1 Introduction

1.1 Dpf3, the third member of the d4 protein family

The transcription factor Dpf3 (d4 protein family, member 3) has initially been identified and cloned from mouse and chicken cDNA libraries (Ninkina *et al.*, 2001). It belongs to the d4 family of proteins, which consists of three members in all vertebrates. The other members of the d4 family are Dpf1 (Neuro-d4) and Dpf2 (Ubi-d4/ Requiem). Dpf1 has been cloned as a neurospecific, developmentally regulated rat gene (Buchman *et al.*, 1992). Recently, Dpf1 has been shown to be expressed in differentiated neurons of the postmitotic zone in the mouse embryo (Lessard *et al.*, 2007a). Dpf2 is a ubiquitously expressed protein implicated in apoptosis caused by IL-3 deprivation of the IL-3 dependent myeloid cell line (Chestkov *et al.*, 1996; Gabig *et al.*, 1994; Mertsalov *et al.*, 2000; Theodore *et al.*, 1998).

Proteins of the d4 family share a high level of sequence similarity. They are characterized by an N-terminal 2/3-domain (encoded by exons 2 and 3), which is unique to this family and of currently unknown function. A putative nuclear localization signal (NLS) is present within this domain, suggesting a role in regulation of subcellular translocation. Furthermore, proteins of the d4 family feature a potential nuclear receptor interaction domain (NID) and a single classical C2H2-type zinc finger ($CX_2CX_{12}HX_2H$). DNA binding potential as well as interactions with RNA and proteins have been described for this domain when found in multiple repeats within a protein (Iuchi, 2001). The C-terminus of proteins of the d4 family is characterized by the presence of a tandem plant-homeodomain (PHD) and has previously been called d4 domain (Chestkov *et al.*, 1996). With regard to the C-terminus, Dpf3 is exceptional to the d4 family, as two splice variants have been identified in the higher vertebrates chicken, mouse and human that give rise to one isoform having the full double PHD finger (Dpf3b) and one isoform having only a truncated version of the first PHD finger (Dpf3a) (Figure 1.1). The role of the plant-homeodomain will be discussed in detail in a separate chapter of the introduction (see Chapter 1.5.3).

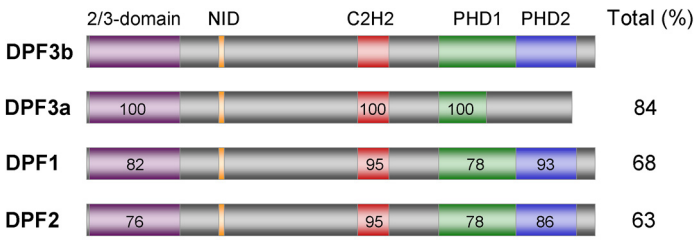

Figure 1.1: Sequence similarities of human d4 famliy proteins compared to DPF3b in %. Purple, 2/3-domain; orange, nuclear receptor interaction domain (NID); red, C2H2-Krüppel-like zinc finger; green, plant-homeodomain 1 (PHD1); blue, plant-homeodomain 2 (PHD2). DPF3b, AAX20019.1; DPF3a, NP_036206; DPF1, EAW56760; DPF2, NP_006259.1.

The d4 family has evolved over a period of around 500 million years, as the founding protein can be traced back to the roundworm *Caenorhabditis elegans* (Nabirochkina *et al.*, 2002). Only a single d4 gene is present in the *C. elegans* genome and it lacks the C2H2-type zinc finger. The first d4 protein that contains all domains can be found in an early member of the deuterostomes, the sea squirt *Ciona intestinales*. Three paralogs of the d4 protein can first be found in the genome of primitive vertebrates such as the zebrafish *Danio rerio*, which is likely the result of a genome duplication event that occurred at the origin of vertebrates (Kasahara, 2007).

Prior to this study, Dpf3 had been poorly characterized and data regarding its function had not been available. Due to the presence of the C2H2-type zinc finger and the PHD fingers, it was annotated as a transcription factor. Human DPF3 was identified by our group in a screen for markers of diseased and normal human myocardium (Kaynak *et al.*, 2003). Both splice variants were shown to be upregulated (Figure 1.2) in myocardial ventricular samples of patients with Tetralogy of Fallot (TOF), a complex congenital heart defect (see chapter 1.3).

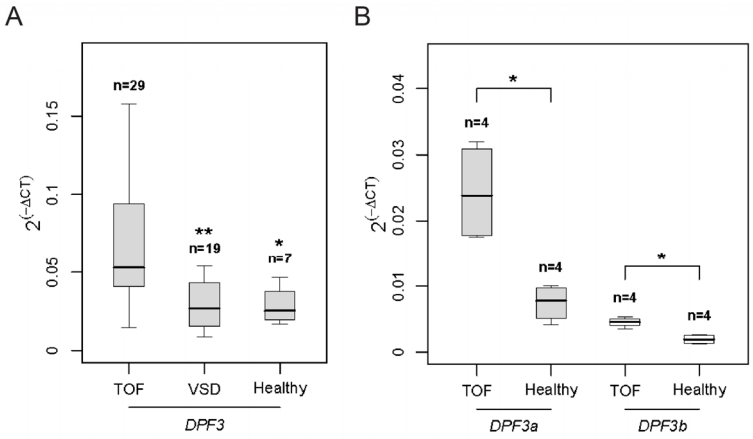

Figure 1.2: Expression of *DPF3* mRNA in malformed and normal human hearts. A) Real-time PCR analysis of *DPF3* mRNA levels in myocardial, right ventricular tissue from patients with Tetralogy of Fallot (TOF), ventricular septal defect (VSD) and healthy controls. B) Analysis of splice variant specific expression of *DPF3a* and *DPF3b* in TOF patients and healthy controls. (* p<0.01; ** p<0.01). Scale-bars represent ± SEM. Figure from accompanied manuscript.

Expression analyses revealed that *Dpf3* is expressed in heart, skeletal muscle and neuronal tissues in mouse embryos (Figure 1.3) (Kaynak, 2005). Whole mount *in situ* hybridization in mouse embryos showed that *Dpf3a* is expressed in the first differentiating cardiomyocytes of the cardiac crescent at E7.5, in the linear heart tube and the first somites at E8.0. During heart looping at E9.0, expression is observed in the inflow tract, primitive atrium, ventricle and outflow tract, as well as the somites. At E9.5, *Dpf3a* is expressed throughout the heart including the sinus venosus, common atrium, left and right ventricles and the outflow tract.

Additional expression was observed in the septum transversum and somites. At E10.5, the intensity of expression in the heart decreases with more pronounced signals observed in the lateral walls of the ventricles and in the atria and weaker signals present in the outflow tract. In addition, *Dpf3a* is expressed in the developing liver and midbrain. At E11.5, strong expression is observed in the region of the forming interventricular septum and the posterior-lateral walls of the ventricles. Section in-situ hybridization revealed that Dpf3a expression is restricted to the myocardial compartment of the heart. Further *in situ* hybridizations using a *Dpf3* probe detecting both splice variants revealed a similar expression pattern.

