteins)[32] or a "SAGA interacting motif LDFS", which are all linked to transactivation function.[33,34]

In this report, we have analysed available structural data of MLL,[6] E2A/HEB[16] and FOXO3[24,25] proteins in respect of shared 9aaTAD structural requirements and sequence variations.

Results and discussion

Structural alignment of HEB and MLL transactivation domains with the KIX domain

The shared sequence characteristics (pattern and clustering of hydrophobic and hydrophilic amino acids) of the 9aaTAD family are based on well characterised transcription factors interacting with general transcriptional mediators TAF9, Med15 and CBP/p300.

So far, only two completely resolved structures have been reported for proteins previously annotated to the 9aaTAD family; MLL and HEB/E2A in complex with the KIX domain of CBP.[6,16]

The structural coordinates for HEB and MLL in complex with the KIX domain of CBP were downloaded from PDB, Protein Data Bank, accession number 2KWF and 2AGH. Here, we analysed the structural data of the HEB/KIX and MLL/KIX complexes and showed a high degree of structural equivalence and conformity of the critical residue positions involved in the binding (Fig. 1A and B).

HEB-Leu20 and MLL-Ile2849 represent the first hydrophobic patch while HEB-Phe23 and MLL-Phe2852 represent the second hydrophobic patch in the 9aaTAD α-helix (*both in green*). Both hydrophobic patches are interspersed by hydrophilic residues HEB-Asp22 and MLL-Asp2851. HEB-Phe23 and MLL-Phe2852 (*both in green*) are located in the pocket (*in yellow*), which is created by Phe612, Ala619 and Arg624 of the KIX C-terminal end of α1 helix, G2 small helix and the N-terminal end of α2 helix. Analogically, HEB-Ser24 and MLL-Val2853 are similarly located between Phe612 (*in yellow*) and Leu664 (*in black*) of KIX α2 helix (Fig. 1A).

The hydrophilic Aspartates of either HEB-Asp22 or MLL-Asp2851 (*in red*) are positioned in the middle of the 9aaTAD α-helix, are exposed to the solvent and generate the hydrophilic
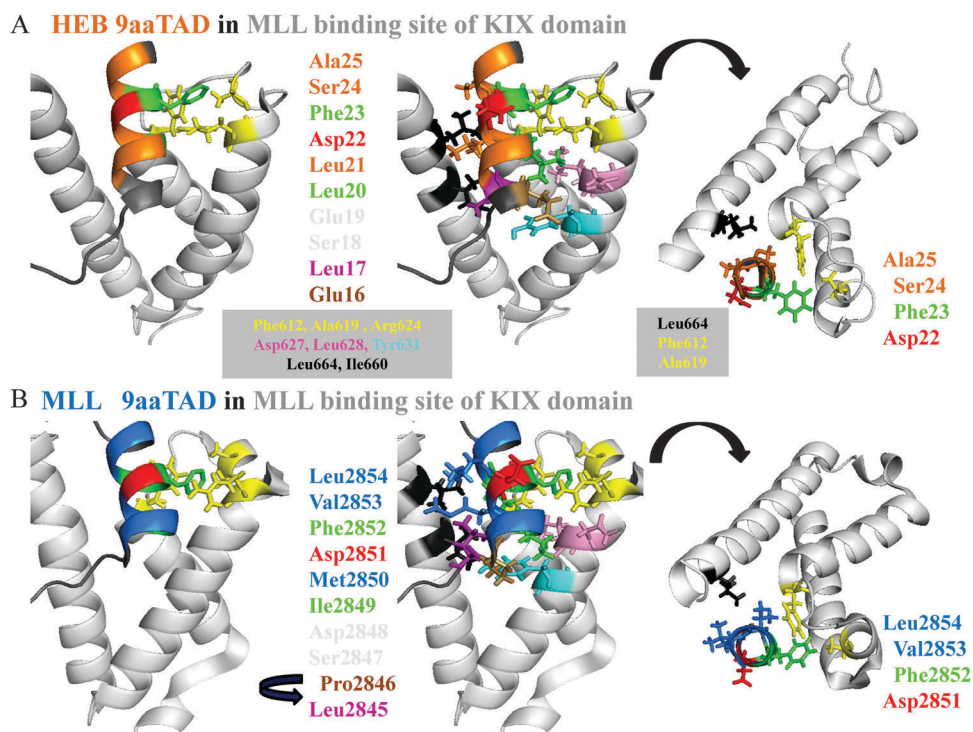

Fig. 1 The 9aaTAD α-helix of E2A/HEB and MLL in the MLL-binding site of the CBP KIX domain. The structure coordinates for HEB and MLL in complex with the KIX domain were downloaded from PDB (2KWF and 2AGH). The 9aaTAD of HEB is labelled in orange (a) and 9aaTAD of MLL in marine blue (b). The main hydrophobic residues of either HEB-Leu20/HEB-Phe23 or MLL-Ile2849 /MLL-Phe2852 facing the KIX α2-helix are in green. The secondary hydrophobic residues of HEB-Leu21/HEB-Ala25 are labelled in orange and MLL-Met2850/MLL-Leu2854 in marine blue facing the KIX α3-helix in black. The hydrophilic Aspartates of either HEB-Asp22 or MLL-Asp2851 are labelled in red, positioned in the middle of the 9aaTAD α-helix and exposed to solvent. HEB-Leu17 and MLL-Leu2845 in magenta are in a proximal position to Ile660 of the KIX α3 helix in black. HEB-Glu16 and MLL-Pro2846 both in brown are in a proximal position to Tyr631 in cyan of the KIX α2 helix. Distinct residues of the MLL-binding site of the CBP KIX domain are highlighted as follows: Phe612, Ala619 and Arg624 in yellow spreading from the α1 to α2-helix of the KIX domain; Asp627 and Leu628 in pink, Tyr631 in cyan as part of the KIX α2-helix and Leu664 and Ile660 in black as part of the KIX α3-helix.

site of the otherwise hydrophobic α-helix (respectively HEB-Asp19 and MLL-Asp2848) (Fig. 1).

