4 Discussion

Prior to this study, experimental data obtained by our group suggested that DPF3 plays a role in cardiac development and disease. First, an upregulation of *DPF3* transcripts was observed in patients with Tetralogy of Fallot (TOF), a congenital heart defect (Kaynak *et al.*, 2003). Moreover, mRNA expression analyses during embryonic development of mouse and chicken revealed that *Dpf3* transcripts are expressed – among other tissues - early in the developing heart. Additionally, *Dpf3* expression was also observed in developing somites, which partially give rise to skeletal muscle, as well as in neuronal cells (Kaynak, 2005).

These findings led to the hypothesis that Dpf3 may be important for heart development. This hypothesis was tested by analysis of Dpf3 function in a combined approach using biochemical and *in vivo* studies to define the molecular properties of Dpf3 and to define its role in developmental pathways of heart development.

4.1 Expression profiles of *Dpf3* splice variants

The onset of *Dpf3* expression had been traced back by *in situ* hybridization to the first stages of cardiomyocyte differentiation at E7.5, which suggested that *Dpf3* is involved in the differentiation process. To define expression profiles of the splice variants *Dpf3a* and *Dpf3b* throughout all stages of mouse cardiac development and adulthood, quantitative real-time PCR (qPCR) of isolated embryonic and adult hearts was performed. The qPCR analysis showed that *Dpf3a* was expressed at high levels early in the developing heart at E9.5 (Figure 3.1). Expression levels of *Dpf3b* were lower compared to *DPF3a* at this early stage, supporting the earlier data from *in situ* hybridization analyses. During chamber morphogenesis and outflow tract development (E11.5-14.5), the expression levels of both splice variants were markedly reduced. Nevertheless, in the maturing embryonic and adult heart, both splice variants were expressed at comparable levels, suggesting that Dpf3 plays a role during embryonic heart development as well as in the adult heart.

An embryonic expression pattern that persists in adulthood has been described for other important cardiac transcription factors such as Mef2a and Nkx2.5, that are important for cardiac development as well as homeostasis (Toko *et al.*, 2002; Wang *et al.*, 2003).

4.2 The role of dpf3 in zebrafish embryonic development

Sequence comparison of the newly cloned full-length *dpf3* transcript to higher vertebrates revealed a high level of sequence conservation in all annotated domains (Chapter 3.2.1.). Accordingly, the comparison of gene expression data from mouse and zebrafish embryos showed that *Dpf3* is expressed in an evolutionary conserved pattern. Mouse and zebrafish, which are separated by around 380 million years of evolution (Kumar and Hedges, 1998), showed expression of *Dpf3* in the developing heart, somites as well as in the brain and neural tube. Interestingly, zebrafish *dpf3* was predominantly expressed in the developing ventricle of the heart, whereas the mouse and chicken orthologs were expressed more broadly throughout the heart. This suggests that *dpf3* is especially important for function of the embryonic ventricle in zebrafish. On the other hand, the fish heart is substantially less complex than that of higher vertebrates and the zebrafish atrium might not require the presence of dpf3 for proper function.

The cardiac expression of *dpf3* enabled the use of the zebrafish embryos for the functional analysis in heart development, which was carried out using morpholino-modified antisense oligonucleotides against *dpf3* (MO^{dpf3}). Injection of the MO^{dpf3} at the one-cell stage and subsequent analysis of the modified *dpf3* transcripts by RT-PCR and sequencing showed the efficacy of the morpholino. As the morpholino targeted the exon4-intron4 boundary, the 5´ located 2/3-domain was still transcribed, whereas the C2H2-type zinc-finger and the double PHD finger were not present in morphant transcripts. Nevertheless, due to the presence of premature termination codons (PTCs), both of the generated morphant transcripts were probably subject to nonsense-mediated mRNA decay (NMD) (Conti and Izaurralde, 2005).

The specificity of the MO^{dpf3} was demonstrated in rescue-experiments by co-injection of mature *dpf3* mRNA, which led to normal development of embryos in a significant number of cases. Moreover, control embryos were injected with a control-morpholino, which had no effect on development. This demonstrated that the observed phenotypes were specifically due to knockdown of *dpf3*.

Knockdown in zebrafish embryos revealed an essential role for *dpf3* in heart and skeletal muscle development as morphant animals frequently displayed morphogenetic defects in cardiac looping and contractility. Moreover, skeletal muscle defects were observed. A detailed description of the phenotype can be founding the accompanied manuscript.

With regard to the heart defects, the ventricle was specifically affected. The strength of the ventricular contraction was weakened compared to normal embryos, while the atrium was unaffected. The fact that only the ventricle was affected is in good agreement with the finding

that *dpf3* is expressed only in this portion of the heart, suggesting that the ventricular dysfunction is a direct effect. In addition to defects in cardiac contractility, structural defects of myofibrils were frequently observed in the heart and skeletal muscle of *dpf3* morphants. Within the embryonic ventricle, very few myofibrils were observed and within somites, the usual parallel alignment of myofibrils was disturbed. Also, transversion of somite boundaries by the thin actin filaments was frequently observed in regions of boundary disruption.

The expression analysis of *dpf3* morphants by microarray and realtime-PCR showed that the myofibrillar defects were due to deregulation of proteins essential for sarcomere function. In particular, the z-disc of sarcomeres appeared to be affected. Z-discs represent the lateral boundaries of the sarcomere where the thin filaments, titin, and nebulin are anchored (Clark *et al.*, 2002). The thin filaments are capped at the ends by Tropomodulin (Tmod) and CapZ, with the latter anchoring them in the Z-disc. In the analysis of *dpf3* morphants, *capZ alpha-1* (*zgc:101755*) and *tropomodulin 4* were downregulated (0.5 fold; 0.6 fold, respectively). It has been shown that inhibition of CapZ-capping activity delays the organization of the Z-discs and thin filaments (Schafer *et al.*, 1995), supporting the idea that CapZ organizes and aligns the barbed ends of the thin filaments. In addition, Tropomodulin caps the pointed ends of the thin filaments and has a high affinity to Nebulin (McElhinny *et al.*, 2001). Interestingly, downregulation of Tropomodulin in cardiomyocytes has been shown to disrupt sarcomeric structures of myofibrils, characterized by unusually long actin filament bundles (Sussman *et al.*, 1998). Consequently, the failure of thin filament attachment at the z-discs of somites is most likely due to downregulation of the actin capping proteins.

Filamin C gamma b (*flncb*) was found to be upregulated in morphants (2.5 fold). Filamin C is exclusively expressed in striated muscle and localized in myofibrillar z-discs, in myotendinous junctions and in intercalated discs of striated muscle, where it interacts with several other muscle proteins, such as delta-and gamma sarcoglycan, myotilin, beta-1D-intergrin and the FATZ family of z-line proteins (Clark *et al.*, 2002) . The crucial role for Filamin C in muscle development is reflected by the severe defects in myogenesis and myotube structure of Filamin C knock-out mice, accompanied by a loss of distinct z-lines (Dalkilic *et al.*, 2006).

At myotendinous junctions and in intercalated discs, Filamin C interacts with Cmya1 (van der Ven *et al.*, 2006), another differentially expressed gene in *dpf3* morphants (2.9 fold upregulation). Structurally, Cmya1 is characterized by the presence of Xin-repeats, which consist of 16 repetitive amino-acids and a putative DNA-binding domain (Jung-Ching Lin *et al.*, 2005). Xin, the ortholog of cmya1, has been shown to be required for cardiac morphogenesis in chicken embryos, as antisense oligonucleotide injection leads to defects in

heart looping and reduced cardiac contractility. Moreover, Xin is a downstream target of Mef2c (Paris *et al.*, 2004; Wang *et al.*, 1999). Xin-alpha null mouse hearts reveal intercalated disc disruption and myofilament disarray (Gustafson-Wagner *et al.*, 2007). Finally, human CMYA1 maps to chromosomal region 3p21.2-21.3, near a dilated cardiomyopathy with conduction defect-2 locus. Although the effect of Xin upregulation on muscle differentiation is not clearly defined yet, one might speculate that its upregulation can have deterious effects on sarcomere function, as Xin repeats compete with Tropomyosin for binding of actin filaments in a concentration-dependent manner *in vitro* (Pacholsky *et al.*, 2004). Thus, excess Xin might interfere with Tropomyosin function, which is essential for muscle contractility (Solaro and Rarick, 1998).

