



Published in final edited form as:

Int J Dev Biol. 2010 ; 54(11-12): 1589–1596. doi:10.1387/ijdb.103194hh.

Structure and functions of powerful transactivators: VP16, MyoD and FoxA

HIROYUKI HIRAI¹, TETSUYA TANI², and NOBUAKI KIKYO^{1,*}

¹Stem Cell Institute, Division of Hematology, Oncology and Transplantation, Department of Medicine, University of Minnesota, USA

²Laboratory of Animal Reproduction, Department of Agriculture, Kinki University, Nara, Japan

Abstract

Induced pluripotent stem cell (iPSC) technology is a promising approach for converting one type of a differentiated cell into another type of differentiated cell through a pluripotent state as an intermediate step. Recent studies, however, indicate the possibility of directly converting one cell type to another without going through a pluripotent state. This direct reprogramming approach is dependent on a combination of highly potent transcription factors for cell-type conversion, presumably skipping more physiological and multi-step differentiation processes. A trial-and-error strategy is commonly used to screen many candidate transcription factors to identify the correct combination of factors. We speculate, however, that a better understanding of the functional mechanisms of exemplary transcriptional activators will facilitate the identification of novel factor combinations capable of direct reprogramming. The purpose of this review is to critically examine the literature on three highly potent transcriptional activators: the herpes virus protein, VP16; the master regulator of skeletal muscle differentiation, MyoD and the “pioneer” factor for hepatogenesis, FoxA. We discuss the roles of their functional protein domains, interacting partners and chromatin remodeling mechanisms during gene activation to understand how these factors open the chromatin of inactive genes and reset the transcriptional pattern during cell type conversion.

Keywords

FoxA; MyoD; nuclear reprogramming; transcriptional activation domain; VP16

Introduction

Current research on the reprogramming of cell types by using defined genes started with the discovery MyoD, a master transcription factor for skeletal muscle differentiation (Davis *et al.*, 1987; Graf and Enver, 2009; Zhou and Melton, 2008). MyoD activates skeletal muscle-specific genes in non-muscle cells and, in some cases, converts differentiated cells into muscle cells without additional exogenous genes (see below for more discussion). Following the discovery of MyoD, several reports documented the conversion of one cell type into another, largely within hematopoietic lineages, by introducing one or a few genes. For instance, enforced expression of the transcription factor C/EBP α converts pre-B cells and pre-T cells into macrophages (Bussmann *et al.*, 2009; Laiosa *et al.*, 2006). However, cell-type conversion by using defined genes is most prominently exemplified by the

establishment of induced pluripotent stem cells (iPSCs), in which four genes (Oct4, Sox2, Klf4 and c-Myc) were used to convert fibroblasts into pluripotent cells (Takahashi and Yamanaka, 2006). This strategy of using defined genes was subsequently employed to reprogram pancreatic exocrine cells into β cells with three genes and to reprogram fibroblasts into neurons or cardiac muscle cells with other three genes (Ieda *et al.*, 2010; Vierbuchen *et al.*, 2010; Zhou *et al.*, 2008). These spectacular studies have taught us that the reprogramming of cell fate (nuclear reprogramming) can be achieved more easily than previously thought once the right combination of transcription factors is identified.

As successfully employed for iPSCs, the first step in the search for the right combination of factors is to identify a list of all candidate genes based on existing information about the genes, such as their expression profile and whether they are required for the development of target cells in a physiological setting. A general strategy to identify candidate transcription factors for nuclear reprogramming is discussed in detail by Zhou and Melton (Zhou and Melton, 2008). The next step is to introduce the genes into host cells in various combinations until the expected cell conversion is observed. Cell conversion is usually determined by identifying cell changes that are representative of the destination cell type, such as activation of key genes, morphological changes and acquisition of new functions. Once some of these parameters are met, the third step is to refine the factor combinations by reducing the number of factors required for conversion until a minimum number is identified. Finally, the authenticity of the destination cells is validated by thoroughly comparing them to their physiological counterpart at the global gene expression level and functional level. This general strategy obviously relies on the assumption that the initial gene list contains all the necessary genes for the cell-fate conversion. Thus, successful artificial nuclear reprogramming is largely dependent on two conditions: initial inclusion of the right transcription factor genes and establishment of an efficient and reliable assay system.

Decisions on which genes to include in the initial list would be improved by a better understanding of the general structure of DNA-binding transcription factors and why some transcription factors are exceptionally more potent than others. DNA-binding transcription factors generally contain a minimum of two domains: a DNA-binding domain (DBD) and a transcriptional activation domain (TAD) (Latchman, 2008; Ptashne and Gann, 2002). The DBD directs the transcription factor to the target gene by recognizing a specific DNA sequence. The DBD is categorized by a handful of structurally conserved motifs, such as the basic helix-loop-helix (bHLH), zinc finger and winged helix. Once recruited to the right target genes, the TAD serves as a scaffold to recruit and assemble additional transcription factors and chromatin remodeling proteins to initiate transcription. In contrast to the DBD, the structural motifs of the TADs are not well defined. The most common, and thus best studied, TAD is characterized by an abundance of acidic and hydrophobic amino acids, but the exact amino acid sequence is quite variable. Overall negative charge, rather than the exact sequence, appears to be the crucial determinant of the potency of the TADs. Reflecting this sequence variability, the acidic TADs do not take a specific conserved three dimensional structure in the absence of binding partners; however, when bound to interacting proteins, they will adopt an α -helical conformation as described in the VP16 section below. Two other TAD motifs, the proline-rich domain and the glutamine-rich domain, are weaker gene activators and less well characterized. Importantly, the TADs can be physically separated from the DBD and linked to heterologous DBDs to activate new target genes. For instance, the fusion protein containing the DBD of the yeast transcription factor Gal4 and the TAD of the herpes simplex virus transcription factor VP16 (Gal4-VP16) activates target genes that contain Gal4-binding sites in their promoters (Sadowski *et al.*, 1988). This system and its variants are frequently used as a potent transcriptional activation (transactivation) model in gene reporter assays and yeast two hybrid assays.