HEB-Leu21 and MLL-Met2850 together with HEB-Ala25 and MLL-Leu2854 (*in orange for HEB and in marine blue for MLL*) are located around Leu664 (*in black*) of KIX α3 helix (Fig. 1). Surprisingly, MLL-Pro2846 (*in brown*) is in the position of HEB-Glu16 (*in brown*) in proximal position to Tyr631 (*in cyan*) of the KIX α2 helix. The MLL-Pro2846 causes structural reposition of MLL-Leu2845 (*in magenta*). MLL-Leu2845 is therefore in a comparable position as the HEB-Leu17 (*in magenta*) and both are in a proximal position to Ile660 of KIX α3 helix (*in black*), which belongs to the KIX hydrophobic core[35] (Fig. 1).

HEB-Glu16 and MLL-Pro2846 share comparable coordinates in the secondary structure but in primary structure (the amino acid sequence) alignment this results in the shift of these two amino acids. Thus, a single amino acid gap occurs in the HEB linear sequence aligned to MLL-Pro2846 (*marked by a star*) where HEB-Leu17 is aligned to MLL-Leu2845 (Fig. 2).

Finally, we recognized two different 9aaTAD formations accomplished by the position of proximal leucine. The position of leucine in the E2A/HEB–KIX complex is determined by the prolonged helical structure from 9aaTAD helix (long helical TAD, red box in Fig. 2). The position of leucine in the MLL–KIX complex is achieved by leucine-turn-helix structural architecture provided by proline (short helical TAD, yellow box in Fig. 2).

We found the 9aaTAD proximal leucines in transcription factors p53, Oaf1, Pdr1, Gcn4 and also in FOXO3, which is now a new member of the 9aaTAD family (along with HEB and MLL with a fully resolved TAD–KIX complex structure). Next, we found that 9aaTAD proximal leucine is completely missing in artificial peptide KBP2.20 reported[28] and Gal4 (Fig. 2).

Position of the 9aaTAD and 9aaTAD proximal leucine of FOXO3 in the KIX domain complex

We used available FOXO3 structural data (downloaded from PDB, accession number 2LQI) and analysed its structure.[24,25] We found overall structural conformity between MLL 9aaTAD, HEB 9aaTAD and FOXO3 TAD in the KIX complex (Fig. 3). FOXO3 has higher structural similarity to TAD of HEB in the KIX complex (with longer helical structure about 12aa) then to MLL. Especially, FOXO3 Glu251/Leu252 and HEB Glu16/Leu17 share both sequence and are in equal position in the KIX complex.

We then generated sequence alignment of FOXO3, MLL and E2A/HEB TADs by using the structural data of the HEB–KIX, MLL–KIX and FOXO3–KIX complexes (Fig. 2). The FOXO3 corresponding sequence to 9aaTADs (MLL and E2A/HEB) also has two hydrophobic clusters interspersed with one hydrophilic