With regard to the defects in cardiac contractility, another protein of importance is Troponin I, the inhibitory subunit of the Troponin complex. In the absence of calcium, Troponin I inhibits actin-myosin-interactions, providing a calcium-dependent regulator of muscle contraction (Chang and Potter, 2005). *Troponin I* was strongly upregulated in *dpf3* morphants (4.7 fold), which frequently displayed impaired cardiac contractility. Notably, mice expressing mutated versions of Troponin I display hypercontractility (James *et al.*, 2000), mirror imaging the *dpf3* morphant phenotype.

As cardiac contractility was mostly impaired in the ventricle, it is reasonable to assume that, in addition to the structural defects of the sarcomeres, the conduction system is affected by knockdown of *dpf3*. In this context, it is striking that members of the *Iroquois* family of transcription factors, namely *irx1b* and irx4a were deregulated in *dpf3* morphants. *Iroquois* genes play a role in regulating neuronal gene expression in the brain and the conduction system of the heart. In *dpf3* morphants, *irx1b* was upregulated 1.7 fold. In the mouse embryo, Irx1 labels the cells of the ventricular conduction system and reduction of *irx1b* levels in zebrafish impairs heart rate (Elaine, 2004). Moreover, *irx1b* expression is disrupted in the silent heart mutant (Sehnert *et al.*, 2002). It remains to be determined whether *irx1b* upregulation directly affects cardiac contractility or whether it reflects an adaptive effect of the diseased organ. Overexpression by *irx1b* mRNA injection could be used to address this question. *Irx4a* is also expressed in the embryonic heart of zebrafish embryos (Virginie Lecaudey, 2005) and was downregulated in *dpf3* morphants (0.6 fold). A specific function for irx4a in zebrafish has not been assigned yet. In the mouse, Irx4 is expressed specifically in the ventricles and has been implicated in ventricular differentiation (Bruneau *et al.*, 2000). Therefore, irx4a is a candidate molecule for ventricular differentiation to be analyzed in *dpf3* morphants.

The basic helix-loop-helix transcription factor *hand2* was downregulated 0.6 fold in *dpf3* morphant hearts. Its essential role in heart development is reflected by the severe cardiac defects and embryonic lethality of Hand2 null mouse and zebrafish embryos (Srivastava *et al.*, 1997; Yelon *et al.*, 2000). In zebrafish, loss of hand2 leads to decreased cardiomyocyte differentiation early in the anterior lateral plate mesoderm (ALPM), whereas in the mouse, Hand2 appears to play a later role in formation of the right ventricle (Trinh *et al.*, 2005). The situation in mouse is furhter complicated by the cardiac expression of Hand1, which may rescue the loss of Hand2 (McFadden *et al.*, 2005). The phenotypes observed in *dpf3* morphants and the zebrafish *hands off* mutant (which lacks *hand2*) are rather distinct, with the latter representing an earlier defect in cardiogenesis. Therefore, it remains to be determined whether *hand2* downregulation is a direct or a secondary effect of the myocardial defects observed in *dpf3* morphants. Interestingly though, the *hands off* phenotype is, in addition to defects in cardiomyocyte differentiation, marked by failure of proper extracelluar matrix deposition. Initial gene expression analyses of the *dpf3* morphant embryos revealed the deregulation of extracelluar matrix molecules implicated in cardiac cell migration. Among these were *matrix metalloproteinase 9* (*mmp9*) and *fibronectin1b*. The family of extracelluar matrix metalloproteinases are zinc-dependent endopeptidases that play an important role in cell migration and tissue remodeling in development and disease. Indeed, induction of Mmp2 and Mmp9 has been shown to be indicative for chronic heart failure due to activation of endocardial-endothelial cell apoptosis (Ovechkin *et al.*, 2005). Moreover, Mmp9 is involved in the degradation and reorganization of collagens, elastin and proteoglycans (Sternlicht and Werb, 2001). With regard to the heart, Mmp9 (and Mmp13) are strongly expressed in the endocardial cushions of the developing valves and human heart valve disease is characterized by increased deposition of extracelluar matrix components (Hinton *et al.*, 2006). *Fibronectin 1b* (*fnb1*) was upregulated 3.0 fold in *dpf3* morphants. Fibronectins play an important role in mediating cell migration. Mouse embryos lacking the Fibronectin gene die around E8.5 due to defects of the neural tube, vasculature, somites and heart. The latter defect is due to impaired cell migration (George *et al.*, 1997). In zebrafish, the *nat* (*natter*) locus, which causes cardia bifida, is encoded by *fibronectin* (Trinh and Stainier, 2004). Loss of fibronectin therefore leads to an early, embryonically lethal phenotype. The effect of fibronectin upregulation has not been analyzed yet. Thus it remains to be determined if fibronectin deregulation is involved in causing the *dpf3* morphant phenotype.

With regard to the functional categories of genes deregulated in *dpf3* morphants, gene ontology analyses showed that upregulated genes were often involved in metabolic processes

as well as transcriptional regulation. In line with that, the cellular component of upregulated genes was mostly found to be in the nucleus. The upregulation of genes controling gene expression may reflect a compensatory feedback loop due to loss of *dpf3*. On the side of downregulated genes, genes involved in iron metabolism were overrepresented and the most common cellular component was the cytosol. It has to be kept in mind that whole embryos were used for the genome wide array analysis. Therefore, as only a subset of differentially expressed genes is of significance for heart or skeletal muscle development, the corresponding GO annotations may not always reflect the biological relevance.

Taken together, dpf3 plays an important role in regulating muscle cell differentiation and the sarcomeric defects in heart and skeletal muscle observed in *dpf3* morphants can be accounted to deregulation of sarcomeric proteins. The defects in contractility are most likely a consequence of the structural abnormalities of the sarcomeres.

4.3 DPF3 is a novel subunit of the BAF chromatin remodeling complex

DPF3a and DPF3b were identified as a novel subunit of the BAF chromatin remodeling complex through a combination of tandem affinity purification and mass spectrometry (Chapter 3.3). Tandem affinity purification involves the introduction and expression of a dually-tagged fusion protein into the host cell followed by two rounds of purifications under native conditions (Gingras *et al.*, 2005). This allows for thorough purification of interaction partners without disrupting the bound complexes. Application of mass spectrometry then allows for identification of novel protein-interaction partners. The threshold set for the identification of interaction partners in the mass spectrometry analysis was a Mascot Score of > 50 and a minimum number of two matching MS/MS spectra. The high values obtained for both criteria in almost all of the identified BAF complex subunits reflect the strength and specificity of the interaction. Moreover, the fact that the entire protein complex was purified shows that the conditions used during the process were well adjusted. Aside from components of the BAF complex, only few other proteins were identified, supporting the specificity of the association.