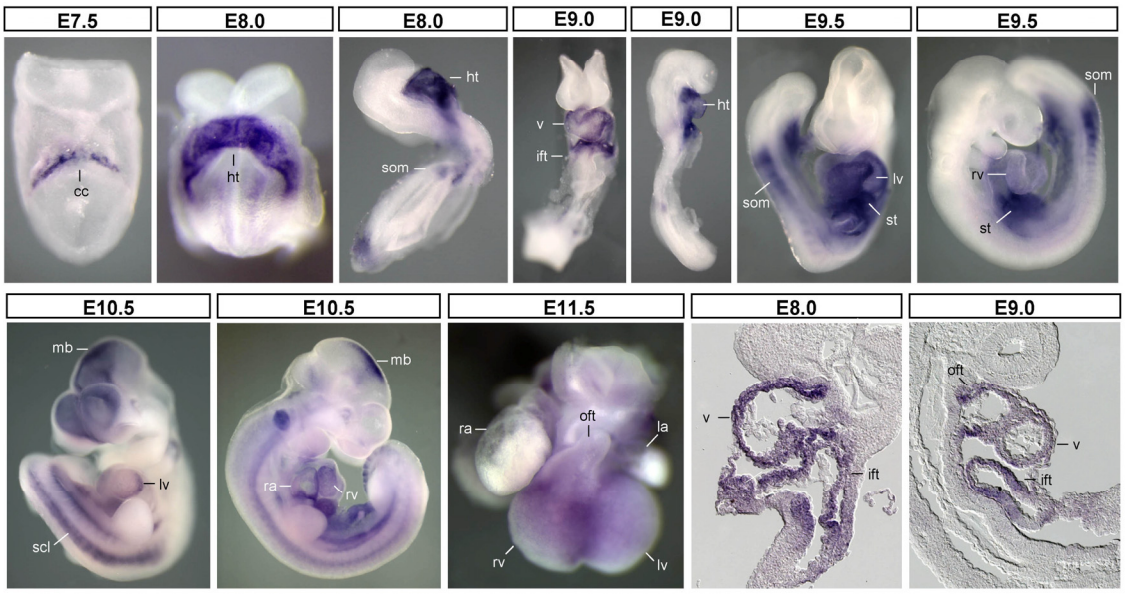

Figure 1.3: Expression pattern of *Dpf3a* mRNA during mouse development (ventral and lateral views and close-ups). cc, cardiac crescent; ht, heart tube; v, ventricle; ift, inflow tract; oft, outflow tract; lv, left ventricle; rv, right ventricle; la, left atrium; ra, right atrium; som, somites; st, septum transversum; mb, midbrain; scl, sclerotome; a, atrium; nt, neural tube; v, ventricle; a, atrium. Figure from accompanied manuscript.

Aside from descriptive analyses, only few publications have been available about DPF3 in the literature. Polymorphisms in the 5' region of *DPF3* have been shown to be associated with increased risk of breast cancer development, lymph node metastases, age of onset, and tumor size in women of european ancestry (Hoyal *et al.*, 2005). Furthermore, *Dpf3* transcripts were enriched in a cell population of Nkx2.5-positive cardiomyocytes from mid-gestation mouse embryos (Masino *et al.*, 2004). Finally, *Dpf3* has been implicated in the transition of neuronal precursors to differentiated neurons (Lessard *et al.*, 2007b).

Due to upregulation in patients with a congenital heart defect (TOF) and the expression pattern of *Dpf3,* its role in heart and skeletal muscle development was analyzed.

1.2 Heart development

The adult heart is a complex muscular pump whose function is essential to the organism. It enables the distribution of blood - and along with it - oxygen, nutrients and waste. During embryonic development the heart is the first organ to form and function from flies to humans (Olson, 2006). Heart development is an intricate process tightly controlled by an evolutionary conserved gene program. Perturbations to this program can have deleterious effects on the embryo often leading to embryonic lethality or congenital heart defects (CHD) (Ransom and Srivastava, 2007). During the course of evolution, the heart has evolved from a simple contractile dorsal vessel found in the fruitfly *Drosophila melanogaster* to a complex, four-chambered organ in higher vertebrates such as birds and mammals (Bodmer and Venkatesh, 1998).

The key steps of cardiogenesis and skeletal muscle development as well as its implications on the development of congenital heart defects will be introduced with special emphasis on higher vertebrates and the zebrafish, a model organism for heart development.

In humans, the embryonic heart begins to pump blood at only three weeks of gestation. The heart is formed through the contribution of different cell types and its assembly is susceptible to perturbations that can ultimately lead to CHD.

By week two of human development (mouse E6.5), cardiogenesis is initiated by the commitment of mesodermal cells to the cardiac lineages, namely the first (FHF) and second heart field (SHF) (Buckingham *et al.*, 2005) (Figure 1.4).

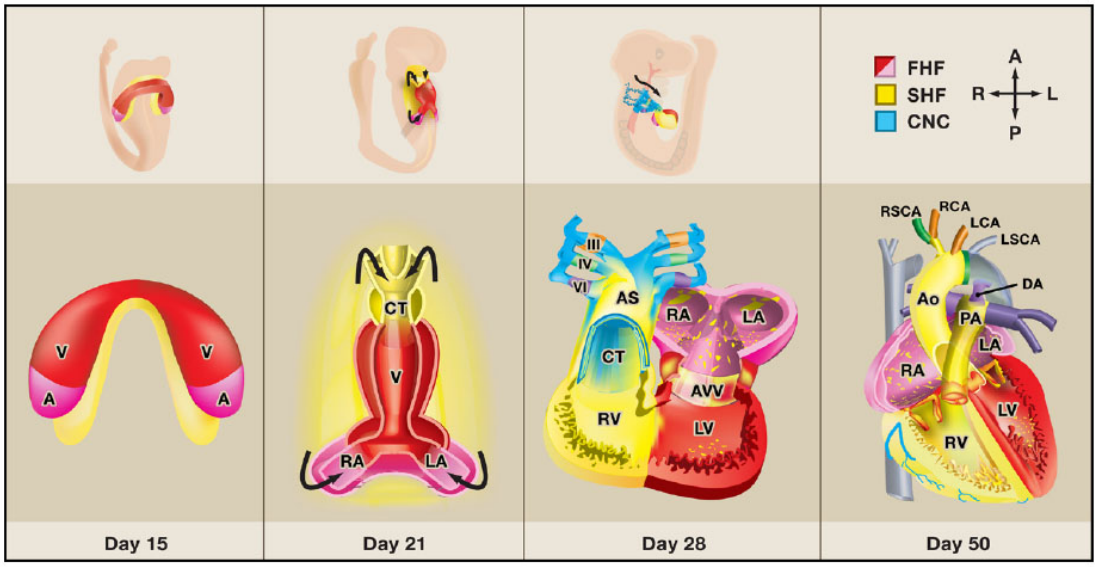

Figure 1.4: Mammalian heart development. Frontal views of cardiac precursors during development in whole mount and close-up views. (First panel) First heart field (FHF) cells form the cardiac crescent with second heart field (SHF) cells medial and anterior to the FHF. (Second panel) SHF cells lie dorsal to the straight heart tube and start to migrate into the ends of the tube to form the right ventricle (RV),

conotruncus (CT), and part of the atria (A). (Third panel) Following heart looping, cardiac neural crest (CNC) cells migrate into the outflow tract to septate the outflow tract. (Fourth panel) Septation of the ventricles, atria, and atrioventricular valves (AVV). V, ventricle; LV, left ventricle; LA, left atrium; RA, right atrium; AS, aortic sac; Ao, aorta; PA, pulmonary artery; RSCA, right subclavian artery; LSCA, left subclavian artery; RCA, right carotid artery; LCA, left carotid artery; DA, ductus arteriosus. Staging indicates days of human development. Figure from (Srivastava, 2006).

The bilaterally positioned cell pools then converge at the midline of the embryo to form the cardiac crescent at 15 days of gestation (mouse E7.5). Subsequently, the beating linear heart tube is formed from the crescent around day 21 (mouse E8.0). It consists of a single layer of endocardial cells and an exterior layer of myocardial cells separated by an extracelluar matrix.

At the ends of the heart tube are the outflow tract (OFT) and the inflow tract (IFT). The OFT will ultimately become the aorta and pulmonary arteries, the IFT will become the atrioventricular canal. Electrical activity can be measured in a posterior-anterior gradient, which ensures that contraction is initiated at the inflow tract in a uni-directional fashion even before the conduction system is present. Anterior-posterior patterning of the linear heart tube is established by a caudo-rostral wave of retinoic acid signaling (Hochgreb *et al.*, 2003). Retinaldehyde-specific dehydrogenase 2 (Raldh2), which is responsible for retinoic acid synthesis, has been shown to be essential for heart development.

Lineage analyses using the Cre-loxP system as well as retrospective clonal analyses revealed that the linear heart tube serves as a scaffold upon which cells of the SHF are added to build the heart chambers (Buckingham *et al.*, 2005). Cells from the SHF, which lies dorsal and medial to the linear heart tube in the pharyngeal mesoderm, migrate and contribute to both ends of the heart tube (Kelly, 2007). The SHF has been identified by a lacZ gene enhancer trap into the *Fgf10* locus in the mouse (Kelly *et al.*, 2001). In this transgenic mouse line, beta-galactosidase activity has been observed in the embryonic right ventricle and OFT of the heart and in the underlying pharyngeal mesoderm. Fate mapping in the form of Dil labeling confirmed that cells from the pharyngeal mesoderm migrate and contribute to the outflow tract, atria and right ventricle of the heart. Cells of the SHF are highly proliferative and start to differentiate only upon addition to the heart tube. The balance between proliferation and differentiation is controlled by an Nkx2.5/Bmp2/Smad1 negative feedback loop (Prall *et al.*, 2007). Only the left ventricle is of FHF origin.