Fig. 2 The alignment of the structural data for HEB/KIX and MLL/KIX complexes. Using the structural data of the HEB/KIX and MLL/KIX complexes we generated sequence alignment based on structural alignment. Kix domain residues and their position on helices, which are in proximal distance to MLL, HEB and FOXO3 residues, are listed. HEB-Leu17 and MLL-Leu2845 in magenta are in the same position proximal to Ile660 of the KIX α3 helix and therefore aligned together in table. HEB–Glu16 and MLL–Pro2846 in brown are in the same position proximal to Tyr631 of the KIX α2 helix. Both Glu22 of HEB and Glu2851 of MLL are facing towards the solvent. The HEB region from Asp16 to Met26 and the MLL region from Ser2846 to Lys2855 have a helical structure with different length (yellow and red boxes). Positions of amphipathic motif (ξφxxφφ), 9aaTAD domain (blue box) and proximal leucines (in magenta) are shown. Other members of the 9aaTAD family, which interact with the KIX domain, are aligned. Colours used are to identify residues corresponding to colours used in Fig. 1, 3, and 4.
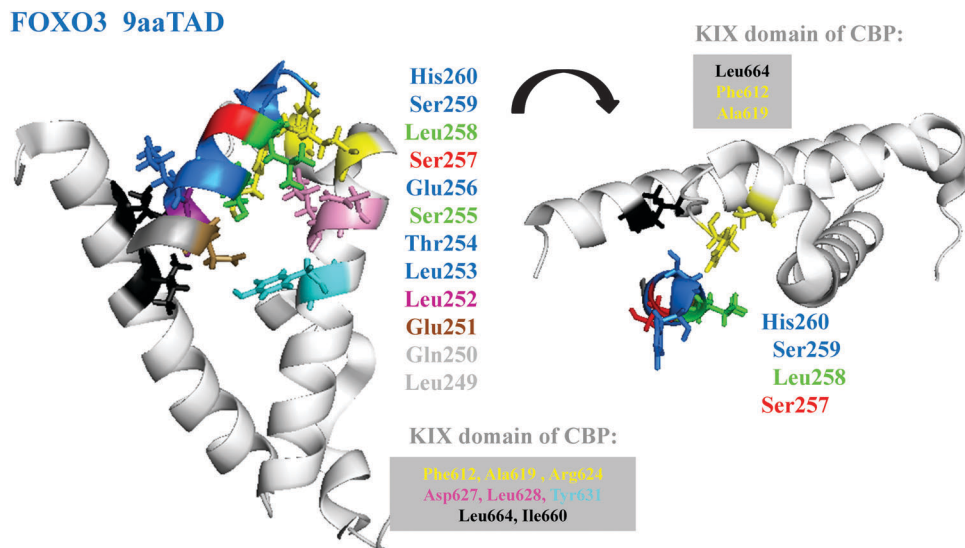
Fig. 3 Shared structure of the FOXO3-KIX complex with the E2A/HEB–KIX and MLL–KIX complexes. The structure coordinates for the FOXO3–KIX complex was downloaded from PDB database (2LQI). The main hydrophobic interaction of FOXO3 9aaTAD is dominated by Leu258 over hydrophilic Ser255 both in green facing the KIX α2-helix. The hydrophilic Ser257 in red is exposed to the solvent and positioned in the middle of the 9aaTAD α-helix. Both Leu249 in magenta and Leu253 in grey face the KIX α3-helix in the opposite direction to Glu251 facing the α2-helix. Distinct residues of the MLL-binding site of the CBP KIX domain in grey are highlighted as follows: Phe612, Ala619 and Arg624 in yellow spreading from the α1 to α2-helix of the KIX domain; Asp627 and Leu628 in pink, Tyr631 in cyan as part of the KIX α2-helix and Leu664 and Ile660 in black as part of the KIX α3-helix.

cluster, and has the proximal leucine, together with all characteristics for 9aaTAD (Fig. 2).

Structural model for p53 interaction with the KIX domain

Using the 9aaTADs sequence alignment data we showed that the first transactivation domain for p53 TAD-I has the highest similarity to 9aaTAD HEB (with longer helical structure about 12aa, red box in Fig. 2). The second transactivation domain p53 TAD-II has the highest similarity to 9aaTAD MLL (with shorter helical structure about 9aa, yellow box in Fig. 2). In the next step, we used reported structural data for the MLL–KIX and E2A/HEB–KIX complexes, respectively, to generate the p53 structural model for their TAD formations in the KIX complex. Using the HEB structural data accession number 2KWF downloaded from PDB database, we replaced the HEB 9aaTAD domain for p53 TAD-I. Similarly, using the MLL structural data accession number 2AGH downloaded from PDB database, we replaced the MLL 9aaTAD domain for p53 TAD-II and generated a model of 9aaTAD domains of p53 binding to the KIX domain (Fig. 4). Two helical regions are predicted in p53 by JPred (17-ETFSDLWKL-25 and 49-DIEQWF-54) (Fig. S1, ESI†) and experimentally observed in p53 complex with another CBP domain (Nuclear Coactivator Binding Domain, NCBD, structural data accession number 2l14 at PDB database, 18-TFSDLWKLL-26 and 46-SPDDIEQWF-54)[36] (Fig. S2, ESI†). From above, the helical structures of both p53 TADs were used in our model.

The main hydrophobic residues of either p53 9aaTAD-I with Phe19 and Leu22 or p53 9aa TAD-II with Ile50 and Trp53 (*in green*) face towards the KIX α2-helix. The hydrophilic residues Asp21 and Gln52 in red of both p53 9aaTADs are exposed to the solvent and positioned in the middle of the 9aaTAD

α-helix. Both Leu14 and Leu45 (*in magenta*) and Gln16 and Ser46 (*in orange*) face towards the KIX α3-helix opposite to Ser15 and Pro47 (*in brown*) facing the α2-helix. Distinct residues of the KIX domain are labelled as follows: Phe612, Ala619 and Arg624 (*in yellow*) spreading from α1 to α2-helix of the KIX domain; Asp627 and Leu628 (*in pink*), Tyr631 (*in cyan*) as part of the KIX α2-helix and Leu664 and Ile660 (*in black*) as part of the KIX α3-helix.

This model showed that the Phe612 (*in yellow*), Arg624 (*in yellow*), Asn627 (*in pink*), Leu664 (*in black*), Ile660 (*in black*) and Tyr631 (*in cyan*) residues of the KIX domain are the most critical in binding of both p53 9aaTAD-I and II domains. Most importantly, our p53 model is in good agreement with the reported chemical shift observed in NMR experiment for the KIX domain upon p53 binding.[6,12] Our model (Fig. 4A and B) clearly shows that all four regions of the KIX domain spreading from α1 to α3 helix labelled in yellow, pink, cyan and black spatially correspond to the reported chemical shift in the KIX domain.[6,12]