A confirmation of the results was carried out by protein pulldown experiments using recombinant GST-DPF3 fusion proteins and HEK293T nuclear protein extracts. As both isoforms of DPF3 were able to pull down the BAF complex – shown by immunodetection of the core component BRG1 as well as BAF60c - the interaction domain of DPF3 must reside within the shared N-terminal region containing the 2/3-domain, nuclear receptor interaction

domain (NID) and C2H2-type zinc finger. In this context, it is interesting to note that the C2H2-type zinc finger can serve as a DNA-binding- as well as protein-protein-interaction domain (Iuchi, 2001). Furthermore, it will be interesting to analyze the potential role of the putative NID. The role of ATP-dependent chromatin remodeling complexes in transcriptional regulation by steroid receptors is well established (Belandia and Parker, 2003). For example, the estrogen receptor interacts with a component found in all SWI/SNF remodeling complexes, namely Baf57, which is essential for the ability of the p160 proteins to activate transcription of estrogen-responsive reporter genes (Belandia *et al.*, 2002). Moreover, it has been shown that components of the BAF complex, such as BAF60c, interact with nuclear hormone receptors and regulate transcriptional activity of retinoic acid-related orphan receptor a1 (Debril *et al.*, 2004). In addition, it is well established that nuclear receptor signaling, for example by retinoic acid, is essential for heart and skeletal muscle development (Hamade *et al.*, 2006; Sucov *et al.*, 1994). Thus, candidate factors to test in interaction studies would include nuclear receptors with overlapping expression patterns or a defined role in muscle development such as RXR*a*, COUP-TF2 and PPAR*gamma* (Barak *et al.*, 1999; Kastner *et al.*, 1997; Pereira *et al.*, 1999). Interestingly, COUP-TF2 has been shown to be a transcriptional regulator of the *DPF3* promoter (see Chapter 3.7.2), further indicating a role of nuclear hormone receptors in mediating DPF3 signaling.

The biological relevance of the identification of DPF3 as a novel component of the BAF complex is reflected by the fact that BAF complexes are essential for heart and skeletal muscle development. The core subunits of the BAF complex are joined by tissue-specific subunits to increase the specificity of the complex in targeting tissue-specific loci (Wang *et al.*, 1996). A direct interaction of GST-DPF3 and components of the BAF complex was identified in pulldown assays with *in vitro* translated Baf60c. The subunits Baf60b and Baf60c are expressed specifically in pancreas and muscle, respectively (Wang *et al.*, 1996), thus representing tissue-specific parts of the complex. Both DPF3a and DPF3b were able to interact with Baf60c, supporting the mass spectrometry results obtained with HEK293T nuclear extracts. It remains to be determined which domains mediate the interactions. The N-terminal part of Baf60c has been shown to bind to the C-terminal part of PPARgamma, whereas the C-terminal part of Baf60c interacts with the N-terminal part of PPARgamma (Debril *et al.*, 2004). In addition to the abovementioned nuclear receptor interactions, Baf60c has been shown to be a promiscuous interactor with cardiac transcription factors Nkx2.5, Gata4 and Tbx5 (Lickert *et al.*, 2004). Pulldown assays using truncated versions of DPF3 and Baf60c would be suitable to address this question in the future.

Another example for tissue-specificity has been shown for the subunit Baf53, which is specifically expressed in proliferating neuronal precursor cells in the developing mouse brain and keeps these cells from differentiating. Once Baf45b replaces this subunit, cells stop proliferating and start differentiating (Lessard *et al.*, 2007a). It is thus tempting to speculate that Dpf3 serves as a tissue-specific BAF subunit that regulates the transition of muscle precursors to differentiating myocytes. Therefore, it would be interesting to monitor the expression profile and subcellular localization of Dpf3 mRNA and protein during muscle differentiation.

The identification of DPF3 as a novel component of the BAF chromatin remodeling complex adds to the increasing appreciation of tissue-specific epigenetic mechanisms. Additional experiments are needed to show to which degree DPF3 is essential for nucleosome remodeling at these sites.

4.4 The Plant-Homeodomains of DPF3b recognize modified histone lysine residues

Plant-homeodomains are frequently found in nuclear proteins and are defined by a stretch of around 60 amino acids containing conserved cysteine and histidine residues (C4-H-C3) that coordinate two zinc ions forming interweaved zinc fingers bridged by two small beta strands (Ruthenburg *et al.*, 2007a). They are known to serve as protein-protein-interaction domains and to bind nuclear phosphoinositides as well as nucleosomes (Bienz, 2006). With regard to nucleosome binding, the specific recognition of methylation marks has been reported (Ruthenburg *et al.*, 2007a). However, in a proteome-wide screen in yeast, only 8 out of 18 PHD fingers showed specific histone methyllysine interactions, indicating additional roles for the plant-homeodomain (Shi *et al.*, 2007).

This study revealed that the double PHD finger of DPF3b (DPF3b-PHD1-PHD2) interacted specifically with mono- and di-methylated lysine 4 on histone 3 (H3K4me1, H3K4me2), whereas unmodified histone 3 (H3, aa1-21) and trimethylated lysine 4 (H3K4me3) were not detected. Histone lysine recognition has been described previously for the PHD finger. In some cases, the trimethylated form of lysine 4 was recognized with higher affinity than the di-methylated form (Li *et al.*, 2006a; Martin *et al.*, 2006; Pena *et al.*, 2006), while the PHD finger of BHC80 specifically recognizes unmethylated lysine 4. Thus, the observed mono- and di-methyllysine specific recognition represents a novel binding specificity of the PHD finger.

In addition to the methyllysines, GST-DPF3b-PHD1-PHD2 also recognized acetylated lysines on histone 3 (H3K14ac, H3K9ac) and histone 4 (H4K5ac, H4K8ac, H4K12ac, H4K16ac), a

property previously only known for the bromodomain (Mujtaba *et al.*, 2007). The fact that phosphorylated serine 1 and methylated arginine 3 on histone 4 were not detected demonstrated the specificity of the binding reaction. The GST-DPF3a-½-PHD1 construct of DFP3a did not bind to any histones, showing that a full PHD finger is necessary for binding.

The binding properties of the plant-homeodomain 1 and 2 were tested alone to see whether single PHD fingers are able to recognize histone modifications. Pulldown assays showed that single DPF3 PHD fingers were sufficient for the interaction with lysine acetylations on histone 4 (H4K5ac, H4K8ac, H4K12ac, H4K16ac), whereas histone 3 acetylations and methylations were only recognized by the double PHD finger (Figure 3.7B).

The structural analysis of PHD finger-methllysine interactions has shown that the formation of a cage of aromatic residues is the basis for binding specificity and that the peptide becomes stabilized through interactions with acidic residues surrounding the cage across the length of the peptide (Ruthenburg *et al.*, 2007a). The essential role of caging tryptophan residues in mediating these interactions has been demonstrated in the case of BPTF, ING2, RAG2-and Yng1-PHD- fingers, where the H3K4 peptide docks across the upper edge of the sheet projecting the R2 and K4 side chains into two pockets that are divided by a conserved tryptophan residue (Li *et al.*, 2006a; Martin *et al.*, 2006; Matthews *et al.*, 2007; Pena *et al.*, 2006). As for methyllysine recognition, aromatic residues are also part of the structural basis for acetyllysine recognition.

To get to the structural basis of the DPF3-PHD-histone interactions, site-directed mutagenesis was performed on a conserved tryptophan residues within PHD1 (W311E) and a tandem repeat of cysteines within PHD2 (C360R/C363R), respectively. The nonpolar, hydrophobic tryptophan residues was replaced by a polar acidic glutamic acid while the neutral cysteines were replaced by basic arginines. These mutant proteins were then tested for their histone peptide binding properties (Figure 3.7B).

GST-pulldown assays using GST-DPF3b-PHD1-PHD2 (C360R/C363R) showed that histone 4 modifications were still detected, while histone 3 modifications were not recognized anymore. This result was in good agreement with results comparing the tandem PHD finger binding properties to the single PHD fingers, which showed that both domains are necessary for the histone 3 interaction. The same result was obtained for GST-DPF3b-PHD1-PHD2 (W311E) while both mutations combined in the same construct GST-DPF3b-PHD1-PHD2 (W311E/C360R/C363R) completely abolished histone recognition.