Combinatorial signaling by BMPs, Shh, Fgfs, Wnts and Notch pathways regulate the cell fates of both lineages. The sources of signaling are located in the surrounding endoderm and mesoderm, although their precise location remains unknown (Zaffran and Frasch, 2002).

At week 4 of development (mouse E8.5), the linear heart tube starts to twist toward the left side of the embryo in a process called cardiac looping (Harvey, 2002). This results in the

parallel arrangement of the future chamber compartments. The molecular pathways controlling heart looping are largely unknown at present. Mice lacking the transcription factors Hand2, Nkx2.5 or Mef2c are embryonically lethal before completion of heart looping (Lin *et al.*, 1997; Lyons *et al.*, 1995; Srivastava *et al.*, 1997). Nevertheless, it is unclear whether these are primary defects or secondary due to the affected myocardium. Defects in heart looping are also often observed as part of a more general defect in the establishment of left-right- (L-R) asymmetry, which affects the whole organism. Asymmetric gene expression and the activity of cilia at the node appear to be the underlying mechanisms controlling L-R asymmetry (Hamada *et al.*, 2002).

Patterning and septation of the OFT are mediated by migratory cells of the cardiac neural crest (CNC) that invade the OFT region at about 6 weeks (mouse E9.0) of gestation (Hutson and Kirby, 2007). Perturbations to neural crest signaling during OFT septation have been shown to cause CHDs (Porras and Brown, 2008). Moreover, OFT remodeling is largely mediated by cells of the SHF. Mice lacking the T-box transcription factor Tbx1, which is expressed in the SHF, die during embryogenesis due to OFT defects (Baldini, 2005). In addition, Pitx2c, Isl1, Mef2c and FoxH1 have been shown to be necessary for OFT remodeling (Buckingham *et al.*, 2005). There is furthermore a clear interdependence between cells of the SHF and CNC. It has been shown that ablation of CNC cells in chicken embryos leads to failure of addition of SHF cells to the outflow tract (Waldo *et al.*, 2005). It appears that signaling from the CNC regulates the expression level of Fgf8 in the pharynx and excess Fgf8 blocks addition of SHF cells to the OFT, resulting in a phenotype that resembles TOF (Hutson *et al.*, 2006).

Development of the cardiac chambers is achieved in a process termed ventricular ballooning (Christoffels *et al.*, 2004). During chamber formation, the primitive myocardium is separated into primary and working myocardium. The primary myocardium is located at the OFT, atrioventricular canal (AVC), and the inflow tract (IFT) and gives rise to the conduction system.

Patterning of the primary myocardium is regulated by the T-box transcription factors Tbx2 and Tbx3 (Hoogaars *et al.*, 2007). The conduction system of higher vertebrates can be subdivided into the nodal components responsible for pulse generation and the His-Purkinje system, which displays high conduction velocities. The nodal compartment consists of the sinuatrial node, located in the right atrium and the atrioventricular node, located at the atrioventricular junction. The His-bundle propagates the signal to the Purkinje fibers, which are embedded in the ventricles.

The working myocardium is located at the chambers and is characterized by high conduction velocity, full force generation as well as good cell-cell coupling and sarcomeric structures (Christoffels *et al.*, 2000). Bmp10 is specifically expressed in the working myocardium and has been shown to be important for proliferation and maturation of the ventricles (Chen *et al.*, 2004).

Morphogenesis of the heart is completed around week 8 of development (E16.5). The cells of the heart become less proliferate and subsequent growth is mainly achieved through hypertrophic growth of already existing cardiomyocytes. Finally, cardiomyocytes completely exit the cell cycle upon birth (Soonpaa *et al.*, 1996).

In summary, the development of the fully functional heart depends on complex network of genetic and protein-protein-interactions that ensure the proper integration of different cell types to form the three-dimensional structure of the heart.

Heart development from the formation of the contractile dorsal vessel of the fruitfly *Drosophila melanogaster* to the four-chambered heart of humans is an evolutionary conserved process. The zebrafish represents an evolutionary intermediate to the aforementioned. Compared to the fruitfly, the zebrafish already possesses clearly defined cardiac chambers, namely the sinus venosus, atrium, ventricle and bulbus arteriosus. These chambers are separated by valves that enable uni-directional blood flow. Cardiac contraction is coordinated by a pace-making conduction system (Chen and Fishman, 2000). Moreover, the zebrafish already has a closed circulatory system that pumps blood, whereas the fruitfly pumps hemolymph through the interstices of tissues rather than through vessels. Also, the zebrafish makes use of gills to efficiently oxygenate its blood. Nevertheless, zebrafish hearts are relatively simple compared to those of higher vertebrates that possess two separate circulatory systems supported by a four-chambered heart.

Despite these anatomical differences, the core molecular networks guiding heart formation have been conserved. Common features between vertebrate and fruitfly cardiogenesis include the induction of cardiac progenitors in the lateral mesoderm through BMP/Dpp signaling, which results in the expression of related cardiogenic transcription factors, particularly Nkx2.5/tinman and GATA4,5,6/pannier (Cripps and Olson, 2002; Zaffran *et al.*, 2002). In addition, the dorsal vessel or tubular heart tube of fruitfly and zebrafish resembles the early linear heart tube of higher vertebrates, making them a model for early vertebrate heart formation. Indeed, the first gene that has been assigned in heart development has been identified in *Drosophila*. Mutation of the *tinman* gene results in failure of dorsal vessel formation (Bodmer, 1993). Soon after, its ortholog Nkx2.5 has been shown to have an essential role in

vertebrate heart formation and congenital heart disease in humans (Lyons *et al.*, 1995; Schott *et al.*, 1998).

1.3 Human congenital heart disease

Congenital heart defects are referred to as structural malformations of the neonatal heart that compromise its function, including abnormalities of the vessels that connect the heart to the circulatory system, the valves that ensure uni-directional blood flow and the walls of the chambers which septate the heart. Morover, congenital heart defects are characterized by cardiac arrhtymias and cardiomyopathies. A complex congential heart defect is known as Tetralogy of Fallot and characterized by four clinical features (Figure 1.5). First, a ventricular septal defect (VSD). Second, an obstruction of the right ventricular outflow tract, characterized by a narrowing at (valvular stenosis) or just below (infundibular stenosis) the pulmonary valve. The stenosis is most likely due to hypertrophy of the septoparietal trabeculae. The degree of stenosis is the primary determinant of disease severity and varies among patients (Bartelings and Gittenberger-de Groot, 1991). The third feature is an overriding aorta. In this case, the aortic valve is not restricted to the left ventricle, but is connected to both chambers. Last, hypertrophy of the right ventricle is observed due to the fact that the right ventricle needs to pump blood past the narrowing of the pulmonary obstruction (Anderson and Weinberg, 2005).

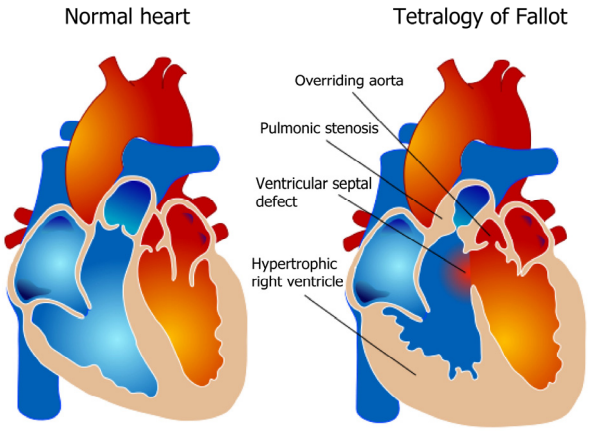

Figure 1.5: Schematic representation of a normal heart and Tetralogy of Fallot (TOF), represented by four clinical features. Overriding aorta, pulmonic stenosis, ventricular septal defect and hypertrophic right ventricle.