Despite the fact that the p53 structure remains unresolved, analysis of MLL and E2A/HEB structural data allowed us to generate the p53 structural models for each 9aaTAD. Both p53 structural models are consistent with positions of major 9aaTAD family determinants. Finally, we generated a model for loss-of-function mutants (Fig. 4C and D) by amino acid substitution of both p53 TADs (L22Q/W23S and W53Q/F54S).[23,37] L22/W23 and W53/F54 represent typical 9aaTAD hydrophobic clusters, which are critical interaction hydrophobic grooves of the KIX domain. Substitution with hydrophilic amino acids (QS) in both positions prevents all possible interaction. What is clearly apparent from our model is the need to eliminate both amino acids in each 9aaTAD simultaneously to achieve complete loss-of-function.
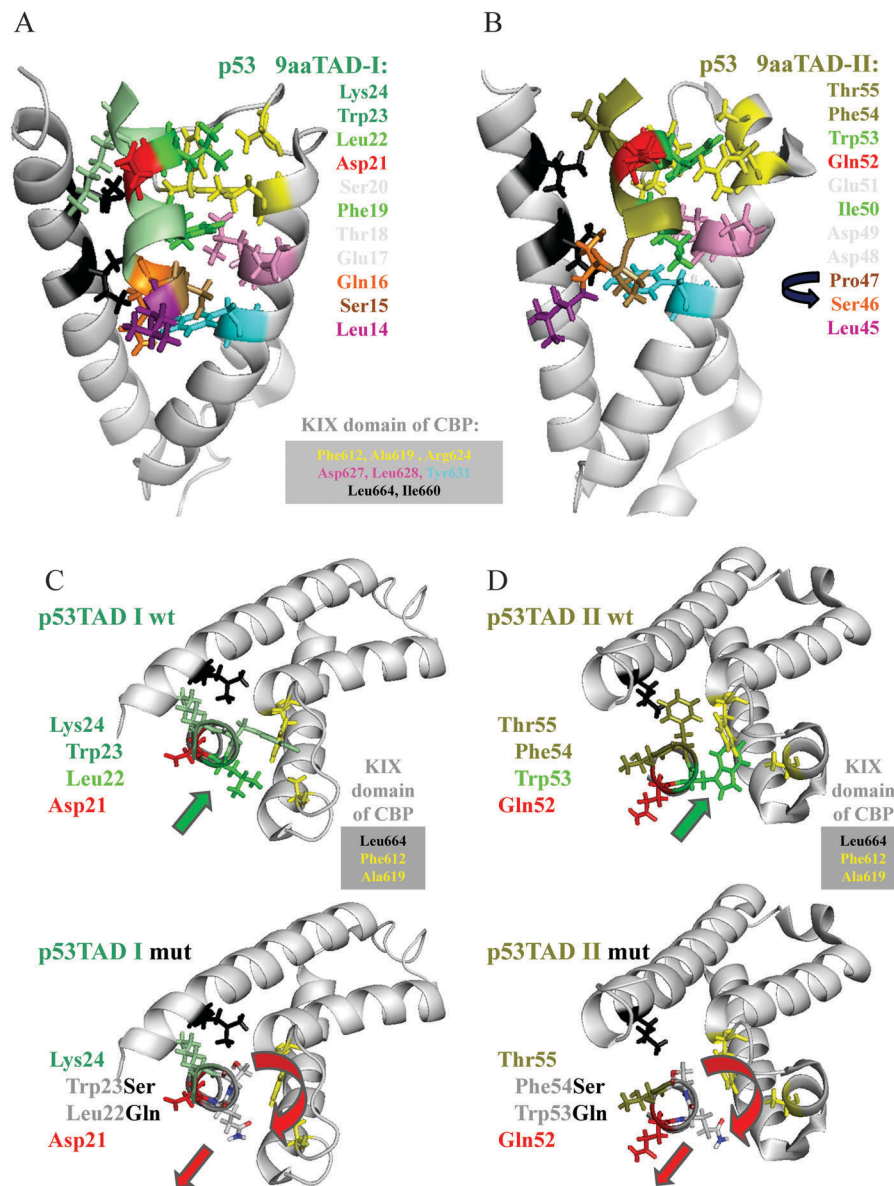
Fig. 4 Structural model of p53 9aaTAD-I and II – derived from structure coordinates for HEB and MLL and positions of their 9aaTADs. The model of 9aaTAD of p53 TAD-I in light-metal green was generated from the resolved HEB structure (2KWF) and 9aaTAD of p53 TAD-II in dark-olive green from the MLL structure (2AGH). The main hydrophobic residues of either p53 9aaTAD-I with Phe19 and Leu22 or p53 9aa TAD-II with Ile50 and Trp53 in green facing the KIX α2-helix. The hydrophilic residues Asp21 and Gln52 in red of both p53 9aaTADs are exposed to the solvent and positioned in the middle of the 9aaTAD α-helix. Both Leu14 and Leu45 in magenta and Gln16 and Ser46 in orange facing the KIX α3-helix opposite to Ser15 and Pro47 in brown facing the α2-helix. Distinct residues of the MLL-binding site of the CBP KIX domain in grey are highlighted as follows: Phe612, Ala619 and Arg624 in yellow spreading from the α1 to α2-helix of the KIX domain; Asp627 and Leu628 in pink, Tyr631 in cyan as part of the KIX α2-helix and Leu664 (A–D) and Ile660 (A, B) in black as part of the KIX α3-helix. The positions of loss-of-function mutations Leu22Gln/Trp23Ser and Trp53Gln/Phe54Ser of both p53 TADs are in black letters. The view is turned so that helical structures of p53 9aaTADs can be seen from the top of their C-terminus (C, D).