An important question that has emerged from the study of TADs is why some are such powerful gene activators that they can initiate the program of cell differentiation. In the current review article, we will discuss three highly potent transactivators – VP16, MyoD and FoxA – and examine the roles of their functional protein domains and interacting partners during gene activation. VP16 is one of the most extensively studied transactivators and serves as a prototype in understanding how transactivators regulate target gene activity. The skeletal muscle-specific transactivator MyoD can activate muscle genes embedded in closed chromatin thereby initiating conversion of non-muscle cells into muscle cells. MyoD is the best characterized transactivator capable of inducing nuclear reprogramming. FoxA is another example of a transactivator that can recognize its target DNA sequence in closed chromatin earlier than any other relevant factors, in this case, during liver development. Unlike MyoD, the unique structure of the DBD of FoxA can potentially explain why this protein can gain access to closed chromatin more efficiently than other transactivators. Understanding the exceptional potency of these factors is important to efforts to improve nuclear reprogramming which is largely dependent on the capability of transactivators to activate a set of new genes embedded in closed chromatin.

VP16: prototype transactivator

VP16 is a transcription factor of herpes simplex virus (HSV) type 1 that is involved in the activation of the viral immediate-early genes (Flint and Shenk, 1997; Wysocka and Herr, 2003). VP16 has 490 amino acids with a core domain in its central region required for indirect DNA binding and a carboxy-terminal TAD located within its last 81 amino acids (Fig. 1A) (Greaves and O'Hare, 1989; Triezenberg *et al.*, 1988). VP16 is originally contained within the virion (virus particle) of the HSV and released into animal cells upon infection. VP16 first binds to the host nuclear protein HCF through its core domain and subsequently binds to another host nuclear protein Oct-1 to form a three-component protein complex (Fig. 1B). This complex then binds to its target DNA sequence TAATGARAT (R is a purine) in the promoters of immediate-early genes. This is achieved through interactions between Oct-1 and the target DNA sequence or a consensus octamer motif that overlaps the 5' portion of this sequence. HCF then stabilizes the interaction between VP16 and Oct1. Once recruited to immediate-early genes, VP16 activates genes through interactions between the TAD and many of the transcription factors described below.

The TAD of VP16 is one of the most potent TADs and is widely fused to host transcription factors to amplify their activity. For instance, the pancreatic transcription factor Pdx1 fused with the VP16 TAD was about 10 times more effective than wild-type Pdx1 in a luciferase reporter assay (Kaneto *et al.*, 2005). VP16 TAD is also frequently fused to DBDs of other transcription factors to study the mechanism of gene regulation. In addition to the GAL4-VP16 fusion protein mentioned earlier, this TAD has been fused with MyoD for domain analyses of the MyoD protein (see below).

The VP16 TAD interacts with numerous proteins involved in gene activation (Fig. 1B). For instance, the interacting partners include basal transcription factors, such as TFIIA (Kobayashi *et al.*, 1995), TFIIB (Lin *et al.*, 1991), TFIIF (Zhu *et al.*, 1994), TFIIF (Xiao *et al.*, 1994) and subunits of TFIID including TBP (Stringer *et al. et al.*, 1990), hTAF_{II}31 (Uesugi, 1997) and hTAF_{II}32 (Klemm *et al.*, 1995). This TAD also binds to the Mediator complex through direct interactions with the two subunits MED17 (Ito *et al.*, 1999) and MED25 (Mittler *et al.*, 2003). Furthermore, the VP16 TAD interacts with the general cofactor PC4 (Ge and Roeder, 1994; Kretzschmar *et al.*, 1994). These multiple interactions indicate that the VP16 TAD facilitates the assembly of the pre-initiation complex at several distinct steps (Choy and Green, 1993). The VP16 TAD also recruits histone acetyltransferases – such as the SAGA complex and the NuA4 complex (Utley *et al.*, 1998;

Vignali *et al.*, 2000), p300 (Kundu *et al.*, 2000) and PCAF (Tumbar *et al.*, 1999), as well as the SWI/SNF ATPase complex – to promoters (Neely *et al.*, 1999). Chromatin decondensation by VP16 during gene activation appears to have a wide-spread effect since it can be detected by fluorescence microscopes (Carpenter *et al.*, 2005; Tumbar *et al.*, 1999).

The VP16 TAD does not usually take a specific three dimensional structure on its own; however, it will adopt an α -helical conformation upon binding to target proteins, such as TFIID (Shen *et al.*, 1996; Uesugi *et al.*, 1997). The TAD of VP16 can be divided into two independent regions, H1 (amino acids 410–452) and H2 (453–490) with distinct protein-protein interactions (Walker *et al.*, 1993). While the H2 region is dispensable for binding to TFIIB (Lin *et al.*, 1991), it is necessary for binding to TFIID (Ingles *et al.*, 1991), TFIIF (Xiao *et al.*, 1994) and CBP (Ikeda *et al.*, 2002).

This wide spectrum of binding partners is not a specific feature to VP16. The TAD of the Tax protein of human T-lymphotropic virus type 1 (HTLV-1) interacts with basal transcription factors (TFIIA and TFIID), histone acetyltransferases (p300/CBP and PCAF), components of the SWI/SNF complexes (BRG1, BAF53, BAF57 and BAF155), the histone methyltransferase SUVH39H1 (methylation of lysine 9 on histone H3) and the histone demethylase JMJD2A (demethylation of lysines 9 and 36 on histone H3) (Boxus *et al.*, 2008). The TAD of the Tat transactivator of human immunodeficiency virus type 1 (HIV-1) also binds to TFIID, p300/CBP and PCAF (Pumfery *et al.*, 2003; Romani *et al.*, 2010). The Gal4 TAD also interacts with TFIIB, TFIID, the SAGA complex and the SWI/SNF complex (Traven *et al.*, 2006). Thus, binding to some of the basal transcription factors, histone acetyltransferases and the SWI/SNF complexes is a common feature of TADs. Nonetheless, these transactivators display a distinct potency for transactivation depending on the assay conditions. For instance, VP16 transcribes a target gene with more than 100-fold higher efficiency than Tat in a specific reporter assay (Blau *et al.*, 1996). VP16 and Tat exert a synergistic effect in another reporter assay, suggesting that they function in different manners (Ghosh *et al.*, 1993). Although the TADs of VP16, Tax and Tat belong to an acidic transactivator group, they do not share homology of amino acid sequences. The differential activity and regulation among different TADs needs to be empirically determined.