TOF is a major cause for the so-called "blue baby syndrome", which represents a cyanotic condition in newborns caused by mixing of oxygenated and deoxygenated blood in the left ventricle through the VSD. Moreover, oxygenated and deoxygenated blood from the

ventricles preferentially flows through the aorta due to the right-to-left shunt. TOF occurs in 3 to 6 out of 10,000 newborns and represents around 5-7 % of all congenital heart defects.
Untreated TOF ultimately leads to cardiac failure with a survival rate of around 60 % after four years (Anderson and Tynan, 1988). Nevertheless, corrective surgery can be applied to completely restore heart function in a very high number of cases.
Congenital heart defects are caused by a complex mixture of genetic defects and environmental factors making CHDs a complex trait. Nevertheless, a variety of heart defects could clearly be associated with single gene defects. It has been estimated that CHDs occur in approximately 8 out of 1000 live births (Hoffman and Kaplan, 2002). As a consequence, CHDs account for the highest mortality rate caused by non-infectious diseases in the first year of life (Hoffman *et al.*, 2004). Thanks to advances in patient diagnosis and corrective surgery, an increasing number of survivors reach childbearing age and can thus transmit the disease at a higher probability than the general population. This allows for a better analysis of the underlying genetic defects of CHDs, but also implicates that improved clinical diagnostics need to be implemented to identitfy populations with increased risk of having offspring with CHD.
The etiology of CHDs can be due to environmental components, genetic defects or a combination of the two. Environmental factors such as exposure of the fetus to teratogens like alcohol, anti-depressants, anti-epileptic drugs or infections during early pregnancy have been shown to cause CHDs and account for about 2 % of CHDs (Starreveld-Zimmerman *et al.*, 1975; Zimmerman, 1991). Moreover, deficiency of retinoic acid, the biologically active form of vitamin A, has been shown to cause CHD (Kastner *et al.*, 1997).
Genetic causes of CHDs have also been identified in syndromic diseases. Syndromes account for about 15-20 % of all CHDs and in many of the syndromes analyzed, the disease-causing gene could be identified(Anderson and Weinberg, 2005). For example, patients with DiGeorge syndrome, a chromosomal deletion of 22q11, have defects in cells derived from the cardiac neural crest and SHF, manifested in interruption of the aortic arch (IAA), persistent truncus arteriosus (PTA), TOF, double outlet right ventricle (DORV) and transposition of the great arteries (TGA). Mouse knock-out studies and mutation analyses in DiGeorge syndrome patients revealed that loss of one copy of *TBX1* is responsible for the heart defects (Jerome and Papaioannou, 2001; Lindsay *et al.*, 2001; Merscher *et al.*, 2001; Yagi *et al.*, 2003).
Holt-Oram syndrome (HOS) leads to limb and heart defects, including atrial septal defect (ASD), TOF and AV conduction defects. HOS is caused by haploinsufficiency of *TBX5*,

representing the first single causative gene for a septation defect (Basson *et al.*, 1997; Li *et al.*, 1997).

Char syndrome, like Holt-Oram syndrome, affects both limbs and heart. The heart defect is characterized by a persistent ductus arteriosus (PDA). *TFAP2ß* has been mapped to the Char syndrome locus and subsequently, a mutation in the PY domain has been identified in patients (Satoda *et al.*, 2000; Zhao *et al.*, 2001).

Nevertheless, most heart defects are non-syndromic and only rare cases have been reported of syndrome-associated gene mutations like *TBX5* to account for isolated heart defects (Basson *et al.*, 1999). A number of other genes have been identified with a clear causal link to CHDs, mostly through linkage analyses of large families. Outflow tract defects, mostly bicuspid aortic valve (BAV), are caused by *NOTCH1* gene mutations, which have been identified in two unrelated families (Garg *et al.*, 2005). Septation of the cardiac chambers is also often affected in CHD and mutations in *NKX2.5*, *GATA4* and *MHC6* have been reported to cause septal defects (Ching *et al.*, 2005; Garg *et al.*, 2003). Among these, *NKX2.5* was the first non-syndromic gene to cause CHD (Schott *et al.*, 1998). Recently, mutations in the transcription factor *CITED2* have been reported in patients with septal defects that lead to impaired function of *TFAP2* coactivation by CITED2 (Sperling *et al.*, 2005). Analysis of these families has provided clear evidence for a genetic cause of CHDs. Nonetheless, a fast amount of all CHD cases are sporadic with no family history. Recent studies suggest that most of the sporadic forms of CHD are oligogenic by nature (McBride *et al.*, 2005). It has been estimated that only about 5 % of CHDs are caused exclusively by gene mutations in *NKX2.5*, *NOTCH1*, *JAGGED1*, *GATA4* or *MHC6* (Basson *et al.*, 1999; Elliott *et al.*, 2003; Goldmuntz *et al.*, 2001; Ikeda *et al.*, 2002; Krantz *et al.*, 1999; McElhinney *et al.*, 2003; Schott *et al.*, 1998; Sperling *et al.*, 2005). This suggests that modifier genes exist which confer to the risk of CHD. One example is the *Vascular Endothelial Growth Factor* (*VEGF)* promoter. The haplotype of three single nucleotide polymorphisms (SNPs) in the 5´UTR of the *VEGF* promoter has been linked as a modifier of DiGeorge syndrome (Awata *et al.*, 2002; Lambrechts *et al.*, 2005; Shahbazi *et al.*, 2002).

In summary, the aforementioned findings demonstrate the complex etiology of CHDs and underline its multifactorial nature.

1.4 Skeletal muscle development

Skeletal muscle development in higher vertebrates is initiated in the paraxial mesoderm by formation of the somites, paired, transient segments on either side of the neural tube (Christ and Ordahl, 1995) (Figure 1.6). Somite formation is initiated by combinatorial effects of Wnt, Shh, Fgf and BMP signaling (Pownall *et al.*, 2002), which leads to expression of the paired-box proteins Pax3 and Pax7 (Buckingham, 2007). These proteins have been shown to activate MyoD and Myf5, key regulators of myogenesis (Rawls and Olson, 1997). The somites are subdivided into the ventrally located sclerotome, which will give rise to cartilage and bone as well as the dorsally located dermomyotome, which gives rise to the derm of the back, the skeletal muscle and the limbs (Tam and Trainor, 1994) .

Before cell differentiation, the muscle progenitor cells delaminate from the dermomyotome and migrate to the rostral and caudal edges to form the myotome (Denetclaw and Ordahl, 2000). Skeletal muscle cell differentiation is regulated by Myf5 and MyoD which fine-tune the balance between proliferating and differentiating cells (Kablar *et al.*, 2000). Moreover, MyoD, Mrf4 and Myogenin are implicated in the subsequent activation of muscle-specific genes during differentiation (Valdez *et al.*, 2000).

The mature muscle fiber types, such as fast and slow muscle, are determined starting around day 14 of mouse embryogenesis. It is also correlated with the innervation of the muscles (Zhang and McLennan, 1998). Slow muscle formation has been shown to be regulated by Ca^{2+}-dependent NFAT signaling as well as Mef2, which can act downstream of calcineurin signaling (Kitzmann and Fernandez, 2001). The signaling pathways regulating fast muscle are not as well characterized (Buckingham, 2001) .

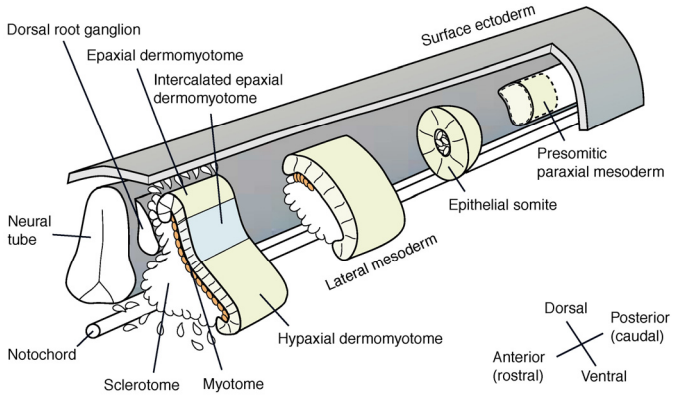

Figure 1.6: Schematic overview of mouse somitogenesis. Somites are formed and mature following a rostro-caudal gradient on either side of the notochord. Figure from (Buckingham, 2006).

Striated muscle can be identified by the presence of alternating light and dark bands, which represent the sarcomeres. The light band is called I-(isotropic) band and the dark band is called A-(anisotropic) band, due to their appearance in polarized light. Striated muscle is a complex cytoskeletal network consisting of myofibers, which in turn consist of myofibrils and the functional sarcomeric units.

In skeletal muscle, the myofiber is generated by fusion of several myoblasts, leading to the formation of multinucleated cells. Skeletal muscle contracts in response to nerve stimulation by acetylcholine, which is released by motor neurons at the neuromuscular junction. Upon stimulation, the sarcoplasmatic reticulum releases calcium ions that induce muscle contraction (Frank, 1964). Contraction is achieved according to the sliding filament model. In the relaxed state, tropomyosin blocks the head-binding sites of actin, preventing binding of myosin to the thin filaments. Binding of calcium by troponin C leads to shifting of tropomyosin, thus exposing the binding sites. Myosin binding activates the Myosin ATPase, which allows for force generation when the extensible region of myosin pulls the filaments across each other.