9aaTAD patterns derived from structural data of E2A/HEB and MLL

The structural data of the E2A/HEB–KIX and MLL–KIX complexes showed a high degree of structural equivalence and conformity with the annotated 9aaTADs (Fig. 1). From E2A and MLL ortholog alignments, we generated 9aaTAD patterns of the E2A/HEB and MLL family (Fig. S3, ESI†) (9aaTAD positions p1 to p9 for E2A: [DE] [LT] LD [FL] [ST] [MKVAIL] [MV],

and MLL: [SNQ] [DE] [LTI] [LMV] [DE] [FL] [STV] [MKVAIL] [MVKN]). Here, the definition and variability of 9aaTADs is restricted to their structural data and ortholog alignments for E2A/HEB and MLL proteins. The positions p1–p2 with amino acids S–D are conserved in almost all E2A/HEB and MLL orthologs (Fig. S3, ESI†). The important similarity of positions p1–p2 present in p53 9aaTADs (E–T and D–D), Gcn4 9aaTADI (D–A), Pdr1 (E–D) and Gal4 (D–D) is shown in Fig. 2.

The positions p3–p4 represent the hydrophobic cluster I conserved in the E2A family to amino acids L–L and in MLL to I–M/V and similarity is present in Gcn4 9aaTAD-I (V–V). The middle part of the hydrophilic cluster at position p5 is conserved to amino acids D/E in both E2A/HEB and MLL orthologs. Similarity in other 9aaTAD are in p53 9aaTAD-I (D) and also in Gcn4 9aaTAD-I (E). The positions p6–p7 represent the hydrophobic cluster II and are well conserved in E2A/HEB and MLL orthologs (F–S/V). The nearest similarities are in FOXO3 (L–S), in p53 (L/W–F/W) and in Gcn4 9aaTAD-I (S–F). The patterns at position p8–p9 are mostly variable in E2A/HEB and MLL orthologs contrary to Gcn4 9aaTAD-I. In the case that a weaker hydrophobic amino acid is present in positions p6–p7, the position p8 is then part of hydrophobic cluster II as shown for Gcn4, Oaf1 and Gal4 (Fig. 2).

In this study, we have analysed available structural data of the MLL,[6] E2A/HEB,[16] and FOXO3[24,25] transactivation domains bound to the KIX domain. First, we asked whether the position and function of 9aaTAD in MLL–KIX and E2A/HEB–KIX complexes is the same and second, whether the FOXO3 possess the 9aaTAD in the same position in the KIX complex as well.

Our results of the structural overlays of MLL, E2A/HEB and FOXO3 revealed binding equivalence of these transcription factors. The sequence alignments of MLL and E2A/HEB annotated 9aaTADs resulted in complete match with structural data. Analogically to the helical 9aaTADs of MLL and E2A, we identified helical 9aaTAD in FOXO3 in equal position in the KIX complex.

The structural overlay of MLL, E2A/HEB, and FOXO3 revealed equal positions of leucine proximal to the 9aaTAD in the KIX complex. The proximal leucine residues were found at the N-terminus of 9aaTADs and are conserved in some members of the 9aaTAD family such as p53, Gcn4, Oaf1, Pip2, Pdr1/3, ELF3 and EBNA2. Thus, leucines proximal to the 9aaTAD are the characteristic feature of a new subgroup of the family, Leu-9aaTAD. However, these leucines were not found in Gal4, NFAT1, CEBP and KBP2.20 (Fig. 2). This suggests that not all transcription factors of the 9aaTAD family require leucine residues to bind with the KIX domain.

Next, we aimed to determine the position of proximal leucines in the KIX complex and identified two different 9aaTAD structural formations. The 9aaTAD proximal leucine residues, MLL-Leu2845 and HEB-Leu17, shared equal coordinates in the KIX complex and faced towards the Ile660 of the KIX α3 helix, which belongs to the KIX hydrophobic core. Similarly, the second 9aaTAD proximal residues, HEB-Glu16 and MLL-Pro2846, shared equal coordinates in the KIX complex and faced towards the Tyr631of the KIX α2 helix. In E2A 9aaTAD formation, the position of glutamic and leucine residues in the E2A/HEB–KIX complex is defined by the prolonged helical structure encompassing both residues and the 9aaTAD. Contrarily, the position of leucine in the MLL–KIX complex is achieved by leucine-turn-helix structural architecture and is not a part of the helix. The MLL proximal leucine (Leu2845) was reported as essential for KIX domain binding.[5] Together, these results suggest that both proximal leucines and other proximal residues (proline or glutamic residues), that are not part

of 9aaTAD, may significantly contribute to the KIX domain binding.

As previous experimental attempts failed to resolve the p53 structure,[12] we used E2A and MLL structural data, partial p53 structural data, p53 helix predictions and p53 9aaTADs sequence alignments in order to generate the p53 TAD–KIX complex. We created a plausible structural model for both p53 TAD-I and TAD-II in complex with the KIX domain (MLL-site). The E2A-9aaTAD-like structural formation with long helical TAD including both proximal leucine and 9aaTAD (Leu-Ser-Gln-9aaTAD) could be proposed for p53TAD-I, whereas leucine-turn-helix structural architecture (Leu-Ser-Pro-9aaTAD) is more preferable for p53TAD-II derived from MLL-9aaTAD structural formation in the KIX complex. In p53 orthologs, both structural formations could be predicted in each TAD; long helical TAD proposed for human p53TAD-I (including both proximal leucine and 9aaTAD) could be predicted for rabbit (Leu-Ser-Ala-9aaTAD), pig and mouse p53TAD-II (3xLeu-9aaTAD) and vice versa leucine-turn-helix structural architecture proposed for human p53TAD-II could be predicted for chicken p53TAD-I (Leu-Glu-Pro-Thr-9aaTAD) (Fig. S6, ESI†).