MyoD: master regulator of muscle differentiation

MyoD is a classic example of a master control gene for cell differentiation in the sense that transduction of this gene is sufficient to activate the whole genetic program of muscle differentiation in non-muscle cells (Berkes and Tapscott, 2005; Tapscott, 2005). Indeed, the key to identifying this gene was its self-sufficient nature. Treatment of the mouse embryonic fibroblast cell line C3H10T1/2 with 5-azacytidine (5-azaC), an inhibitor of DNA methylation, induced differentiation of these cells to skeletal muscle cells (Constantinides *et al.*, 1977). This observation led to the hypothesis that demethylation and the resulting activation of unidentified genes were responsible for the conversion to muscle cells. Subsequently, a transfection experiment with genomic DNA fragments prepared from 5-azaC-induced myoblasts suggested that only one gene that was activated by 5-azaC was sufficient for the conversion (Lassar *et al.*, 1986). Finally, a subtractive hybridization experiment comparing untreated and 5-azaC-treated fibroblasts led to the identification of a cDNA encoding MyoD (Davis *et al.*, 1987). The myogenic function of MyoD was confirmed by the conversion of several fibroblast cell lines to skeletal muscle cells after transfection of the gene.

Several studies have examined the ability of MyoD to fully activate the muscle differentiation program using both *in vitro* and *in vivo* model systems. Retroviral transduction of MyoD was found to activate such skeletal muscle-specific genes as desmin

and myosin heavy chain in non-muscle cells, including melanoma and neuroblastoma cells (Weintraub *et al.*, 1989). During this process MyoD does not suppress the genes specific to the parent cells; thus, the MyoD-induced muscle program and the parent cell-specific program co-exist in these cells. In some cases, however, MyoD fully converts non-muscle cells, such as pigmented epithelial cells, to functional myotubes in about 10 days (Choi *et al.*, 1990). The frequency of the conversion is not high, ranging from 1 to 5% of cells, as calculated by dividing the number of nuclei within the myotubes, which contain multiple nuclei due to cell fusion, by the total number of nuclei in a given culture dish. Although MyoD demonstrates an impressive capability for activating muscle-specific genes in tissue culture cells, the same is not necessarily true when ectopically expressed in embryos. For instance, injection of MyoD mRNA into *Xenopus* embryos activates muscle-specific genes in prospective endoderm cells but fails to initiate muscle differentiation (Hopwood and Gurdon, 1990). Similarly, micro-injection of MyoD cDNA into fertilized mouse oocytes activates some muscle genes in non-muscle cells but does not induce overt myogenic conversion of the cells (Faerman *et al.*, 1993). The current interpretation of these results is that MyoD activates some of its target genes when overexpressed but additional signaling mechanisms are required to activate the entire muscle differentiation program *in vivo*. This conclusion is not surprising given the complexity of cell-cell interactions *in vivo*, unlike the homogenous and simple culture system *in vitro*.

MyoD is one of the four members of the bHLH proteins that are involved in skeletal muscle differentiation (Pownall *et al.*, 2002). During the early phase of myogenesis, MyoD and Myf5 are essential for the establishment and maintenance of myogenic precursor cells. Myogenin is important for myogenic terminal differentiation, and MRF4 is involved in both early and terminal myogenesis. All four myogenic bHLH proteins contain four conserved domains: a TAD in the amino terminal region, a histidine/cysteine rich domain (H/C domain) and a bHLH in the central region and an amphipathic α helix domain (helix III) in the carboxy terminal region (Fig. 2A) (Berkes *et al.*, 2004; Gerber *et al.*, 1997). The basic amino acids in the bHLH are important for DNA binding as well as for the maintenance of the TAD conformation, which is critical for the activation of muscle genes (Brennan *et al.*, 1991; Huang *et al.*, 1998; Molkentin *et al.*, 1996). In addition, the basic region contains an alanine-threonine sequence called the myogenic code, which is conserved in all myogenic bHLH proteins from worm to human and allows myogenic bHLH proteins to specifically activate muscle genes (Brennan *et al.*, 1991; Davis *et al.*, 1990). When the corresponding sequence of the non-myogenic bHLH protein E12 is replaced with alanine-threonine/lysine, this replacement confers myogenic capability to E12 (Davis and Weintraub, 1992).

The HLH is responsible for homo- or hetero-dimerization with ubiquitously expressed bHLH proteins called E2A proteins. The dimerized bHLH proteins then bind to the E box consensus sequence (CANNTG), which is found in the promoters and enhancers of muscle-specific genes. The TAD, which contains 12 acidic amino acids out of a total of 54 amino acids, belongs to the classic acidic TAD groups. This TAD can activate a reporter gene when fused to the DBD of the Gal4 protein (Weintraub *et al.*, 1991). Importantly, this fusion protein can activate the reporter gene 20-fold higher than the fusion protein of full-length MyoD and the Gal4 domain, suggesting that inhibitory roles are played by the non-TAD regions of MyoD. It is not known whether this inhibitory effect is due to specific interactions between the non-TAD regions and other molecules.

The TAD is also important for the interaction between MyoD and the histone acetyltransferase p300/CBP (Sartorelli *et al.*, 1997). MyoD interacts with two histone acetyltransferases, PCAF and p300/CBP, which play distinct roles in transactivation by MyoD. Although p300 and CBP are different proteins, they are frequently referred to as p300/CBP and thought of as a single entity because they are considered homologs. Both

PCAF and p300/CBP can acetylate lysines in MyoD, which increases the affinity of MyoD to its target DNA sequences as well as to p300/CBP (Polesskaya and Harel-Bellan, 2001; Polesskaya *et al.*, 2001; Sartorelli *et al.*, 1999). p300/CBP then acetylates histones H3 and H4 at the MyoD-binding loci, facilitating gene activation (Dilworth *et al.*, 2004). Besides binding to histone acetyltransferases, the TAD of VP16 and that of MyoD share similarities and differences. For instance, the TAD of VP16 functionally replaces that of MyoD in a reporter assay, but the VP16 TAD is not influenced by the presence of the basic amino acids in the bHLH, unlike the MyoD TAD (Weintraub *et al.*, 1991). While transactivation by the MyoD TAD is suppressed by the MEK1 kinase, the VP16 TAD is not influenced by the kinase (Perry *et al.*, 2001).

MyoD and Myf5 can remodel the chromatin of suppressed muscle genes and activate the genes more efficiently than myogenin. The source of this difference primarily lies in the sequences of the H/C region and the helix III region (Bergstrom and Tapscott, 2001; Gerber *et al.*, 1997). These two regions allow MyoD to bind to its target promoters through interaction with an adjacent complex containing the homeodomain proteins, Pbx and Meis (Berkes *et al.*, 2004). However, the H/C domain and helix III appear to be essential for activation of only a subset of MyoD target genes (de la Serna *et al.*, 2001); it remains to be studied how other target genes are differentially activated by MyoD and myogenin.