To analyze the effect of amino acid exchanges on single DPF3-PHD fingers, the same mutations were introduced into the GST-DPF3b-PHD1 and GST-DPF3b-PHD2 constructs.

Pulldown assays revealed that histone 4 binding was completely abolished in the GST-DPF3b-PHD1 (W311E) and GST-DPF3b-PHD2 (C360R/C363R) mutants, showing that intact single PHD fingers are necessary and suffiencent for histone 4 acetyllysine recognition.

These results led to the following conclusions. First, an intact double PHD finger is sufficient and necessary for histone 3 methyl- and acetyl-interactions. The mutual exclusive binding to histone 3 suggests that the binding pocket for histone 3 peptides is composed of residues from PHD1 and PHD2, which contribute to either formation of the aromatic cage or stabilization of the peptide through hydrogen bonds along the peptide. In this context it is interesting to note that the crystal structure of the double PHD finger of Mll3 has been resolved recently. In this structure, PHD1 and PHD2 are separated by a narrow cleft, which may represent a binding pocket for histone 3, to which residues from both domains contribute.

Second, single PHD fingers are necessary and sufficient for histone 4 recognition, suggesting a different mode of recognition for H3 and H4, respectively. For example, distinct binding pockets might exist for each histone tail. The necessity of an intact PHD finger for histone recognition is demonstrated by failure of the ½-PHD1 of DPF3a to recognize H4. Histone methyllysine binding properties similar to DPF3 have been described for a mutated form of BPTF-PHD (Y17E), which preferentially recognizes H3K4me2 over H3K4me1. Selective binding of lower lysine methylation states can be achieved through size selection, as the bulky trimethylammonium can be excluded from a narrow binding pocket. Although structurally unrelated, both domains achieve methyllysine binding through formation of a cage consisting of aromatic residues in which bulky amino acids selectively exclude the trimethylammonium, whereas di- or mono-methylated side chains are able to enter the pocket (Li *et al.*, 2007; Min *et al.*, 2007). Experiments using crystallography or NMR spectroscopy will determine the structural basis for the histone tail recognition by DPF3b.

The high level of sequence similarity between DPF3 and its paralogs DPF1 and DPF2 suggests that comparable histone binding properties exist. GST-pulldown assays using DPF1 and DPF2 fusion proteins would be suitable to address this question in the future.

With regard to DPF3-histone interactions, it may be interesting to note that a stretch of amino acids only found in the C-terminus of DPF3a (AA345-348, SGRG) is similar to the N-terminal tail of histone 4 (AA1-4, SGRG). The serine of histone 4 is subjected to post-translational modification by Casein Kinase II, which phosphorylates this residue (Cheung *et al.*, 2005). CKII also phosphorylates non-histone proteins. Thus, a combined regulation by phosphorylation of histone 4 and DPF3a might exist. Experimental evidence for post-

translational modification comes from Western Blot analyses of DPF3a and DPF3b. Despite the lower calculated molecular weight of DPF3a, it has been shown to migrate slower in a polyacrylamide gel, indicative of a post-translational modification.

4.5 ChIP-chip analysis of Dpf3 downstream targets

Chromatin immunoprecipitation is a technique to identify regions of chromatin bound by a particular protein *in vivo* and has been applied to identify genes directly regulated by Dpf3 in mouse C2C12 skeletal muscle cells.

A custom muscle specific promoter array with 740.000 probes covering 10 kb upstream and 3 kb downstream of ~12.000 transcripts was designed and used (see accompanied manuscript). The transcripts consisted of a comprehensive set of muscle-expressed genes (Fischer *et al.*, 2008). This array enabled the analysis of genes of interest with a much higher degree of tiling and sequence coverage than standard whole-genome arrays would provide. Due to lack of isoform-specific antibodies suitable for immunoprecipitation, experiments were carried out by transfecting cells with Flag-epitope tagged expression vectors for Dpf3a and DPF3b and immunoprecipitation was performed using an anti-Flag antibody. As negative control, immunoprecipitation was performed using untransfected cells and the anti-Flag antibody to exclude unspecific target amplification.

A total of 1,201 DPF3 targets were identified, 460 in the case of DPF3a and 979 for DPF3b. The number of shared targets was 238. Despite the fact that DPF3a does not directly interact with histone tails (see Chapter 3.4), the ChIP-chip analysis was successful in identifying downstream targets. This can be explained because DPF3a is associated with the BAF complex, which in turn binds to chromatin through the bromodomain of BRG1 or BRM as well as through the PHD fingers of DPF3b (see Chapter 3.4). Therefore, it is likely that loci identified for DPF3a are indirectly bound through other components of the BAF complex. An indirect binding mechanism is also supported by the lower number of identified targets, as crosslinking of more than two proteins via the complex is not as stable as the direct protein–histone interaction during affinity purification. On the other hand, DPF3b directly interacts with chromatin through histone tails, providing a thorough interaction surface that is more stable during the tandem affinity purification process. Therefore, the number of targets is almost twice the size of that for DPF3a. These differences may also reflect isoform-specific functions, which are brought about by specific interaction partners. This may be of significance, bearing in mind the slightly earlier onset of *Dpf3a* expression in the mouse heart.

Finally, a direct interaction between the co-expressed isoforms DPF3a and DPF3b may account for DPF3a targets identified by ChIP.

It would be interesting to compare data obtained from ChIP-analyses of other components of the BAF complex with those of DPF3 to see the degree of overlap and divergence, which might point to additional functional roles of DPF3.

In line with the previously discussed role of DPF3 in chromatin remodeling, the gene ontology analysis of Dpf3 downstream targets revealed an overrepresentation of terms such as nucleosome and chromatin assembly for both isoforms. In addition, Dpf3b GO terms included developmental processes such as actin cytoskeletal organization and heart development. Among the ChIP-chip targets were many genes essential for heart and skeletal muscle development, such as the signaling molecules *Myogenic factor 6* (*Myf6*) and *Bone morphogenetic protein 2* (*Bmp2*) (Pownall *et al.*, 2002; Prall *et al.*, 2007). Transcription factors with a clear role in heart development bound by DPF3 included *Cited2*, *Foxc2*, *Mef2a*, *Mef2d*, *Pitx2*, *Tbx2*, *Tbx3* and *Shox2 (Blaschke et al., 2007; Campione et al., 1999; Hinits and Hughes, 2007; Hoogaars et al., 2007; Kume et al., 2001; Weninger et al., 2005)*. Finally, a set of muscle structural genes were also identified as direct targets of DPF3. These included *alpha-Actinin*, *Myosin heavy chain 9* (*Myh9*), *Myosin light chain 1* (*Myl1*), *Myosin 10* (*Myo10*) and *Troponin C* (*Tnnc1*) (Clark *et al.*, 2002).

These results show that DPF3 is regulating the expression of muscle genes in two ways. First, indirectly by regulating the activity of signaling pathways controlling myogenesis and second, by directly affecting the expression of muscle structural proteins. This is in good agreement with the results obtained from the zebrafish analysis, which revealed an essential role of *dpf3* in muscle development.

The BAF chromatin complex has been shown to regulate myogenesis by integrating signals from the MKK6/p38 and IGF1/PI3K/AKT pathway (Serra *et al.*, 2007). Nevertheless, direct downstream targets had not been identified previously. Therefore, this study expands the knowledge about how myogenesis is controlled by the BAF complex.

4.6 Dpf3 binding sites co-localize with acetylated and methylated histone tails

DPF3a and DPF3b are novel components of the BAF chromatin remodeling complex (see Chapter 3.3). Moreover, the double plant-homeodomain of DPF3b has been shown to interact with acetylated and methylated histone lysine residues (see Chapter 3.4). Therefore, it

is reasonable to assume that DPF3b serves as a reader of these modifications and guides the BAF complex to the modified loci. To analyze the degree of overlap between Dpf3 targets and modified chromatin regions on a genome-wide level, the datasets of the Dpf3 ChIP-chip experiments were combined with data obtained in previous analyses of the modification status of histones 3 and 4 in the same cell line (see accompanied manuscript).