Nevertheless, p53TAD-II structural formation could be helical when encompassing proline 47 as reported for another complex (Nuclear Receptor Coactivator Binding Domain).[36] These two p53 structural formations are in good agreement with that reported as "noise" for p53 in the KIX complex,[6,12] which are generally caused by two or more fuzzy binding. Further supporting observation for optional p53 TADII helical structure is modifications of native p53 TAD-II (LSP–DD–IE–Q–WF–TE) into p53 variant E-Cap-LL (LTE–EE–LI–E–LW–FT), which increases p53 helical property and renders 20-fold more potent transcriptional activity.[38] Noteworthily, the E-Cap-LL shows perfect match with the 9aaTAD and resembles the p53 TAD-I (LSQ–ET–FS–D–LW–KL) and also Gcn4 more than the original p53 TAD-II (Fig. 2).

To further evaluate our p53 structural model, we showed that Phe612 and Arg624 residues of the KIX domain are the most critical in binding of both p53 9aaTAD-I and II domains. As reported by Wright lab, these residues showed the most prominent chemical shift in the KIX domain after p53 binding.[6,12] Also, our model clearly explains previously reported loss-of-function mutations L22Q/W23S and W53Q/F54S of both p53 TADs,[23,37] where substitution with hydrophilic amino acids (QS) in both positions prevents all possible interaction and the need to eliminate both amino acids in each 9aaTAD simultaneously to achieve complete loss-of-function. Moreover, similar models could be generated for other 9aaTADs as we have shown already for transcription factor Sox18.[27] Importantly, such structural models offer the means (i) to design loss-of-function mutations, (ii) to design artificial TADs, and (iii) for revision of previous experimental data from the 1990's.

As the KIX domains of CBP are structurally highly similar to the KIX of MED15 (Fig. S4, ESI†), we aimed to compare the MLL–KIX complex with Oaf1–KIX, Pdr1–KIX and Gcn4–KIX of MED15 complexes. The critical CBP region (from F612 to R624) for interaction with MLL[6,12] corresponds to the MED15 region

(from M29 to T41), for which the highest chemical shift was observed after Gcn4.[18] Similarly, although less precise and involving largely diffuse regions, chemical shift of the MED15 KIX domain was observed after Pdr1 and Oaf1 9aaTADs binding.[9,10]

We then focused on Gcn4 that contains clusters of aromatic or bulky hydrophobic residues essential for transactivation function (at positions 9/16, 45/48, 67/69, 89/97/98, 107/110/113, 120/123/124).[1,3] The clusters of hydrophobic amino acid first truly define a large family of transactivation domains.[1,5,6,12,18–23] Two annotated 9aaTADs and two further predicted pseudo 9aaTADs for Gcn4 cover half of those clustered bulky hydrophobic residues (at positions 42–50, 63–71 and 91–99) (Fig. S5, ESI†).[39] Similarly to E2A, Gcn4 9aaTAD-I (at positions 91–99) also has proximal Leu89 to 9aaTAD (sequence: ELD-DAVVESFFS), which is essential for transactivation of Gcn4 TAD-I. As reported previously, the Gcn4 has the second 9aaTAD-II at positions 118–126, which has an unusual Lys and is out of predictive recognition by the 9aaTAD pattern (sequence: KEWTSLFDN).[39] We identified a number of variations of Leu-9aaTAD in the Gcn4 orthologs (Fig. S6, ESI†), which clearly represent the natural variability of the 9aaTAD family. Noteworthily, using JPred, we predicted a helical structure for Gcn4 9aaTAD-I (at position 89–99, data not shown). The second helical structure for Gcn4 9aaTAD-II was reported despite extensive experimental "noise", the helical structure was identified to position 116–125.[40] These data resemble our reported 9aaTAD-loop-9aaTAD motifs of p53[41] and reported structure.[36]

Similarly to p53, the Gcn4–KIX model could be generated to interpret numerous mutants and to design more powerful Gcn4 variants from (i) Gcn4 9aaTAD-I similarity to E2A (Fig. 2), (ii) structural overlay of KIX/CBP and KIX/MED15 complexes (Fig. S4, ESI†), (iii) E2A resolved structure (Protein Data Bank, accession number 2KWF), (iv) Gcn4 9aaTAD-I predicted helical structure (JPred, at position 89–99), (v) altered structure of KIX/MED15 by Gcn4 binding,[18] and (vi) Gcn4 9aaTAD-I residues, which are critical for transactivation (positions 89/97/98).[1,3]

In conclusion, the 9aaTAD of either E2A or MLL has significant similarities, but largely or completely differ from other 9aaTADs (FOXO3, p53, Oaf1, Pdr1 and Gcn4) that bind to the KIX domain (Fig. S3 and S6, ESI†). Furthermore, we suppose that conservation of 9aaTAD within one protein family reflects more likely specific structural and functional protein requirements (e.g. intramolecular masking) then 9aaTAD sequence requirements for transactivation function (e.g. binding to multiple mediators such as CBP, MED15 and TAF9).