To integrate our understanding of these protein interactions, the following cascade of chromatin events has been proposed during target gene activation by MyoD (Fig. 2B) (de la Serna *et al.*, 2005). MyoD is recruited to its target gene promoters via interaction with the Pbx-Meis complex which is constitutively bound to the genes in fibroblasts and myoblasts. PCAF and/or p300/CBP are recruited to the target promoters through MyoD and then acetylate histones H3 and H4, relaxing chromatin. Subsequently, a BRG1-containing SWI/SNF complex is also recruited to the promoters which remodels nucleosomes and stabilizes the DNA binding of MyoD. The order of events may vary depending on the MyoD target genes; nonetheless, this model provides a framework to further investigate the detailed mechanisms of gene activation by MyoD. It is not known whether the Pbx-Meis complex constitutively exists at muscle gene promoters in other non-muscle cells. If it is not present, another mechanism must exist by which MyoD detects its target genes while they are still embedded in closed chromatin. The nature of this mechanism is one of the remaining and most critical questions in gene activation by MyoD.

FoxA: pioneer factor for hepatic differentiation

VP16 and MyoD use their TADs as the primary domain for the activation of their target genes. In contrast, the FoxA family proteins use their DBDs as the central player for gene activation. The FoxA proteins belong to the large forkhead box (Fox) gene family characterized by the presence of a winged helix DBD in its central region (Carlsson and Mahlapuu, 2002; Friedman and Kaestner, 2006). This domain comprises three α -helices placed in a helix-turn-helix configuration flanked by a loop on each side, like wings. The terms –winged helix domain proteins and forkhead box proteins – are interchangeably used. Forkhead is named after the *Drosophila* gene *forkhead* whose mutation displays defects in head fold involution (Weigel *et al.*, 1989). Because mammalian FoxA proteins were first identified in the liver cell nucleus as hepatocyte nuclear factor-3 (HNF-3), FoxA1, 2 and 3 used to be called HNF-3 α , β and γ , respectively, until 2000 when the standardized nomenclature was introduced (Kaestner *et al.*, 2000).

The FoxA proteins regulate many liver-specific genes and control the development of liver as well as the metabolism of adult liver (Friedman and Kaestner, 2006; Zaret, 2008). Double-knockout mice of Foxa1 and 2 (“a” is used for mouse nomenclature) completely

lack liver from the earliest stage of development, with the liver bud failing to form in the foregut region and no expression of α fetoprotein, an early marker of hepatogenesis (Lee *et al.*, 2005). At the molecular level, *in vivo* footprinting experiments detected protein occupancy of the FoxA-binding site and the GATA4-binding site in the enhancer of the liver-specific albumin gene, supposedly by FoxA and GATA4, respectively, in the liver precursor region of the gut. Chromatin binding of these proteins takes place earlier than any other proteins in the region and before transcription of the albumin gene commences (Bossard and Zaret, 1998; Gualdi *et al.*, 1996). In addition, FoxA opens compacted chromatin in nucleosome arrays containing the albumin enhancer *in vitro*, independent of the SWI/SNF chromatin remodeling ATPases which are frequently required for chromatin opening (Cirillo *et al.*, 2002). Because of its ability to open chromatin and thus increase the accessibility of other transcription factors to their target genes, FoxA is called a “pioneer” factor (Zaret, 2008).

FoxA proteins contain a DBD in a central winged helix domain and a TAD in both its amino terminal region and carboxy terminal region (Fig. 3A) (Qian and Costa, 1995). The winged helix domain binds to the consensus DNA sequence A(A/T)TRTT(G/T)RYTY, where R indicates purine and Y pyrimidine (Cereghini, 1996). FoxA proteins are thought to open chromatin through replacement of linker histone using its winged helix domain whose structure is highly similar to that of linker histone (Fig. 3B) (Cirillo *et al.*, 1998; Clark *et al.*, 1993; Ramakrishnan *et al.*, 1993). The presence of the carboxy terminal region (amino acids 295–468 of FoxA3), which is longer than the TAD, is also required for transactivation, probably through binding to core histones H3 and H4 (Cirillo *et al.*, 2002). Direct binding to core histones is consistent with an earlier finding that FoxA proteins bind more stably to nucleosome core particles than to free DNA. The histone binding is not affected by histone acetylation but is facilitated by dimethylation of histone H3 lysine 4 (Cirillo and Zaret, 1999; Lupien *et al.*, 2008).

In addition to the albumin gene, FoxA has been shown to function as a pioneer factor for other genes, such as α fetoprotein and the mouse mammary tumor virus long terminal repeat (MMTV LTR) (Crowe *et al.*, 1999; Holmqvist *et al.*, 2005). However, of the 43 members in the mammalian Fox gene family, it is not known if other members besides FoxA function as pioneer factors (Kato, 2004).

Conclusion

Comparison of VP16, MyoD and other transactivators mentioned above demonstrates that each TAD interacts with an overlapping but different set of proteins during gene activation. Despite extensive research on a variety of TADs, we still do not know why some TADs are more potent than others. Although several assay systems have empirically shown that the TAD of VP16 is one of the most potent, the reason for its extraordinary potency is unclear. To address this fundamental question, it would be necessary to comprehensively compare a diverse range of TADs using a common assay. For such an assay to be complete, it would need to be both plasmid-based (transient transfection with a reporter construct) and chromatin-based (activation of an endogenous gene). A plasmid-based approach is needed to measure the capability of a TAD to recruit basal transcription factors, Mediators and other coactivators, while a chromatin-based approach is needed to reveal additional interactions with histone- and chromatin-modifying enzymes. The chromatin-based approach is especially important if we are to apply the knowledge on TAD functionality to nuclear reprogramming. Once a robust comparison of different TADs is complete with these two approaches, the next step would be the identification and comparison of each TAD's interacting partners. This second step is critical to understand the molecular basis for the differential potency of each TAD. This comprehensive comparison of the functions and

binding proteins of a diverse range of TADs would benefit not only the study of stem cell biology and nuclear reprogramming but also research on transcriptional regulation in general.