The group of cells activated by a motor neuron is called motor unit. These units are used to regulate the strength of the contraction.

Skeletal muscle and cardiac muscle are both striated. However, there are substantial differences between the two. Cardiac muscle is made from single cells. Nevertheless, they also contract according to the sliding filament model. In contrast to skeletal muscle, cardiac contraction in higher vertebrates is mediated by a group of pacemaker cells which reside in the sinoatrial node (SAN), located in the right atrium (Keith and Flack, 1907). These cells have the potential of automaticity, thus they are self-excitable without the need for an electrical impulse by the central nervous system. Nevertheless, the CNS can modulate the strength and speed of contraction via the autonomic nervous system through the sympathetic and parasympathetic nervous system. The cells of the SAN can be distinguished from the surrounding atrial cells by expression of unique ion channels and gap junction proteins (Boyett *et al.*, 2000). Cells of the myocardium are tightly connected by intercalated discs, a double cell membrane that includes gap junctions, which allow for the transmission of electric potential from one cell to the other through ion channels. The adherens junctions of the intercalated discs ensure the structural integrity of cell-cell contacts even when high force is applied.

1.4.1 The sarcomere

The unifying element of cardiac and skeletal muscle is the sarcomere. It consists of the myosin-containing thick filaments and the actin-containing thin filaments (Figure 1.7). The

third filament type is titin, which spans half of the sarcomere and represents the largest known human protein. Moreover, Nebulin spans the length of the actin filaments and is the fourth filament type (Wang, 1982).

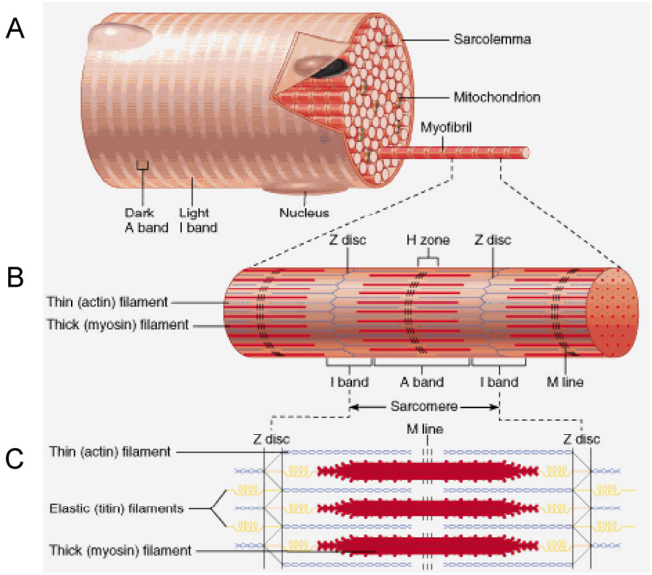

Figure 1.7: Schematic representation of striated muscle architecture. A) Cross section of myofiber reveals aligned myofibrils, mitochondria and sarcolemma. B) Single myofiber, C) Sarcomeric structures. Figure from (Cummings, 2001).

The thin filaments are attached to the z-disc and span the I-band up to the middle of the sarcomere. The actin filaments are intertwined with myosin filaments at the A-band. Moreover, nebulin, titin and the tropomyosin-troponin complexes are associated with the thin filament throughout the cell (Steinmetz *et al.*, 1997). At the ends of the thin filaments, the actin-capping proteins CapZ and Tropomodulin (Tmod) are located. CapZ localizes to the z-disc, where it also binds alpha-actinin and anchors the thin filaments (Papa *et al.*, 1999). Tmod binds to the pointed end of actin filaments and is important for thin filament length regulation, organization, and contractile activity (Gregorio *et al.*, 1995).

The thick filament myosin and its associated proteins are located at the A-band. Connections between myosin and actin are established through the globular heads of myosin. Myosins are the motor proteins of the sarcomere. They use free energy from ATP-hydrolysis to rotate the myosin head, which is attached to actin, thus leading to muscle contraction (Vale and Milligan, 2000). Several thick filament-associated proteins have been identified. Among them are Myosin Binding Proteins C and H (MyBP-C and MyBP-H), which appear to play multiple roles in muscle fiber function. For example, MyBP-C has been proposed to link the thick and thin filaments in the A-band (Winegrad, 1999). Moreover, both seem to play a role in muscle contraction (Winegrad, 2005).

The ends of titin molecules of individual sarcomeres overlap at the z-disc and the M-line, thus forming an intercellular protein complex (Maruyama, 1986). The I-band region of titin consists of an elastic region that acts as a molecular spring needed for proper muscle contraction. Due to its repetitive motif structure and interactions with sarcomeric proteins, it has been suggested to serve as a sarcomeric stabilizer (Clark *et al.*, 2002) .

The thin, titin and nebulin filaments are anchored in the z-disc, the lateral walls of the sarcomere. Neighboring z-discs are aligned and enable coordinated contraction among individual myofibrils. The connection between myofibrils is established by intermediate filament proteins such as desmin, which are located at the z-disc (Lazarides, 1980). The z-disc is the first structure to be formed during muscle development, indicating a role in organizing the sarcomere.

The identification of several z-disc associated proteins led to the insight that the z-disc consists of a complex protein network. The most prominent member is alpha-actinin, which serves as an actin crosslinking protein (Blanchard *et al.*, 1989). Many of the other z-disc associated proteins directly bind to alpha-actinin (Faulkner *et al.*, 2001). Among those is the group of filamins, which appear to play a critical role in linking the membrane and the sarcomeric cytoskeleton (Thompson *et al.*, 2000; van der Ven *et al.*, 2000).

1.4.2 Muscle disease in humans

Proteins of the sarcomere are also implicated in disease. Over 100 mutations in nine genes have been linked to hypertrophic cardiomyopathy (HCM) (Towbin and Bowles, 2002). Therefore, familial HCM has been coined as "the disease of the sarcomere" (Thierfelder *et al.*, 1994). The group of proteins mutated in dilated cardiomyopathy (DCM) is more diverse and also includes cytoskeletal proteins. Moreover, muscular dystrophy is a group of inherited disorders that result in muscle wasting and weakness accompanied by changes in muscle fiber size, muscle necrosis and increased amounts of fat and connective tissue (Emery, 2002).

The first protein identified to be causative for MD was dystrophin and shown to be localized to the sarcolemma (Hoffman *et al.*, 1987). Since then, mutations in more than 17 other genes have been identified to cause the disease. Notably, all of the gene products are localized to the sarcomere or sarcolemma.

Regarding proteins of the z-disc, mutations in *ALPHA-ACTININ* have been shown to cause DCM (Mohapatra *et al.*, 2003). Interestingly, mutations in *ACTIN* can cause both DCM as well as HCM, depending on the protein domain affected (Mogensen *et al.*, 1999; Olson *et al.*,

1998). Mutations in the dimerization domain of *FILAMIN C* have been shown to cause myofibrillar myopathy (MFM) (Vorgerd *et al.*, 2005).

1.5 Epigenetic regulation of gene expression

Epigenetics has classically been defined as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs *et al.*, 1996). Nowadays, epigenetics is more mechanistically and molecularly defined as "the sum of alterations to the chromatin template that collectively establish and propagate different patterns of gene expression and silencing from the same genome" (Allis *et al.*, 2007b). The main features of epigenetic gene regulation are DNA methylation, small non-coding RNAs and changes to the chromatin structure.

In higher eukaryotes, the combined length of all chromosome pairs is about 2 meters. Consequently, the DNA has to be compacted about 10,000-fold to fit into a cell's nucleus. The vast majority of genomic DNA in eukaryotes is wrapped around a set of basic core histone proteins, which build the nucleosomes. The core histones are histone H2A, H2B, H3 and H4, around which 147 bp of DNA are wrapped (Luger *et al.*, 1997). Histone proteins consist of a globular domain and a flexible histone tail, which projects outward from the nucleosome. These tails are subject to a large variety of covalent modifications that are correlated with transcriptional activity (see Chapter 1.5.3). The nucleosomes are connected by a short stretch of linker DNA, to which histone H1 binds and are the subunits of the higher-order chromatin structure (Figure 1.8).