Out of 546 Dpf3b binding sites, 265 overlapped with histone 3 acetylation, 220 with histone 4 acetylation and 294 overlapped with histone 3 methylation marks. Thus 66 % of Dpf3b binding sites overlap with acetylation marks and 54 % with methylation marks, which is significantly more than one could expect from random permutations (min 26 %, max 39 %). Therefore, the linker function of DPF3 between the BAF complex and modified histone residues is likely to have a general implication for gene transcription in muscle cells.

The *Pitx2* locus represents an example of co-occurrence of Dpf3 binding as well as histone modification. The high impact of the modification status of histones (acetylation vs. deacetylation and methylation vs. demethylation) on transcription and on the phenotype is well characterized, e.g. class II histone deacetylases control cardiac growth and gene expression in response to stress stimuli (Backs and Olson, 2006).

Thus, DPF3 potentially represents the missing link to explain the high impact of the histone modification status on recruitment of the BAF complex to chromatin target sites. A model of the role of DPF3 in recruitment of the BAF complex is shown below (Figure 4.1).

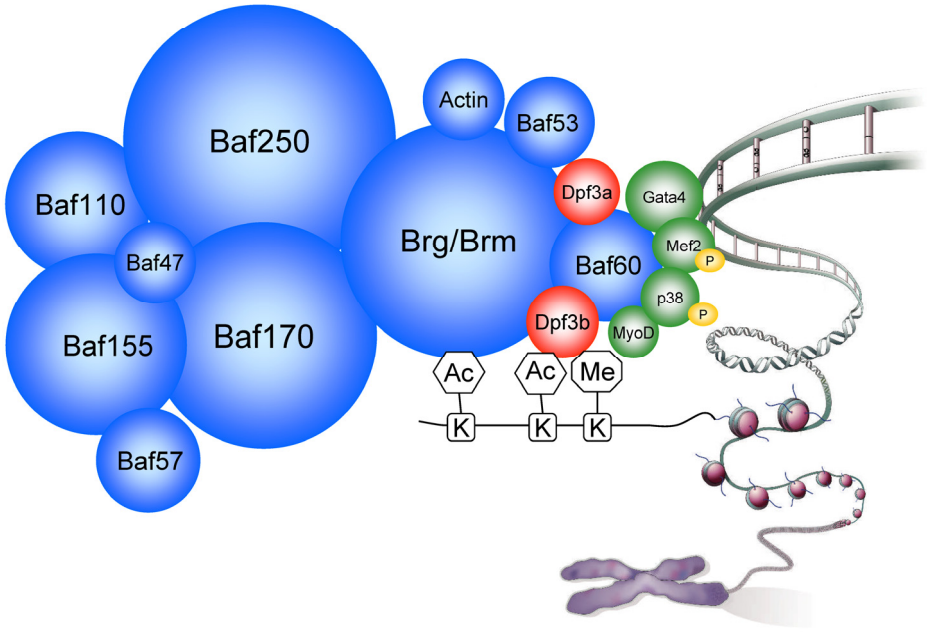

Figure 4.1: Model of BAF chromatin remodeling complex recruitment to target loci by combinatorial read-out of DNA-binding interaction partners and histone modification readers. Proteins of the BAF complex are colored in blue and interaction partners in green. K, lysine; Ac, acetylation; Me, methylation.

4.7 The role of DPF3 in developmental signaling pathways

Gene expression patterns of DPF3 are evolutionary conserved in vertebrates. Therefore, it was assumed that its transcriptional regulatory elements are also well conserved among species. Consequently, the human *DPF3* promoter region was analyzed for the presence of evolutionary conserved elements and potential transcription factor binding sites.

Analysis of the upstream regulatory elements revealed two regions of high evolutionary conservation when promoter sequences from human were compared to other vertebrates.

The conserved region proximal to the transcription start site was around 1 kb (core promoter), in which a minimal region of around 385 bp (minimal promoter) sufficient for transcriptional activation in reportergene assays resided. The size of the *DPF3* minimal promoter was in good agreement with genome-wide data obtained from human, where it has been shown that 91 % of putative promoters derived from 550 bp of genomic sequence immediately upstream of a collection of full-length cDNA clones have promoter activity when assayed using luciferase-based transfection in four human cultured cell types (Trinklein *et al.*, 2003). Moreover, analysis of 900 putative human transcriptional promoters in the ENCODE regions showed that deletions in the -350 to -40 promoter region upstream of transcription start sites resulted in decreased reporter gene signals (Cooper *et al.*, 2006). This region roughly corresponds to the region that is conserved between the promoters of orthologous genes in mice and humans (Taylor *et al.*, 2006).

A closer investigation of the level of sequence conservation showed that the *DPF3* minimal promoter is conserved in species from human to fish (Figure 4.2). To a lesser extend, this was also true for the *DPF3* enhancer region. Sequence conservation was observed from human to chicken. The fact that this region is only conserved in higher vertebrates indicates that it might be involved in regulating the more complex expression pattern of Dpf3 in higher vertebrates.

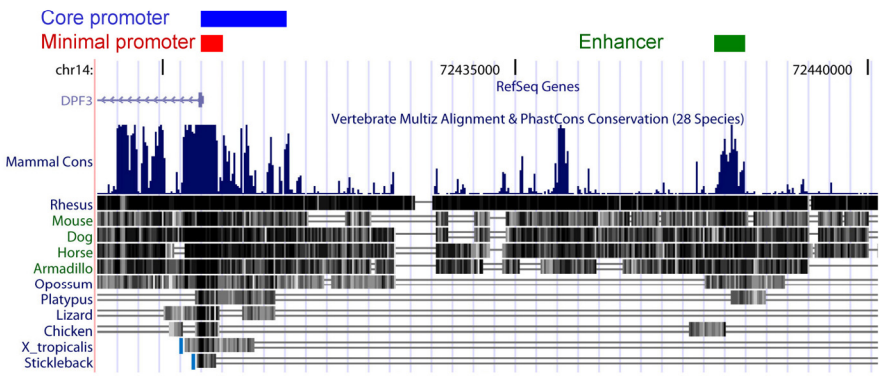

Figure 4.2: UCSC Genome Browser Vertebrate Multiz Alignment of the *DPF3* promoter identifies regions of evolutionary conservation.

The *DPF3* core promoter was analyzed for the presence of potential transcription factor binding sites.

Among the identified transcription factor binding sites in the *DPF3* minimal promoter, those for Nkx2.5, COUP-TFs (NR2F2s) and Mef2a were shown to activate the reporter construct, whereas Sall4 acted as a repressor.

A prerequisite for regulatory interactions *in vivo* is co-expression of transcription factor and target gene. Comparison of gene expression patterns showed that Nkx2.5, COUP-TFs, Mef2a as well as Sall4 are partially co-expressed with Dpf3 in the developing heart and skeletal muscle of mouse embryos (Edmondson *et al.*, 1994; Kaynak, 2005; Koshiba-Takeuchi *et al.*, 2006; Pereira *et al.*, 1999; Prall *et al.*, 2007).

Nkx2.5 is a homeobox transcription factor essential for heart development (Lyons *et al.*, 1995). Mice lacking Nkx2.5 have severe defects in heart looping and chamber formation. Moreover, differentiation of cardiac progenitor cells from the second heart field is affected (Prall *et al.*, 2007). Nkx2.5 is expressed at the cardiac crescent stage at E7.5 in cardiomyocytes and expression persists throughout development of the heart.