The 9aaTAD could not be defined solely by primary amino acid sequences (9aaTAD pattern, Fig. S3 and S6, ESI†), but rather by two hydrophobic clusters that are interspersed with one hydrophilic cluster.[26] The positions of hydrophobic clusters in the 9aaTADs (position p3–p4 and p6–p8) were derived from Oaf1 homologs[26] and resemble Phenylalanine clusters reported for Gcn4.[1] The middle hydrophilic cluster (around position p4–p6) is not strictly acidic as reported for Gal4,[2] but rather the hydrophilic region comprised of one to three amino acids. The hydrophobic interaction between 9aaTAD and the KIX domain is provided by two hydrophobic clusters (Fig. 1). The opposite side of the 9aaTAD helix is hydrophilic and provides interaction

with the solvent. This asymmetric hydrophobic and hydrophilic distribution in 9aaTAD secures driving forces for the TAD–KIX interaction.

Because of variation in the size of these clusters, the position of 9aaTAD could be either determined by structural studies or predicted by ortholog alignments with one inexact aa position (aa + 8aa or 8aa + aa = 9aaTAD). Nevertheless, fuzzy motifs such as 9aaTAD and their predictions are largely illusive but they can be still useful and are also reported by others.[42,43] The 9aaTAD prediction service is available online (www.piskacek.org), which helps solely to identify the clustered hydrophobic and hydrophilic amino acids with distinct distribution typical for 9aaTADs described above and also have potential to identify other TADs. The families of transcription factor E2A, MLL, p53, Oaf1, Pdr1 and Gcn4 build a library of KIX binding 9aaTADs (Fig. S6, ESI†), which are based on shared composition and structural features in the KIX complexes.

Conclusions

We promote in this study the 9aaTAD–KIX complex structural model, which gains the first molecular understanding for empirical observations and sheds more light on the ambiguous 9aaTAD domain. Such structural models may be useful for the 9aaTAD prediction, design of loss-of-function mutations, artificial TADs, and comprehensively define TAD structural requirements and amino acid involvement, sequence variations for the 9aaTAD family and final definition of transactivation domains pursued over the years.

Materials and methods

Structural data

The structure coordinates for the KIX domain interaction have been downloaded from the Protein Data Bank for HEB/KIX (PDB 2KWF) referring to TCF4 (11-GTDKELSDLLDFSAMF-26), which sequence is identical to HEB/TCF12 (12-GTDKELSDLLDFSAMF-27) and similar to E2A/TCF3 (11-GTDKELSDLLDFSMMF-26) and MLL/KIX (PDB 2AGH). Helical structures were predicted by JPred3 at The University of Dundee (www.compbio.dundee.ac.uk/www-jpred/).[44] Molecular graphics were generated using PyMOL (www.pymol.org).

Prediction of the 9aaTAD

The 9aaTAD prediction service and Service description were previously available at EMBnet/ExPASy (2007–2012) and now have moved to www.piskacek.org. The 9aaTAD pattern was employed using the ScanProsite algorithm and refinement criteria CR1-12 defined for the 9aaTAD domains were applied as described previously.[26]

Acknowledgements

References

1 B. M. Jackson, C. M. Drysdale, K. Natarajan and A. G. Hinnebusch, *Mol. Cell. Biol.*, 1996, **16**, 5557–5571.

2 J. Ma and M. Ptashne, *Cell*, 1987, **51**, 113–119.

3 C. M. Drysdale, E. Dueñas, B. M. Jackson, U. Reusser, G. H. Braus and A. G. Hinnebusch, *Mol. Cell. Biol.*, 1995, **15**, 1220–1233.

4 J. S. Fassler and F. Winston, *Mol. Cell. Biol.*, 1989, **9**, 5602–5609.

5 P. Ernst, J. Wang, M. Huang, R. H. Goodman and S. J. Korsmeyer, *Mol. Cell. Biol.*, 2001, **21**, 2249–2258.

6 N. K. Goto, T. Zor, M. Martinez-Yamout, H. J. Dyson and P. E. Wright, *J. Biol. Chem.*, 2002, **277**, 43168–43174.

7 M. Tollinger, K. Kloiber, B. Agoston, C. Dorigoni, R. Lichtenecker, W. Schmid and R. Konrat, *Biochemistry*, 2006, **45**, 8885–8893.

8 D. P. Teufel, S. M. Freund, M. Bycroft and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7009–7014.

9 J. K. Thakur, H. Arthanari, F. Yang, S.-J. Pan, X. Fan, J. Breger, D. P. Frueh, K. Gulshan, D. K. Li, E. Mylonakis, K. Struhl, W. S. Moye-Rowley, B. P. Cormack, G. Wagner and A. M. Näär, *Nature*, 2008, **452**, 604–609.

10 J. K. Thakur, H. Arthanari, F. Yang, K. H. Chau, G. Wagner and A. M. Näär, *J. Biol. Chem.*, 2009, **284**, 4422–4428.

11 P. Schanda, B. Brutscher, R. Konrat and M. Tollinger, *J. Mol. Biol.*, 2008, **380**, 726–741.

12 C. W. Lee, M. Arai, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *Biochemistry*, 2009, **48**, 2115–2124.

13 C. W. Lee, J. C. Ferreon, A. C. M. Ferreon, M. Arai and P. E. Wright, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 19290–19295.