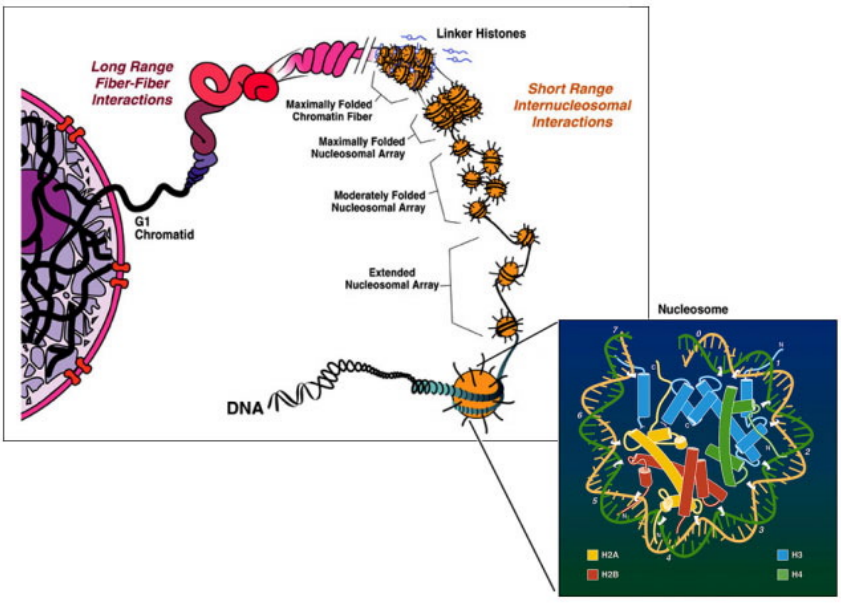

Figure 1.8: Schematic representation of Chromatin structure. Figure from (Shogren-Knaak, 2008).

The chromatin structure has been shown to be highly dynamically regulated and accounts for a large part of epigenetic gene regulation. As a consequence, structurally "open" chromatin is generally associated with active genes as these regions are easily accessible for the transcriptional machinery. On the other hand, the closed chromatin structure represents an effective barrier for these proteins. Changes to the chromatin structure can be achieved in different ways: by shuffling of histone variants, posttranslational histone modifications and ATP-dependent chromatin remodeling.

1.5.1 Chromatin remodeling

In the nucleus, DNA is wrapped around the histone octamers and is further condensed into nucleosomes, which ultimately build up the chromatin structure (Felsenfeld and Groudine, 2003; Khorasanizadeh, 2004). Regions of open chromatin structure (euchromatin) exist, in which transcription factors and the basal transcription machinery can easily access DNA. On the other hand, regions of closed chromatin (heterochromatin) are tightly compacted, excluding access to DNA. The chromatin state is highly dynamic and transition from closed to open chromatin is in part brought about by the activity of chromatin remodeling complexes. These complexes use free energy from ATP-hydrolysis to non-covalently change chromatin and nucleosome composition. Based on their characteristic domain features, four different subfamilies of remodeling factors have been recognized: the CHD, Ino80-, SWI2- and ISWI-subfamilies (Eberharter and Becker, 2004).

The CHD-subfamily is characterized by the chromodomain (Marfella and Imbalzano, 2007). The Ino80-subfamily is less well characterized, but has been implicated in the exchange of canonical histone H2A by a H2A variant (Krogan *et al.*, 2003). The ISWI family is characterized by the presence of C-terminal SANT-like domains and includes the NURF complex. These complexes mobilize and slide nucleosomes along the DNA (Tsukiyama *et al.*, 1995; Varga-Weisz *et al.*, 1997). The forth group consists of the brahma or SWI/SNF family, originally identified in yeast as "mating type switching and sucrose non-fermenting" mutants (Martens and Winston, 2003). These complexes alter the structure of nucleosomes, thus exposing DNA-histone contacts. This is achieved by peeling up about 50 b of DNA from the edge of the nucleosome and sliding the histone beyond the DNA ends. Then, the end of the looped out DNA is bound to the histone octamer, which leads to the formation of a DNA-accessible site. Formation of this stable complex is achieved at the cost of less than ten ATPs in one second (Kassabov *et al.*, 2003).

The SWI/SNF complexes in mammals are referred to as Brg1/Brm-associated factor (BAF) complexes. The complex is about 2 MDa in size and consists of a distinct ATPase and at least 10 subunits that vary depending on cell type and stage (Olave *et al.*, 2002; Wang *et al.*, 1996) (Figure 1.9). Due to their dynamic subunit composition, they have been shown to be involved in a diverse set of biological processes (Eberharter and Becker, 2004). A role for these complexes has been shown in control of gene expression, cell-cycle and oncogenesis (Kingston and Narlikar, 1999; Muchardt and Yaniv, 1999; Vignali *et al.*, 2000). The ATPase subunits are either Brahma-related gene 1 (Brg1) or Brm (derived from *Drosophila* brahma protein) and contain a bromodomain which recognizes acetylated histone lysine residues (Tamkun *et al.*, 1992). The bromodomain is a distinguishing feature of the SWI/SNF class and not found in ISWI-, CHD- or Ino80-type ATPases. Nevertheless, bromodomains are present in other subunits of these complexes.

Despite their high similarity, Brg1 and Brm appear to have distinct functions, reflected by the differences observed in gene inactivation studies in mice. Brm knock-out mice appear grossly normal and have only slight changes in control of cell proliferation while Brg1 null mice die during early embryogenesis (Bultman *et al.*, 2000; Reyes *et al.*, 1998). In addition to a distinct ATPase subunit, mammalian BAF complexes contain either BAF250 (BAF) or BAF180 (PBAF). BAF can contain either Brg1 or Brm, while PBAF always contains Brg1 (Simone, 2006). Aside from altering nucleosome structures, Brg1 can directly interact with specific transcription factors, like members of the Gata family (Kadam and Emerson, 2003). Protein-protein interaction domains are also found in BAF250, BAF180, BAF170, BAF155, BAF60 (a, b, c1/2), BAF53 (a, b), BAF47, and G-actin. Moreover, the subunits BAF250, BAF180, BAF170, BAF155 and BAF57 contain DNA-binding domains for direct Protein-DNA interactions (Mohrmann and Verrijzer, 2005).

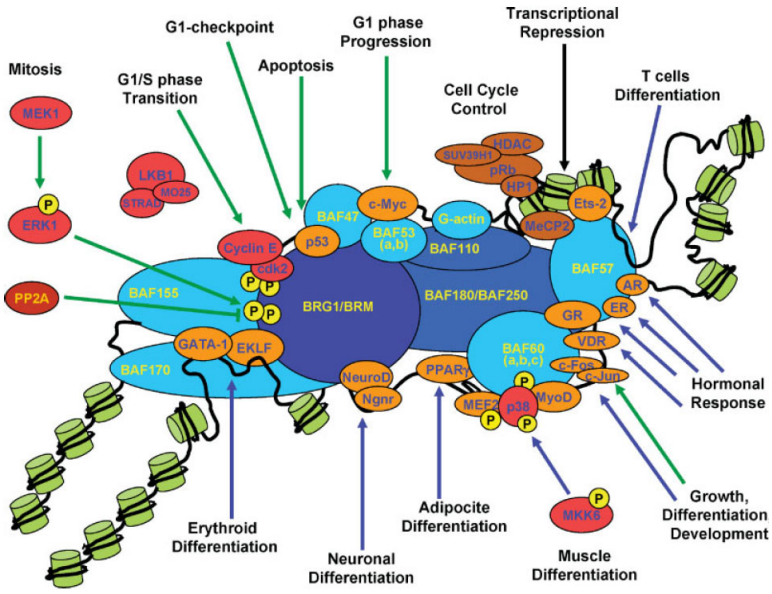

Figure 1.9: Signaling pathways converging on the mammalian BAF chromatin remodeling complex. Figure from (Simone, 2006).

1.5.2 Histone shuffling

Histone protein variants exist that can replace the canonical histones at centromeres, active or silent loci. For example, histone H3 is replaced by histone H3.3 at active loci (Hendzel and Davie, 1990). The resulting variant nucleosomes have different properties both in terms of their structure (Fan *et al.*, 2002) as well as their posttranslational modifications (Henikoff and Ahmad, 2005). Another well described example is the exchange of histone 2A (H2A) by marcroH2A. This vertebrate-specific H2A variant, which consists of an additional C-terminal globular domain, is enriched on the mammalian inactive X-chromosome. A silencing mechanism through attraction of proteins via the globular domain has been proposed (Allen *et al.*, 2003).