Acknowledgments

We thank Michael Franklin of the University of Minnesota for critical reading of the manuscript and the NIH (R01 DK082430), the Leukemia Research Fund and the Academic Health Center of the University of Minnesota for their support of N.K.

Abbreviations used in this paper

5-azaC	5-azacytidine
bHLH	basic helix-loop-helix
DBD	DNA-binding domain
HNF-3	hepatocyte nuclear factor-3
HSV	herpes simplex virus
H/C domain	histidine/cystein rich domain
iPSCs	induced pluripotent stem cells
TAD	transcriptional activation domain
VP	herpes virus protein

References

- BERGSTROM DA, TAPSCOTT SJ. Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family. *Mol Cell Biol.* 2001; 21:2404–2412. [PubMed: 11259589]
- BERKES CA, BERGSTROM DA, PENN BH, SEAVER KJ, KNOEPFLER PS, TAPSCOTT SJ. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol Cell.* 2004; 14:465–477. [PubMed: 15149596]
- BERKES CA, TAPSCOTT SJ. MyoD and the transcriptional control of myogenesis. *Semin Cell Dev Biol.* 2005; 16:585–595. [PubMed: 16099183]
- BLAU J, XIAO H, MCCracken S, O'HARE P, GREENBLATT J, BENTLEY D. Three functional classes of transcriptional activation domain. *Mol Cell Biol.* 1996; 16:2044–2055. [PubMed: 8628270]
- BOSSARD P, ZARET KS. GATA transcription factors as potentiators of gut endoderm differentiation. *Development.* 1998; 125:4909–4917. [PubMed: 9811575]
- BOXUS M, TWIZERE JC, LEGROS S, DEWULF JF, KETTMANN R, WILLEMS L. The HTLV-1 Tax interactome. *Retrovirology.* 2008; 5:76. [PubMed: 18702816]
- BRENNAN TJ, CHAKRABORTY T, OLSON EN. Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc Natl Acad Sci USA.* 1991; 88:5675–5679. [PubMed: 1648228]
- BUSSMANN LH, SCHUBERT A, VU MANH TP, DE ANDRES L, DESBORDES SC, PARRA M, ZIMMERMANN T, RAPINO F, RODRIGUEZ-UBREVA J, BALLESTAR E, GRAF T. A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell.* 2009; 5:554–566. [PubMed: 19896445]
- CARLSSON P, MAHLAPUU M. Forkhead transcription factors: key players in development and metabolism. *Dev Biol.* 2002; 250:1–23. [PubMed: 12297093]

- CARPENTER AE, MEMEDULA S, PLUTZ MJ, BELMONT AS. Common effects of acidic activators on large-scale chromatin structure and transcription. *Mol Cell Biol.* 2005; 25:958–968. [PubMed: 15657424]
- CEREGHINI S. Liver-enriched transcription factors and hepatocyte differentiation. *FASEB J.* 1996; 10:267–282. [PubMed: 8641560]
- CHOI J, COSTA ML, MERMELSTEIN CS, CHAGAS C, HOLTZER S, HOLTZER H. MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc Natl Acad Sci USA.* 1990; 87:7988–7992. [PubMed: 2172969]
- CHOY B, GREEN MR. Eukaryotic activators function during multiple steps of preinitiation complex assembly. *Nature.* 1993; 366:531–536. [PubMed: 8255291]
- CIRILLO LA, LIN FR, CUESTA I, FRIEDMAN D, JARNIK M, ZARET KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell.* 2002; 9:279–289. [PubMed: 11864602]
- CIRILLO LA, MCPHERSON CE, BOSSARD P, STEVENS K, CHERIAN S, SHIM EY, CLARK KL, BURLEY SK, ZARET KS. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* 1998; 17:244–254. [PubMed: 9427758]
- CIRILLO LA, ZARET KS. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell.* 1999; 4:961–969. [PubMed: 10635321]
- CLARK KL, HALAY ED, LAI E, BURLEY SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature.* 1993; 364:412–420. [PubMed: 8332212]
- CONSTANTINIDES PG, JONES PA, GEVERS W. Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. *Nature.* 1977; 267:364–366. [PubMed: 68440]
- CROWE AJ, SANG L, LI KK, LEE KC, SPEAR BT, BARTON MC. Hepatocyte nuclear factor 3 relieves chromatin-mediated repression of the alpha-fetoprotein gene. *J Biol Chem.* 1999; 274:25113–25120. [PubMed: 10455192]
- DAVIS RL, CHENG PF, LASSAR AB, WEINTRAUB H. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell.* 1990; 60:733–746. [PubMed: 2155707]
- DAVIS RL, WEINTRAUB H. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science.* 1992; 256:1027–1030. [PubMed: 1317057]
- DAVIS RL, WEINTRAUB H, LASSAR AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell.* 1987; 51:987–1000. [PubMed: 3690668]
- DE LA SERNA IL, CARLSON KA, IMBALZANO AN. Mammalian SWI/SNF complexes promote MyoD-mediated muscle differentiation. *Nat Genet.* 2001; 27:187–190. [PubMed: 11175787]
- DE LA SERNA IL, OHKAWA Y, BERKES CA, BERGSTROM DA, DACWAG CS, TAPSCOTT SJ, IMBALZANO AN. MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol Cell Biol.* 2005; 25:3997–4009. [PubMed: 15870273]
- DILWORTH FJ, SEAVER KJ, FISHBURN AL, HTET SL, TAPSCOTT SJ. *In vitro* transcription system delineates the distinct roles of the coactivators pCAF and p300 during MyoD/E47-dependent transactivation. *Proc Natl Acad Sci USA.* 2004; 101:11593–11598. [PubMed: 15289617]
- FAERMAN A, PEARSON-WHITE S, EMERSON C, SHANI M. Ectopic expression of MyoD1 in mice causes prenatal lethality. *Dev Dyn.* 1993; 196:165–173. [PubMed: 8400402]
- FLINT J, SHENK T. Viral transactivating proteins. *Annu Rev Genet.* 1997; 31:177–212. [PubMed: 9442894]
- FRIEDMAN JR, KAESTNER KH. The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci.* 2006; 63:2317–2328. [PubMed: 16909212]
- GE H, ROEDER RG. Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell.* 1994; 78:513–523. [PubMed: 8062391]