14 P. R. Cook, N. Polakowski and I. Lemasson, *J. Mol. Biol.*, 2011, **409**, 384–398.

15 W. C. Pomerantz, N. Wang, A. K. Lipinski, R. Wang, T. Cierpicki and A. K. Mapp, *ACS Chem. Biol.*, 2012, **7**, 1345–1350.

16 C. M. Denis, S. Chitayat, M. J. Plevin, F. Wang, P. Thompson, S. Li, H. L. Spencer, M. Ikura, D. P. Lebrun and S. P. Smith, *Blood*, 2012, **120**, 3968–3977.

17 S. Gianni, A. Morrone, R. Giri and M. Brunori, *Biochem. Biophys. Res. Commun.*, 2012, **428**, 205–209.

18 I. Jedidi, F. Zhang, H. Qiu, S. J. Stahl, I. Palmer, J. D. Kaufman, P. S. Nadaud, S. Mukherjee, P. T. Wingfield, C. P. Jaroniec and A. G. Hinnebusch, *J. Biol. Chem.*, 2010, **285**, 2438–2455.

19 R. N. De Guzman, N. K. Goto, H. J. Dyson and P. E. Wright, *J. Mol. Biol.*, 2006, **355**, 1005–1013.

20 M. Arai, H. J. Dyson and P. E. Wright, *FEBS Lett.*, 2010, **584**, 4500–4504.

21 M. Uesugi, O. Nyanguile, H. Lu, A. J. Levine and G. L. Verdine, *Science*, 1997, **277**, 1310–1313.

22 M. Uesugi and G. L. Verdine, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 14801–14806.

23 J. Lin, J. Chen, B. Elenbaas and A. J. Levine, *Genes Dev.*, 1994, **8**, 1235–1246.

24 F. Wang, C. B. Marshall, G.-Y. Li, K. Yamamoto, T. W. Mak and M. Ikura, *ACS Chem. Biol.*, 2009, **4**, 1017–1027.

25 F. Wang, C. B. Marshall, K. Yamamoto, G.-Y. Li, G. M. C. Gasmi-Seabrook, H. Okada, T. W. Mak and M. Ikura, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 6078–6083.

26 S. Piskacek, M. Gregor, M. Nemethova, M. Grabner, P. Kovarik and M. Piskacek, *Genomics*, 2007, **89**, 756–768.

27 J. Sandholzer, M. Hoeth, M. Piskacek, H. Mayer and R. de Martin, *Biochem. Biophys. Res. Commun.*, 2007, **360**, 370–374.

28 J. V. Frangioni, L. M. LaRiccia, L. C. Cantley and M. R. Montminy, *Nat. Biotechnol.*, 2000, **18**, 1080–1085.

29 C. Y. Majmudar, B. Wang, J. K. Lum, K. Håkansson and A. K. Mapp, *Angew. Chem., Int. Ed.*, 2009, **48**, 7021–7024.

30 A. C. Joerger and A. R. Fersht, *Annu. Rev. Biochem.*, 2008, **77**, 557–582.

31 M. Bista, S. M. Freund and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 15752–15756.

32 J. Zhang, M. Kalkum, S. Yamamura, B. T. Chait and R. G. Roeder, *Science*, 2004, **305**, 1286–1289.

33 M. E. Massari, P. A. Grant, M. G. Pray-Grant, S. L. Berger, J. L. Workman and C. Murre, *Mol. Cell*, 1999, **4**, 63–73.

34 M. E. Massari, P. A. Jennings and C. Murre, *Mol. Cell. Biol.*, 1996, **16**, 121–129.

35 S. Brüschweiler, R. Konrat and M. Tollinger, *ACS Chem. Biol.*, 2013, **8**, 1600–1610.

36 C. W. Lee, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *Biochemistry*, 2010, **49**, 9964–9971.

37 C. Venot, M. Maratrat, V. Sierra, E. Conseiller and L. Debussche, *Oncogene*, 1999, **18**, 2405–2410.

38 C. Langlois, A. Del Gatto, G. Arseneault, J. Lafrance-Vanasse, M. De Simone, T. Morse, I. de Paola, M. Lussier-Price, P. Legault, C. Pedone, L. Zaccaro and J. G. Omichinski, *J. Am. Chem. Soc.*, 2012, **134**, 1715–1723.

39 M. Piskacek, *Nature Precedings*, 2009, DOI: 10.1038/npre.2009.3984.1.

40 P. S. Brzovic, C. C. Heikaus, L. Kisselev, R. Vernon, E. Herbig, D. Pacheco, L. Warfield, P. Littlefield, D. Baker, R. E. Klevit and S. Hahn, *Mol. Cell*, 2011, **44**, 942–953.

41 M. Piskacek, *Nature Precedings*, 2009, DOI: 10.1038/npre.2009.3488.2.

42 A. Ernst, B. A. Appleton, Y. Ivarsson, Y. Zhang, D. Gfeller, C. Wiesmann and S. S. Sidhu, *J. Mol. Biol.*, 2014, **426**, 3509–3519.

43 S. Hui, X. Xing and G. D. Bader, *BMC Bioinf.*, 2013, **14**, 27.

44 C. Cole, J. D. Barber and G. J. Barton, *Nucleic Acids Res.*, 2008, **36**, W197–W201.