1.5.3 Histone modifications

Transcriptional regulation is partially achieved through covalent modifications of histone tail residues (Kouzarides, 2007). More than 70 sites of modification and seven distinct modifications have been reported (Berger, 2007). These include the addition of small moieties such as acetyl-, methyl-, or phosphate-groups on specific tail residues. So far, lysine (K), arginine (R), serine (S), threonine (T), tyrosine (Y), histidine (H) and glutamic acid (E) have been identified as sites of modifications (Taverna *et al.*, 2007). The wide range of modifications as well as their association with specific genomic loci and transcriptional activity led to the

formulation of the histone code hypothesis, which states that "distinct histone modifications, on one or more tails, act sequentially or in combination to form a 'histone code' that is read by other proteins to bring about distinct downstream events" (Strahl and Allis, 2000). In other words, the combination of the modifications at distinct loci create functional subdomains, which serve as signaling marks in the initiation of transcriptional regulation.

The main modifications are acetylation and methylation, which are added to the histone residue by a so-called "writer" (Ruthenburg *et al.*, 2007a). Hypoacetylation is generally associated with the repressive chromatin state, while hyperacetylation represents active chromatin. Methylation of histone residues is observed in both cases and depends on the individual position. The transfer of acetyl-groups is mediated by a group of enzymes called K-acetyltransferases (KATs, formerly known as HATs), which are in a dynamic interplay with lysine deacetylases (HDACs); enzymes that remove acetyl-groups, thus opposing KAT effects (Allis *et al.*, 2007a; Thiagalingam *et al.*, 2003).

The transfer of methyl-groups is mediated by a group of enzymes called K-methyltransferases (KMTs, formerly known as lysine methyltransfereases) Most KMTs contain a SET domain, which catalyzes the addition of methyl-groups to the specific lysine residue (Dillon *et al.*, 2005; Martin and Zhang, 2005). Depending on the residue, lysines can be mono-, di- or trimethylated. As in the case of acetylation, the methyl-group can be removed from the histone tail residue by K-demethylases (KDMs, formerly known as lysine demethylases) and the methylation status can be highly dynamic (Allis *et al.*, 2007a; Shi, 2007). The ability to dynamically change the modification status of chromatin allows for intricate regulation of gene expression.

The consequence of such histone modifications can be two-fold (Jenuwein and Allis, 2001; Strahl and Allis, 2000). In the "direct effect" model, a certain modification directly leads to alteration of chromatin structure. This has been shown to be the case for acetylation of histone tail residues, as the acetyl-group removes parts of the positive charge of the basic lysine residue, thus weakening its interaction with the negatively charged DNA backbone. A similar mechanism has been shown for phosphorylations (Cosgrove *et al.*, 2004; Shogren-Knaak *et al.*, 2006; Ura *et al.*, 1997; Wolffe and Hayes, 1999).

In the "effector-mediated" model, a certain histone modification is read by a so-called "effector molecule", which is either part of a larger protein complex that brings about changes to the chromatin architecture or recruits such complexes (Ruthenburg *et al.*, 2007a; Seet *et al.*, 2006). Among these complexes are the chromatin remodeling complexes, which alter chromatin structure in an ATP-dependent manner. These complexes will be discussed in a

separate chapter of the introduction (see Chapter 1.5.1). Histone tail modifications are read by a large variety of protein domains, depending on the site and nature of modification.

Histone lysine acetylation is recognized by the bromodomain (Dhalluin *et al.*, 1999; Zeng and Zhou, 2002). This domain is frequently found in chromatin-associated proteins, namely lysine acetyltransferases (KATs) such as Gcn5p, general transcription factors including TAFII250 and chromatin remodeling complexes such as the SWI/SNF complex (Jacobson *et al.*, 2000; Owen *et al.*, 2000; Shen *et al.*, 2007). The structural basis of acetyllysine binding by the bromodomain has been thoroughly analyzed. The bromodomain is around 110 amino acids and consists of four alpha-helices with inter-helical ZA and BC loops which interact with the acetyl-lysine. The interaction of the bromodomain with acetyl-lysine is mediated through highly conserved tyrosine residues. Conserved asparagines contribute to binding by formation of hydrogen bonds with the acetyl-group. Sequence specificity of the binding reaction is dependent on the sequence environment surrounding the acetylated lysine and differs among proteins (Mujtaba *et al.*, 2007).

Histone lysine methylation at particular residues is enriched in certain regions of chromatin, indicating that the marks are specialized and have distinct biological functions (Martin and Zhang, 2005). For example, tri-methylation of lysine 4 of histone 3 (H3K4me3) is enriched around the 5´region of active genes and correlates with transcription rates, active polymerase II occupancy, and histone acetylation (Bernstein *et al.*, 2005; Fischer *et al.*, 2007; Ng *et al.*, 2003; Santos-Rosa *et al.*, 2002; Schneider *et al.*, 2004; Schubeler *et al.*, 2004).

Histone lysine methylation is recognized by two groups of domains. The Royal superfamily fold and the PHD finger superfamily (Li *et al.*, 2006a; Maurer-Stroh *et al.*, 2003; Wysocka *et al.*, 2006).

The Royal superfamily includes the chromo-, tudor-, and malignant brain tumor- (MBT) domain. For example, higher methylation states such as H3K4me3 are read by the double chromodomains of CHD1 or the double tudor domain of JMJD2A (Flanagan *et al.*, 2005; Huang *et al.*, 2006). In addition, lower methyl-lysine states are read of proteins like L3MBTL1, which recognizes H4K20me2 (Li *et al.*, 2007).

The plant-homeodomain was originally described in *Arabidopsis thaliana* as a conserved stretch of amino acids with regularly spaced cysteines (Schindler *et al.*, 1993). The name plant-homeodomain had been assigned to this domain due to its reminiscence of a different class of homeodomain proteins in this species. The PHD finger superfamily is classified by the presence of an interweaved dual-zinc finger and the C4HC3 sequence (Mellor, 2006) (Figure

1.10A). It is similar to the RING finger and FYVE domain, which serve as protein ubiquitin E3 ligases and phosphoinositide binding elements, respectively (Joazeiro and Weissman, 2000; Kutateladze, 2006).

Evidence for the presence of the PHD finger in all eukaryotes analyzed was provided by Aasland *et al.* (Aasland *et al.*, 1995). It has been estimated that around 150 proteins containing a plant-homeodomain are present in the human genome. Moreover, a striking co-occurrence of nuclear protein domains was found by analyzing data from the SMART database (Ruthenburg *et al.*, 2007b). The most frequent combination of domains was observed between the PHD finger and the bromodomain (Figure 1.10B).

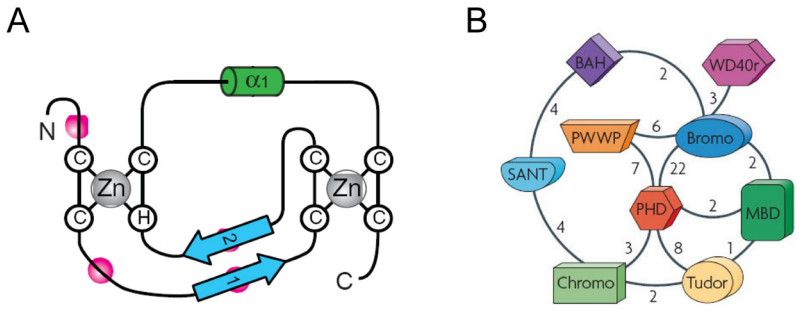

Figure 1.10: A) Topology of the PHD finger fold. Blue arrows, two small beta-strands that bridge the interweaved zinc finger motifs; labeled white circles, zinc-coordinated cysteine and histidine residues; green cylinder, short alpha-helix (alpha1) near the C-terminus; pink circles, the caging residues for readout of methyllysine or unmodified lysine marks. B) Co-occurrence of the PHD finger with other chromatin-associated domains based on SMART database entries. Numbers indicate instances of two domains found in one protein. Figure from (Ruthenburg *et al.*, 2007b).

The fact that PHD fingers are frequently found in nuclear protein complexes involved in chromatin remodeling implicated a role for the plant-homeodomain in transcriptional regulation. In 2006, four papers were published simultaneously that described the molecular function and structural basis of histone tail recognition by the PHD finger.

As for the Royal superfamily, PHD fingers have been shown to recognize methylation states of histone lysines.