- GERBER AN, KLESERT TR, BERGSTROM DA, TAPSCOTT SJ. Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev.* 1997; 11:436–450. [PubMed: 9042858]
- GHOSH S, SELBY MJ, PETERLIN BM. Synergism between Tat and VP16 in trans-activation of HIV-1 LTR. *J Mol Biol.* 1993; 234:610–619. [PubMed: 8254663]
- GRAF T, ENVER T. Forcing cells to change lineages. *Nature.* 2009; 462:587–594. [PubMed: 19956253]
- GREAVES R, O'HARE P. Separation of requirements for protein-DNA complex assembly from those for functional activity in the herpes simplex virus regulatory protein Vmw65. *J Virol.* 1989; 63:1641–1650. [PubMed: 2538647]
- GUALDI R, BOSSARD P, ZHENG M, HAMADA Y, COLEMAN JR, ZARET KS. Hepatic specification of the gut endoderm *in vitro*: cell signaling and transcriptional control. *Genes Dev.* 1996; 10:1670–1682. [PubMed: 8682297]
- HOLMQVIST PH, BELIKOV S, ZARET KS, WRANGE O. FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. *Exp Cell Res.* 2005; 304:593–603. [PubMed: 15748903]
- HOPWOOD ND, GURDON JB. Activation of muscle genes without myogenesis by ectopic expression of MyoD in frog embryo cells. *Nature.* 1990; 347:197–200. [PubMed: 1697650]
- HUANG J, WEINTRAUB H, KEDES L. Intramolecular regulation of MyoD activation domain conformation and function. *Mol Cell Biol.* 1998; 18:5478–5484. [PubMed: 9710631]
- IEDA M, FU JD, DELGADO-OLGUIN P, VEDANTHAM V, HAYASHI Y, BRUNEAU BG, SRIVASTAVA D. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell.* 2010; 142:375–386. [PubMed: 20691899]
- IKEDA K, STUEHLER T, MEISTERERNST M. The H1 and H2 regions of the activation domain of herpes simplex virion protein 16 stimulate transcription through distinct molecular mechanisms. *Genes Cells.* 2002; 7:49–58. [PubMed: 11856373]
- INGLES CJ, SHALES M, CRESS WD, TRIEZENBERG SJ, GREENBLATT J. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature.* 1991; 351:588–590. [PubMed: 1646402]
- ITO M, YUAN CX, MALIK S, GU W, FONDELL JD, YAMAMURA S, FU ZY, ZHANG X, QIN J, ROEDER RG. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol Cell.* 1999; 3:361–370. [PubMed: 10198638]
- KAESTNER KH, KNOCHEL W, MARTINEZ DE. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* 2000; 14:142–146. [PubMed: 10702024]
- KANETO H, NAKATANI Y, MIYATSUKA T, MATSUOKA TA, MATSUHISA M, HORI M, YAMASAKI Y. PDX-1/VP16 fusion protein, together with NeuroD or Ngn3, markedly induces insulin gene transcription and ameliorates glucose tolerance. *Diabetes.* 2005; 54:1009–1022. [PubMed: 15793239]
- KATO M. Human FOX gene family. *Int J Oncol.* 2004; 25:1495–1500. [PubMed: 15492844]
- KLEMM RD, GOODRICH JA, ZHOU S, TJIAN R. Molecular cloning and expression of the 32-kDa subunit of human TFIID reveals interactions with VP16 and TFIIB that mediate transcriptional activation. *Proc Natl Acad Sci USA.* 1995; 92:5788–5792. [PubMed: 7597030]
- KOBAYASHI N, BOYER TG, BERK AJ. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol Cell Biol.* 1995; 15:6465–6473. [PubMed: 7565798]
- KRETZSCHMAR M, KAISER K, LOTTSPREICH F, MEISTERERNST M. A novel mediator of class II gene transcription with homology to viral immediate-early transcriptional regulators. *Cell.* 1994; 78:525–534. [PubMed: 8062392]
- KUNDU TK, PALHAN VB, WANG Z, AN W, COLE PA, ROEDER RG. Activator-dependent transcription from chromatin *in vitro* involving targeted histone acetylation by p300. *Mol Cell.* 2000; 6:551–561. [PubMed: 11030335]

- LAIOSA CV, STADTFELD M, XIE H, DE ANDRES-AGUAYO L, GRAF T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity*. 2006; 25:731–744. [PubMed: 17088084]
- LASSAR AB, PATERSON BM, WEINTRAUB H. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell*. 1986; 47:649–656. [PubMed: 2430720]
- LATCHMAN, D. Eukaryotic transcription factors. 5. San Diego: Academic Press; 2008.
- LEE CS, FRIEDMAN JR, FULMER JT, KAESTNER KH. The initiation of liver development is dependent on Foxa transcription factors. *Nature*. 2005; 435:944–947. [PubMed: 15959514]
- LIN YS, HA I, MALDONADO E, REINBERG D, GREEN MR. Binding of general transcription factor TFIIB to an acidic activating region. *Nature*. 1991; 353:569–571. [PubMed: 1922364]
- LUPIEN M, ECKHOUTE J, MEYER CA, WANG Q, ZHANG Y, LI W, CARROLL JS, LIU XS, BROWN M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008; 132:958–970. [PubMed: 18358809]
- MITTLER G, STUHLER T, SANTOLIN L, UHLMANN T, KREMMER E, LOTTSPEICH F, BERTI L, MEISTERERENST M. A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *EMBO J*. 2003; 22:6494–6504. [PubMed: 14657022]
- MOLKENTIN JD, BLACK BL, MARTIN JF, OLSON EN. Mutational analysis of the DNA binding, dimerization, and transcriptional activation domains of MEF2C. *Mol Cell Biol*. 1996; 16:2627–2636. [PubMed: 8649370]
- NEELY KE, HASSAN AH, WALLBERG AE, STEGER DJ, CAIRNS BR, WRIGHT AP, WORKMAN JL. Activation domain-mediated targeting of the SWI/SNF complex to promoters stimulates transcription from nucleosome arrays. *Mol Cell*. 1999; 4:649–655. [PubMed: 10549297]
- PERRY RL, PARKER MH, RUDNICKI MA. Activated MEK1 binds the nuclear MyoD transcriptional complex to repress transactivation. *Mol Cell*. 2001; 8:291–301. [PubMed: 11545732]
- POLESSKAYA A, HAREL-BELLAN A. Acetylation of MyoD by p300 requires more than its histone acetyltransferase domain. *J Biol Chem*. 2001; 276:44502–44503. [PubMed: 11577095]
- POLESSKAYA A, NAGUIBNEVA I, DUQUET A, BENGAL E, ROBIN P, HAREL-BELLAN A. Interaction between acetylated MyoD and the bromodomain of CBP and/or p300. *Mol Cell Biol*. 2001; 21:5312–5320. [PubMed: 11463815]
- POWNALL ME, GUSTAFSSON MK, EMERSON CP JR. Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu Rev Cell Dev Biol*. 2002; 18:747–783. [PubMed: 12142270]
- PTASHNE, M.; GANN, A. *Gene & Signals*. New York: Cold Spring Harbor Laboratory Press; 2002.
- PUMFERY A, DENG L, MADDUKURI A, DE LA FUENTE C, LI H, WADE JD, LAMBERT P, KUMAR A, KASHANCHI F. Chromatin remodeling and modification during HIV-1 Tat-activated transcription. *Curr HIV Res*. 2003; 1:343–362. [PubMed: 15046258]
- QIAN X, COSTA RH. Analysis of hepatocyte nuclear factor-3 beta protein domains required for transcriptional activation and nuclear targeting. *Nucleic Acids Res*. 1995; 23:1184–1191. [PubMed: 7739897]
- RAMAKRISHNAN V, FINCH JT, GRAZIANO V, LEE PL, SWEET RM. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*. 1993; 362:219–223. [PubMed: 8384699]
- ROMANI B, ENGELBRECHT S, GLASHOFF RH. Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *J Gen Virol*. 2010; 91:1–12. [PubMed: 19812265]
- SADOWSKI I, MA J, TRIEZENBERG S, PTASHNE M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature*. 1988; 335:563–564. [PubMed: 3047590]
- SARTORELLI V, HUANG J, HAMAMORI Y, KEDES L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol Cell Biol*. 1997; 17:1010–1026. [PubMed: 9001254]
- SARTORELLI V, PURI PL, HAMAMORI Y, OGRYZKO V, CHUNG G, NAKATANI Y, WANG JY, KEDES L. Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program. *Mol Cell*. 1999; 4:725–734. [PubMed: 10619020]