To test if the predicted binding sites are indeed responsible for activation, site-directed mutagenesis was performed for the Nkx2.5 and COUP-TF binding sites. This resulted in reduced transcriptional activation, showing that these were indeed the functional binding sites. Binding of Nkx2.5 and COUP-TF2 to these regions was furthermore confirmed in electromobility shift assays.

The essential role for nuclear hormone receptor COUP-TF2 in heart development is reflected by defects in inflow tract development of COUP-TF2 knockout mice (Pereira *et al.*, 1999). Moreover, this factor is important for proper development of skeletal muscle (Lee *et al.*, 2004). COUP-TF2 is expressed in the posterior region of the inflow tract, the somites and skeletal muscle.

Sall4 belongs to the Spalt-family of transcription factors and has been described as a transcriptional repressor as well as activator (Sweetman and Munsterberg, 2006). Overexpression of Sall4 led to transcriptional repression of the *DPF3* minimal promoter. Sall4 is expressed throughout the embryonic mouse heart, with lower expression levels in the interventricular septum and Sall4-null embryos die due to heart- and limb defects, in which Sall4 is expressed during development (Koshiba-Takeuchi *et al.*, 2006). The specific binding properties of Sall4 could not be further described due to lack of a known binding site. Therefore, it is currently unknown how Sall4 mediates its repressive effect. It has been shown that Sall4 can interact with other transcription factors, such as Tbx5.

MEFs are DNA-binding transcription factors crucial for heart and skeletal muscle development (Black and Olson, 1998). Mef2a is active throughout the heart starting at E8.5 and in the myotome from E9.5. Knockdown of mef2a in zebrafish embryos impairs cardiac contractility and sarcomere assembly (Wang *et al.*, 2005). Mef2a-deficient mice commonly die within the first week of life due to dilation of the right ventricle, myofibrillar fragmentation and mitochondrial disorganization (Wang *et al.*, 2005).

The transcriptional regulation of *DPF3* by Mef2a was also tested in luciferase reportergene assays. The previously characterized *DPF3* mininmal promoter contains a putative Mef2 binding site was tested either alone or in combination with 4 consecutive repeats of additional putative Mef2 binding sites (Mef2.1 and Mef2.2) located in the ChIP-enriched region upstream of the core promoter. Co-transfections in HEK293T cells revealed an endogenous activation of the core promoter alone, which was additionally enhanced by Mef2a overexpression. Fusion constructs of the core promoter and the Mef2.1 and Mef2.2 sites showed that transcriptional activity was additionally enhanced only by the Mef2.1 site, supporting a role for Mef2a as a regulator of *DPF3* through combinatorial effects on the Mef2.1 and Mef2.3 sites.

Chromatin immunoprecipitation showed that Mef2a binds to the *Dpf3* promoter in HL-1 mouse cardiomyocytes. In support of this result, siRNA-mediated knockdown of *Mef2a* led to reduced expression of *Dpf3* in the same cell line, indicating that Mef2a is an upstream regulator of *Dpf3*.

The *DPF3* enhancer region was also tested for transcriptional activation in reportergene assays. Despite the presence of several potential transcription factor binding sites, no regulatory factors could be identified. Nevertheless, reportergene assays monitoring the basal activity of the construct in different cell lines showed a preferential activation in muscle cells.

Therefore, the enhancer is probably targeted by signaling molecules of myogenic pathways such as the p38 MAPK-, BMP-, FGF and Wnt-pathways (Pownall *et al.*, 2002).

In summary, the promoter analyses resulted in a model, in which the *DPF3* core promoter is regulated through combinatorial action of activating and repressive transcription factors. The precise signaling molecules mediating the muscle-specificity of the DPF3 enhancer remain to be determined (Figure 4.3).

As a consequence of transcriptional activation, DPF3 can fulfill its role as a component of the BAF chromatin remodeling complex. DPF3 has been shown to localize to the chromatin

region of genes encoding sarcomeric proteins, cardiac (and muscle) transcription factors as well as other nucleosomal proteins.

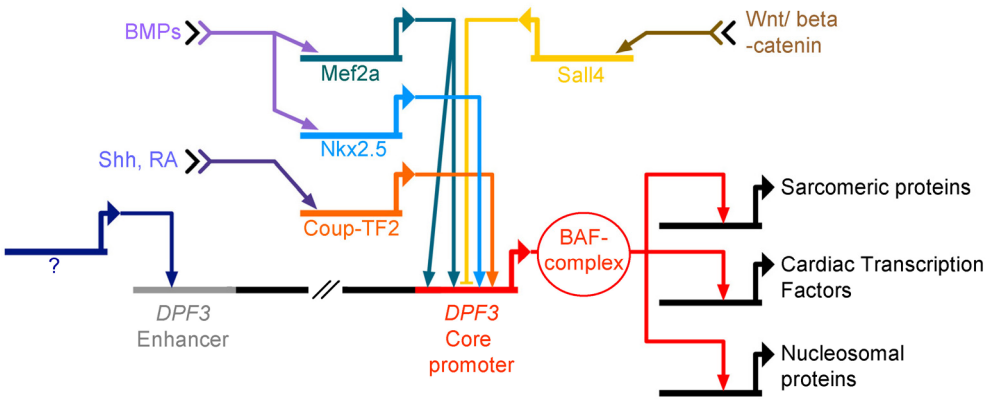

Figure 4.3: Model of the signaling pathways implicated in transcriptional regulation of *DPF3* and its downstream targets. Double arrows indicate diffusible morphogens that have been shown to regulate expression of the transcription factors Mef2a, Nkx2.5, Coup-TF2 and Sall4 (Bohm *et al.*, 2006; Karamboulas *et al.*, 2006; Leng *et al.*, 1996; Prall *et al.*, 2007). Model created with Biotapestry (Longabaugh *et al.*, 2005).

The concept of multifactorial regulation of genes is well established and has been studied in detail. For example, the regulatory elements of the *Nkx2.5* gene have been studied *in vitro* and *in vivo*, resulting in the precise delineation of promoter and enhancer elements. It has been shown that the spatiotemporal expression pattern of *Nkx2.5* is regulated via distinct promoters and enhancers which are targets of Bmp signaling (Chi *et al.*, 2005). Nkx2.5 expression at crescent stage is initiated by proximal regulatory elements, whereas expression during chamber septation and maturation is maintained by distal enhancers located around 12 kb upstream of the transcription start site. Regulation of the *Nkx2.5* gene reflects the concept of modular transcriptional regulation (Schwartz and Olson, 1999).

Bearing in mind the high level of sequence conservation between transcription factors in species as distinct as fruitfly and mouse, the addition of regulatory elements allows for the acquisition of further expression domains. This may enable the integration of additional cells types into the heart, ultimately leading to the development of additional cardiac structures. Thus, expansion of regulatory elements reflects one of the evolutionary tools used for development of the four-chambered heart (Firulli and Olson, 1997).

In order to evaluate the results from reportergene assays *in vivo*, transgenic mice were generated that expressed beta-galactosidase under the control of the aforementioned *DPF3* promoter regions and their mutated derivates. Thus, stable mouse lines were generated for the wildtype *DPF3* minimal promoter, those containing the Nkx2.5 or COUP-TF binding site deletions, as well as the *DPF3* enhancer.

Two transgenic mouse lines were obtained for the wildtype *DPF3* minimal promoter. Even though both lines (1001, 1004) showed expression of the transgene in the developing heart, the distribution of signal was not identical. Transgene expression in line 1001 was restricted to the anterior region of the linear heart tube and future right ventricle early in development and subsequently became restricted to the interventricular septum (IVS) of the heart. Expression was observed in the myocardial compartment of the heart.

On the other hand, the more widely distributed cardiac expression pattern of line 1004 was only observed in cells of the endocardium. In addition, robust transgene expression was observed from the earliest stages of somitogenesis and onwards as well as in the developing limbs.

The fact that distinct expression patterns are generated using the same transgene DNA is partially due to the integration site within the host genome, which is achieved by random integration. For example, a transgene can be silenced upon integration into a region of condensed chromatin (heterochromatin) or be influenced by other, adjacent promoter elements. This is commonly referred to as a position effect (Palmiter and Brinster, 1986).

To evaluate the possible influence of the integration site on transgene expression, the genomic loci containing the transgene were determined by inverse PCR. This modification of the PCR allows for determination of unknown sequence adjacent to a region of known DNA sequence (Ochman *et al.*, 1988). Using genomic DNA from tail cuts of the individual transgenic animals, integration sites could be determined for 4 out 6 lines analyzed. With regard to the differences in expression between wildtype *DPF3* minimal promoter lines 1001 and 1004, it was interesting to see that the transgene of line 1001 was integrated into intronic region of the *Clock* gene. This gene has been shown to be ubiquitously expressed, suggesting that the IVS-specific expression in this line is not mediated by the genomic locus. As the transgene might disrupt at least one copy of the *Clock* gene, it is interesting to note that mutation of the *Clock* gene affects circadian rhythm, but is non-lethal (Vitaterna *et al.*, 1994). Therefore, line 1001 appeared phenotypically normal.

The transgene in line 1004 had integrated in the upstream region of a locus encoding two transcription factors of the Forkheadbox family, namely Foxc2 and Foxl1. Comparison of the transgene expression pattern of line 1004 with the endogenous mRNA expression patterns of the F-box proteins revealed a striking overlap between line 1004 and Foxc2 (Figure 4.4). This strongly suggested that the transgene was, at least partially, under the control of regulatory elements of the Foxc2 gene.

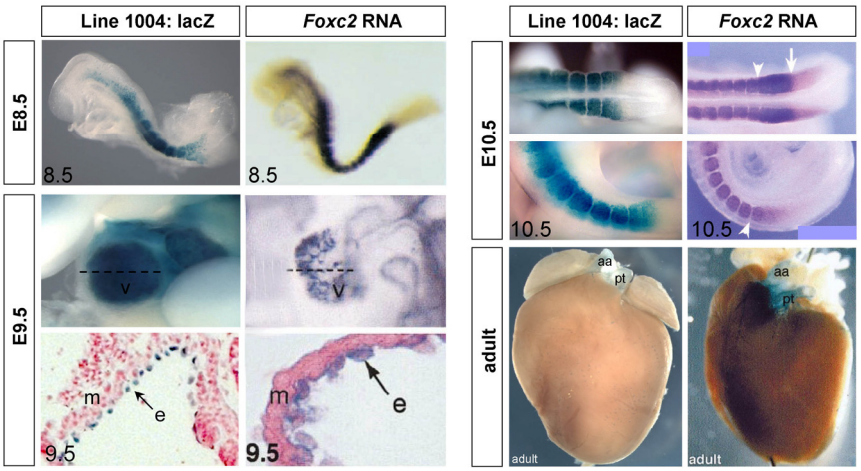

Figure 4.4: Comparison of transgenic line 1004 and *Foxc2* RNA *in situ* hybridization. *Foxc2* RNA *in situ* hybridization images modified from (Kume *et al.*, 2001; Winnier *et al.*, 1999).

The transgene integration sites of Coup-TF deletion lines 2002 and 2027 were also analyzed. In the latter, the insertion was found in a gene-depleted region, thus excluding influence of surrounding regulatory elements.

On the other hand, the integration site in line 2002 was found in the intron 5 of an unannotated gene similar to Cadherin-11 (Osteoblast-cadherin, OSF-4). Interestingly, Cdh-11 is expressed in a proximo-distal gradient within the limb buds as well as the trabeculae of the outflow tract of the heart, resembling the observed expression pattern of line 2002. Therefore, as in the case of line 1004, it is likely that the lacZ expression pattern is influenced by regulatory elements of the gene similar to Cadherin-11 (Simonneau *et al.*, 1995).

These discrepancies in expression patterns suggested that more transgenic lines would be needed. Therefore, transgenic mice were generated and directly analyzed in FO animals, meaning that no stable cell lines were established this time. Unfortunately, out of 23 animals analyzed, offspring of only one mouse carried the transgene. As in lines 1001 and 1004, the transgene was expressed in the developing heart of the embryos. The pattern of expression within the heart was restricted to the anterior portion of the developing right ventricle adjacent to the IVS. Outside the heart, the transgene was again expressed in the somites.

Analysis of the mutated versions of the *DPF3* minimal promoter in transgenic mice was performed by generating two stable mouse lines for each construct.

Deletion of the Coup-TF binding site within the *DPF3* minimal promoter in transgenic mice led to complete loss of transgene expression in line 2027, aside from faint expression in the developing eyes. On the other hand, the transgene was still active in line 2002. Specifically, transgene expression was observed in the outflow tract region of the heart throughout

development, manifested in strong expression in the aorta and pulmonary artery at E15.5. In addition to the heart, expression was observed in the developing limbs, specifically in the zone of polarizing activity (ZPA) at E11.5 and in the progression zone of the outgrowing fingertips at E15.5. Interestingly, the genomic integration site of the transgene was determined to be in the locus of an unannotated gene 6820431F20Rik on chr8:19,991,120-19,991,510, with high similarity to Cadherin-11. As Cadherin-11 expression has been reported in the trabeculae of the cardiac outflow tract, it would be interesting to test the endogenous expression pattern of 6820431F20Rik by *in situ* hybridization and compare these date to the expression profile of line 2002.

Neither of the two mouse lines containing the Nkx2.5-binding site deletion (lines 2511 and 2521) construct expressed the transgene in the heart. Line 2511 had an expression domain in the midbrain and the septum transversum, which will give rise of the liver, whereas line 2521 showed faint expression in midbrain only.

To exclude position effects in the analysis of transgenic constructs in mice, it is possible to make use of the Rosa26R locus, which can be efficiently targeted without affecting the host genome (Kisseberth *et al.*, 1999). Nevertheless, the process of targeted introduction of DNA into the mouse genome is substantially more laborious.

Another possibility is to increase the number of analyzed mice using transient transgenic embryos, which can be analyzed as early as one week after generation, excluding the lengthy process of breeding mouse lines. The low efficiency in generating transient lines of the *DPF3* minimal promoter may have been due to the quality of the obtained eggs or insufficient transgene DNA quality.

The transgenic mouse line 17 obtained for the *DPF3* enhancer showed an interesting cardiac expression pattern that was restricted to the left ventricle of the developing heart. Regional-specific expression of genes during cardiogenesis reminiscent of the *DPF3* enhancer have been described previously. The transcription factors Hand1 and Hand2 are expressed in a complementary fashion and are restricted to regions of the heart tube fated to form the left and right ventricles, respectively. Hand1 is predominantly expressed at the outer curvature of the left ventricle in the looping heart during mouse development (Srivastava *et al.*, 1997). Therefore, it acts as a chamber-specific transcriptional regulator of downstream targets such as *Nppa* (*ANF*), the major secretory product of the heart (Morin *et al.*, 2005).

A direct influence of Hand1 on expression of the *DPF3* enhancer could not be determined in reportergene assays. Nevertheless, the combinatorial action of this or other TFs may be involved in *DPF3* enhancer regulation.

The combination of these results and data obtained from reportergene assays suggested that the *DPF3* minimal promoter is able to drive transgene expression in the developing heart, although the expression domains within the heart varied among the different lines.

This suggests that additional regulatory domains are necessary for the precise spatio-temporal regulation of expression. As the *DPF3* enhancer is also able to drive transgene expression in the developing heart, it would be interesting to combine these two regions in a transgene construct and analyze its expression pattern. Furthermore, it should be noted that a third region of moderate evolutionary conservation in between the *DPF3* core promoter and enhancer could be identified (Figure 4.), which might contribute regulatory sequences in higher vertebrates. This combinatorial mode of regulation would fit into well-established concepts of transcriptional regulation through distinct modules (Schwartz and Olson, 1999).