- SHEN F, TRIEZENBERG SJ, HENSLEY P, PORTER D, KNUTSON JR. Transcriptional activation domain of the herpesvirus protein VP16 becomes conformationally constrained upon interaction with basal transcription factors. *J Biol Chem.* 1996; 271:4827–4837. [PubMed: 8617752]
- STRINGER KF, INGLES CJ, GREENBLATT J. Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature.* 1990; 345:783–786. [PubMed: 2193231]
- TAKAHASHI K, YAMANAKA S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006; 126:663–676. [PubMed: 16904174]
- TAPSCOTT SJ. The circuitry of a master switch: MyoD and the regulation of skeletal muscle gene transcription. *Development.* 2005; 132:2685–2695. [PubMed: 15930108]
- TRAVEN A, JELICIC B, SOPTA M. Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep.* 2006; 7:496–499. [PubMed: 16670683]
- TRIEZENBERG SJ, KINGSBURY RC, MCKNIGHT SL. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* 1988; 2:718–729. [PubMed: 2843425]
- TUMBAR T, SUDLOW G, BELMONT AS. Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol.* 1999; 145:1341–1354. [PubMed: 10385516]
- UESUGI M, NYANGUILE O, LU H, LEVINE AJ, VERDINE GL. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science.* 1997; 277:1310–1313. [PubMed: 9271577]
- UTLEY RT, IKEDA K, GRANT PA, COTE J, STEGER DJ, EBERHARTER A, JOHN S, WORKMAN JL. Transcriptional activators direct histone acetyltransferase complex to nucleosomes. *Nature.* 1998; 394:498–502. [PubMed: 9697775]
- VIERBUCHEN T, OSTERMEIER A, PANG ZP, KOKUBU Y, SUDHOF TC, WERNIG M. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature.* 2010; 463:1035–1041. [PubMed: 20107439]
- VIGNALI M, STEGER DJ, NEELY KE, WORKMAN JL. Distribution of acetylated histones resulting from Gal4-VP16 recruitment of SAGA and NuA4 complexes. *EMBO J.* 2000; 19:2629–2640. [PubMed: 10835360]
- WALKER S, GREAVES R, O'HARE P. Transcriptional activation by the acidic domain of Vmw65 requires the integrity of the domain and involves additional determinants distinct from those necessary for TFIIB binding. *Mol Cell Biol.* 1993; 13:5233–5244. [PubMed: 8395001]
- WEIGEL D, JURGENS G, KUTTNER F, SEIFERT E, JACKLE H. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell.* 1989; 57:645–658. [PubMed: 2566386]
- WEINTRAUB H, DWARKI VJ, VERMA I, DAVIS R, HOLLENBERG S, SNIDER L, LASSAR A, TAPSCOTT SJ. Muscle-specific transcriptional activation by MyoD. *Genes Dev.* 1991; 5:1377–1386. [PubMed: 1651276]
- WEINTRAUB H, TAPSCOTT SJ, DAVIS RL, THAYER MJ, ADAM MA, LASSAR AB, MILLER AD. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad Sci USA.* 1989; 86:5434–5438. [PubMed: 2748593]
- WYSOCKA J, HERR W. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends Biochem Sci.* 2003; 28:294–304. [PubMed: 12826401]
- XIAO H, PEARSON A, COULOMBE B, TRUANT R, ZHANG S, REGIER JL, TRIEZENBERG SJ, REINBERG D, FLORES O, INGLES CJ, et al. Binding of basal transcription factor TFIID to the acidic activation domains of VP16 and p53. *Mol Cell Biol.* 1994; 14:7013–7024. [PubMed: 7935417]
- ZARET KS. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nat Rev Genet.* 2008; 9:329–340. [PubMed: 18398419]
- ZHOU Q, BROWN J, KANAREK A, RAJAGOPAL J, MELTON DA. *In vivo* reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature.* 2008; 455:627–632. [PubMed: 18754011]
- ZHOU Q, MELTON DA. Extreme makeover: converting one cell into another. *Cell Stem Cell.* 2008; 3:382–388. [PubMed: 18940730]

ZHU H, JOLIOT V, PRYWES R. Role of transcription factor TFIIF in serum response factor-activated transcription. *J Biol Chem.* 1994; 269:3489–3497. [PubMed: 8106390]

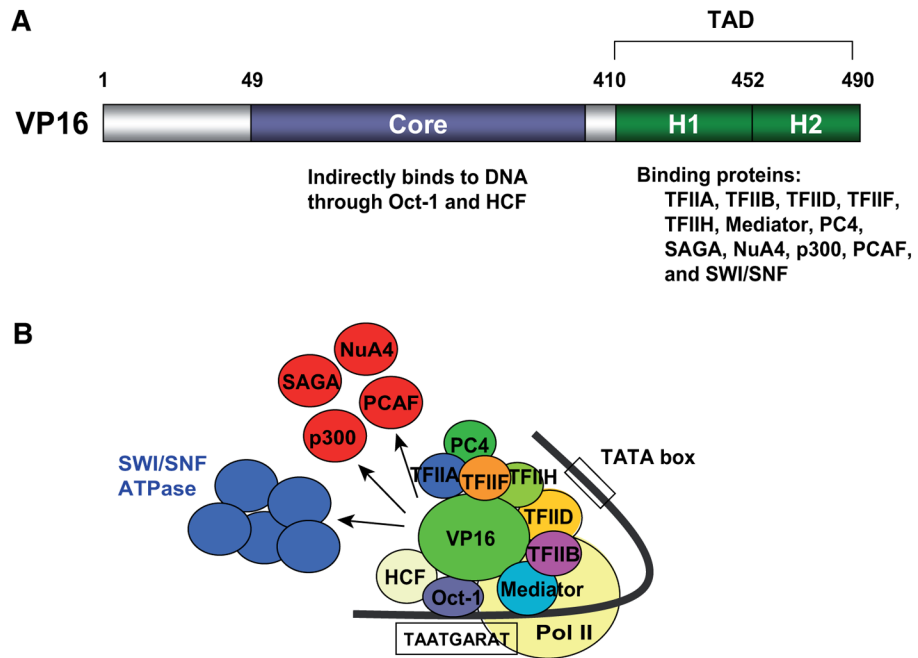


Fig. 1. Domain structure and interacting proteins of herpes virus protein VP16

(A) Domain structure of VP16. The amino acid number for the TAD sequence varies slightly depending on the literature source. The numbers used here were taken from Walker *et al.* (Walker *et al.*, 1993). (B) Proteins that interact with VP16 during gene activation.

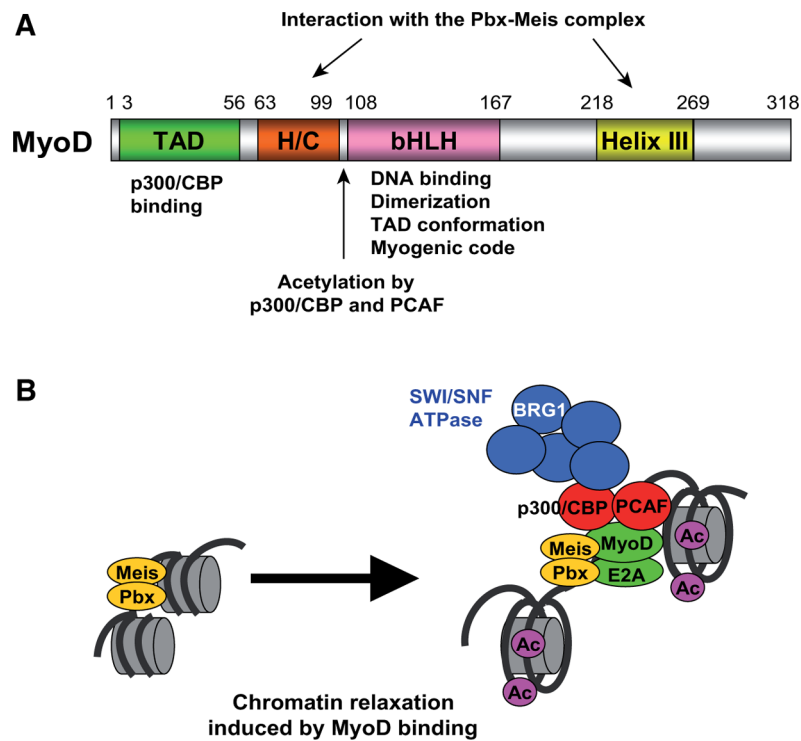


Fig. 2. Domain structure and functions of MyoD

(A) Structure and functions of MyoD domains. Amino acid numbers are shown above the protein domains. Amino acid numbers are taken from Bergstrom and Tapscott (Bergstrom and Tapscott, 2001; Gerber *et al.*, 1997). (B) Recruitment of binding proteins and resulting chromatin relaxation by MyoD during gene activation.

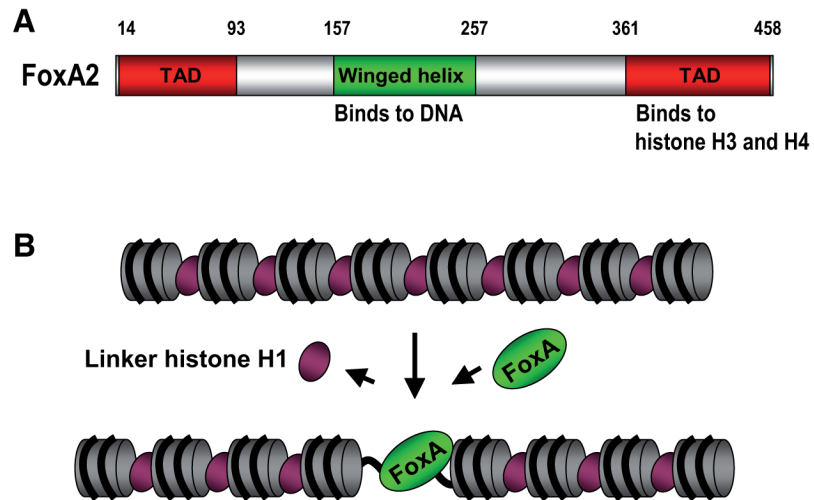


Fig. 3. Domain structure of FoxA and its function as a pioneer factor

(A) Domain structure of FoxA2. FoxA1 and A3 share similar structures. The amino acid number for the carboxyl terminus sequence varies depending on the literature source. The numbers used here were taken from Qian and Costa. (Qian and Costa, 1995). (B) Replacement of the linker histone H1 by FoxA and subsequent relaxation of chromatin during gene activation.