In addition to mouse lines expressing lacZ, the *DPF3* minimal promoter was fused to *Cre* Recombinase in an attempt to generate a mouse line expressing Cre Recombinase in the interventricular septum of the developing heart. Malformations of the IVS are part of the TOF phenotype. The developmental cues regulating the formation of the IVS are only beginning to emerge (Franco *et al.*, 2006). Specific expression of Cre-Recombinase in the IVS would allow for IVS-specific ablation of gene function in the mouse embryo and be helpful for the study of IVS development. The activity of the transgene in mice was tested using a reportergene system. Upon crossing of Cre-Recombinase males mice with female mice of the Rosa26R lacZ strain, Cre-Recombinase activity can be monitored by X-Gal staining. This is possible, because the normally inactive form of beta-galactosidase in Rosa26R mice restored by Cre. Opposite to direct transcriptional readout of transgene activity, the analysis of *Cre*-Recombinase activity results in a fate-map of all cells of the embryo in which the transgene was active, because the restored Rosa26R locus drives ubiquitous expression in the embryo. Unfortunately, the IVS specific expression of transgenic line 1001 could not be replicated. Despite expression within the IVS, the domains of activity were expanded to the atria and sclerotome. Therefore, the line was considered unsuitable for IVS-specific gene ablation.

4.8 Evolution of the d4 protein family

Protein sequence comparisons can be used to analyze the evolutionary origin of proteins and protein families. Consequently, the origin of the d4 protein family can be traced back to the roundworm *C. elegans,* which contains the most derived d4 sequence. In invertebrate species, only a single d4 protein is found, defined by the presence of the unique N-terminal 2/3-domain.

The earliest members of the d4 family in the roundworm *C. elegans* and the fruit fly *D. melanogaster* do not have the C2H2 type zinc finger, whereas d4 proteins of sea squirt *C. intestinalis* and sea urchin *S. purpuratus* already contain this domain.

C. elegans and *D. melanogaster* belong to the protostomes, which are relatively primitive invertebrates. On the other hand, the sea squirt (*C. intestinalis*) and sea urchin (*S. purpuratus*) belong to the deuterostomes and are more recent species in terms of evolution.

This suggests that the C2H2-type zinc-finger has been added later in evolution of the d4 protein family.

Opposite to the 2/3-domain, the other protein domains of the d4 family, namely the C2H2-type zinc finger and the PHD fingers are found in all eukaryotes, including unicellular organisms such as yeast (*S. cerevisiae).* Both types of domains are present in a broad range of protein families.

The yeast C2H2-type zinc finger protein with the highest similarity to the C2H2-type zinc finger of human DPF3 is Sfp1. This protein has been shown to play a role in transcriptional regulation of ribosomal proteins as well as cell size control in response to nutrient- and stress stimuli (Cipollina *et al.*, 2008). It has been shown that Sfp1 translocates to the nucleus upon activation of the TOR pathway, leading to modulation of ribosomal protein gene expression (Marion *et al.*, 2004). Interestingly, Sfp1 has been genetically associated with the function of SWI4 (Jorgensen *et al.*, 2002). After SWI/SNF chromatin remodeling complexes (and SAGA) have modified chromatin, SBF (Swi4p–Swi6p) can bind to the remodeled DNA, where it presumably recruits the general transcription machinery (Aalfs and Kingston, 2000).

The yeast PHD finger proteins with the highest similarity to the human DPF3 PHD fingers are Sant-domain-containing 2 (Snt2) for human PHD1 and jumanji homeodomain 2 protein (Jhd2p) for human PHD2.

Snt2 has been implicated in stress response signaling in yeast (Rokhlenko *et al.*, 2007). Aside from the PHD finger, it contains a Bromo-adjacent homology domain (BAH), which has been shown to have DNA and nucleosome binding properties (Connelly *et al.*, 2006) .

Yeast Jhd2p is a histone 3 lysine 4 trimethyldemethylase (Liang *et al.*, 2007). The jumonji family regulates chromatin, gene expression, and controls development through various signaling pathways (Takeuchi *et al.*, 2006).

Taken together, the functional roles of the yeast proteins containing domains with the highest similarity to d4 proteins already show an implication in transcriptional regulation on the chromatin level. The combination of the different, previously separate domains within d4 proteins will have allowed for new mechanisms of transcriptional regulation during evolution.

The generation of novel proteins can be achieved in two ways. Either by gene-, genome-, or chromosome-duplication followed by divergence through genetic drift and mutation or through recombination of sequences encoding certain protein domains (Schmidt and Davies, 2007). Both events can be achieved in one generation, although the long-lasting process of sequence divergence is necessary to obtain novel protein functions in the first scenario.

As the C2H2-type zinc finger and the two PHD fingers of *C. elegans* d4 have clear orthologous domains in yeast, it is reasonable to assume that these domains have been combined in a recombination event. On the other hand, the unique 2/3-domain has no sequence similarity to any yeast protein and thus represents a domain in which previously non-functional sequence probably has acquired a novel function.

As the 2/3-domain is an invention of the metazoans, it is interesting to note that genetic experiments in mice have shown that the 2/3-domain and the C2H2-type zinc finger of the related protein PHF10 play a role in maintenance of neural stem cell proliferation (Lessard *et al.*, 2007a), indicating a role in the regulation of multicellular organisms.

The N-terminal 2/3-domain and the C2H2-type zinc finger of the d4 family are especially interesting with regard to the unusual isoform DPF3a, which does not bind to histones.

Three paralogs of the d4 protein exist in all vertebrates. These proteins have been generated by genome duplication events, which have occurred at the origin of vertebrate species (Babushok *et al.*, 2007). After duplication, the paralogs underwent adaptive evolution, acquiring novel expression domains. This is reflected by the lack of sequence conservation between promoter elements of *DPF1*, *DPF2* and *DPF3*, which probably accounts for differences in the expression patterns. In terms of molecular function, the members of the d4 family are probably all involved in chromatin remodeling and transcriptional regulation, due to the high level of sequence and domain conservation.

Sequence comparison of proteins of the d4 family showed that the shorter isoform DPF3a, which contains only a truncated form of the first PHD finger, is exclusively present in the higher vertebrate species from chicken to human. The early cardiac expression pattern of

Dpf3a during embryogenesis suggests that it is involved in the development of the four-chambered heart in these species, which is substantially different from that of lower vertebrates such as reptiles and amphibians. Therefore, the specific function of Dpf3a could not be determined in zebrafish and will have to be addressed in a higher vertebrate model organism.

Sequence comparison furthermore revealed that the d4 protein of sea urchin (*S. purpuratus*) is the protein most closely related to human DPF3 before the occurrence of a gene duplication event (Figure 3.41). To analyze the expression domains of sea urchin d4, *in situ* hybridization was performed. This allowed for comparison of the initial expression pattern of the d4 protein with that of vertebrate d4 family members. *d4* mRNA expression was observed in the vegetal plate at blastula stages as well as early gastrula. During gastrulation, *d4* expression was restricted to the anterior and posterior ends of the fore- and hindgut, a region home to cells of mesodermal and endodermal origin. Notably, no expression was observed in ectodermal cells. On the other hand, vertebrate d4 family members Dpf1 and Dpf2 and Dpf3 are all expressed in neuronal cells suggesting that these expression domains have been aquired later in evolution. This implicates that the expression of Dpf3 in heart and skeltetal muscle, which are of mesodermal origin, is evolutionary highly conserved. Taken together, *d4* expression in sea urchin was observed in tissues of mesodermal origin, whereas no expression was observed in cells of the ectoderm, which give rise to neuronal tissues.