BPTF (bromodomain and PHD finger transcription factor) is the largest subunit of the nucleosome remodeling factor (NURF) complex, an ISWI-containing ATP-dependent chromatin remodeling complex, and preferentially binds to H3K4me3 (Li *et al.*, 2006a; Wysocka *et al.*, 2006). NURF-mediated ATP-dependent chromatin remodeling has been shown to be directly coupled to H3K4 trimethylation in maintenance of Hox gene expression during development.

Identical histone tail binding properties have been described for ING2 (Pena *et al.*, 2006; Shi *et al.*, 2006), a subunit of a repressive mSin3a-HDAC1 histone deacetylase complex (Doyon *et al.*,

2006). Upon DNA damage, this complex is recruited to active gene loci and mediates gene repression, indicating a possible role in tumor suppression.

PHD fingers not only play a role in chromatin remodeling complexes. It has been shown recently that the basal transcription machinery, namely TFIID, is recruited to H3K4me3 via its subunit TAF3-PHD finger and that TAF3 can act as a transcriptional co-activator in a PHD finger-dependent manner (Tora, 2002; Vermeulen *et al.*, 2007).

A common structural mechanism for H3K4me3 binding has been identified that consists of the formation of a cage within the PHD finger lined by aromatic residues. The di-or tri-methylammonium group of H3K4 is positioned within this cage and stabilized by cation-π interactions (Li *et al.*, 2006a). The integration of structural data of 17 PHD fingers alone and three in complexes with H3K4me3 allowed for the identification of essential amino acids that mediate interactions. For example, there is a conserved methionine that forms the base of the aromatic cage (substituted for a tyrosine in BPTF) and an essential tryptophan that serves to ensure the correct spacing within the cage. Further aromatic residues such as tyrosines or tryptophanes contribute to the structural integrity of the PHD finger.

Nevertheless, the analysis of available protein sequences revealed that the majority of PHD fingers lack the residues essential for H3K4 methyl binding. This indicates that either alternative modes of binding exist or that the PHD finger has additional functions.

Alternative functions of a PHD finger have been described for the KAP1 co-repressor, in which a tandem PHD finger is located adjacent to a bromodomain. Here, the PHD finger serves as an intramolecular E3 ubiquitin ligase that mediates the sumoylation of lysine residues in the bromodomain (Ivanov *et al.*, 2007). Sumoylation is required for KAP1-mediated gene silencing and functions by directly recruiting the SETDB1 histone methyltransferase and the CHD3/Mi2 component of the NuRD complex via SUMO interacting motifs. Sumoylated KAP1 stimulates the histone methyltransferase activity of SETDB1. These results showed a mechanistic explanation for the cooperation of PHD and bromodomains in gene regulation. Bearing in mind that 22 known proteins share a PHD finger and a bromodomain, the aforementioned cooperative activity might represent a more general mechanism. Moreover, the PHD finger of ING2 has been shown to serve as a nuclear receptor for phosphoinositides, which mediates p53-dependent apoptotic pathways upon DNA damage (Gozani *et al.*, 2003).

Proteins of the d4 family from fruitfly to human contain a tandem repeat of the plant-homeodomain. In addition, isoforms of Dpf3 with a truncated version of the PHD1 (½-

PHD1) and no PHD2 have been identified in higher vertebrates from human to chicken. The ½-PHD1 of Dpf3a only has three of the zinc-chelating cysteines (Figure 1.11).

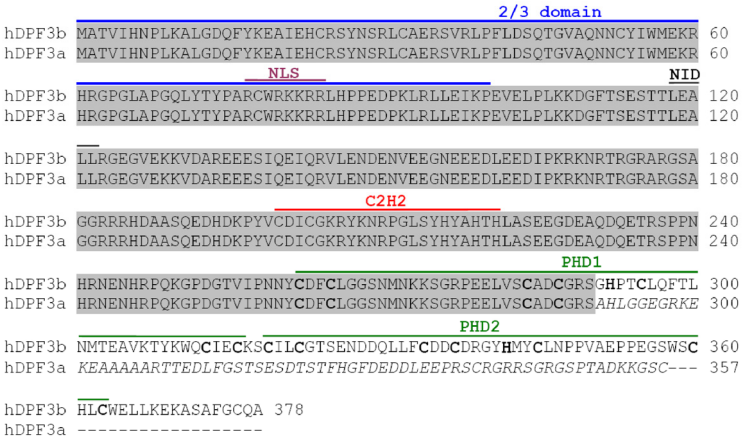

Figure 1.11: Sequence conservation and divergence of human DPF3 isoforms. DPF3a (AAX20019.1) and DPF3b (NP_036206) contain an N-terminal 2/3 domain, putative nuclear localization signal (NLS), nuclear receptor interaction domain (NID) and C2H2-Krüppel-like zinc finger. Cysteine and histidine residues of the plant-homeodomains are bold.

Two copies of a PHD finger can be found in several other proteins. For example, the proteins BPTF, CHD3/4, BRPF3, PHF1/12/17 JADE1/2, PHF19, G2E3, MYST3/4, JMJD1/2 and PCL2 have two PHD fingers positioned directly adjacent to each other (Brooks *et al.*, 2007; Coulson *et al.*, 1998; Ivanov *et al.*, 2007; Jones *et al.*, 2000; Li *et al.*, 2006b; Pelletier *et al.*, 2003; Tan *et al.*, 2007; Tzouanacou *et al.*, 2003), while JARID1, TRIM33/66 and BRD1 contain two PHD fingers separated by other domains (He *et al.*, 2006; Lee *et al.*, 2007; Severinsen *et al.*, 2006). Multiple PHD fingers have been found in the proteins MLL2/3, JARID1A, PHF14 and NSD2/3 (Iwase *et al.*, 2007; Ruault *et al.*, 2002; Wang *et al.*, 2007) .

PHD finger proteins have also been implicated in human disease. The William-Beuren syndrome transcription factor (WSTF), which is a component of the ISWI- and SWI/SNF-based chromatin remodeling complexes, has been implicated in development of the disease due to impaired recruitment of the vitamin D receptor (VDR) to vitamin D regulated promoters (Bozhenok *et al.*, 2002; Kitagawa *et al.*, 2003). Chromodomain protein 3 (CHD3) and CHD4 are dermatomyositis-specific autoantigen markers (Targoff and Reichlin, 1985) and belong to the NuRD chromatin remodeling complexes (Becker and Horz, 2002).

Also implicated in the development of autoimmune disease is the autoimmune regulator (AIRE) protein 1. AIRE1 mutations are causative for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (Villasenor *et al.*, 2005).

Finally, mutations in ATRX, a member of the SNF2 family of proteins consistently cause ATR-X syndrome in humans, characterized by mental retardation, genital abnormalities and facial dysmorphism (Gibbons *et al.*, 1995; Picketts *et al.*, 1996).

1.5.4 Epigenetic regulation of heart and skeletal muscle development

Heart and skeletal muscle development is regulated by a combination of genetic mechanisms, such as DNA-binding transcription factors and epigenetic factors. Small non-coding RNAs are increasingly being reported to play key roles regulating the activity of genes in cardiogenesis. MicroRNAs (miRNAs) regulate gene activity by inhibition of translation or mRNA degradation (Zhao and Srivastava, 2007). The miRNA *mir-1* is expressed specifically in the developing heart and skeletal muscle throughout development and adulthood. *Mir-1* has been shown to be transcriptionally regulated by SRF, Mef2 and MyoD, which are key regulators of myogenesis. Mechanistically, *mir-1* regulates the activity of Hand2, another transcription factor important for heart development (Zhao *et al.*, 2005).

Chromatin remodeling factors in heart and skeletal muscle development

The role of chromatin remodeling factors in heart and skeletal muscle development has been addressed by several studies. Using dominant negative versions of Brm1 or Brg1, it has been shown that the BAF complex is essential for initiation of MyoD-mediated muscle differentiation (de la Serna *et al.*, 2001). It has later been shown that the BAF complex is targeted to muscle-specific loci by the p38 MAPK pathway. Mechanistically, upon MKK6 activation, p38 localizes to chromatin and phosphorylates the BAF subunit Baf60, thus recruiting the complex to target loci (Simone *et al.*, 2004). In addition, the insulin growth factor 1 (IGF1)-induced Pi3K/AKT pathway promotes the association of MyoD with p300 and PCAF acetyltransferases. The two pathways ultimately converge on the promoters of muscle-specific loci to induce differentiation (Serra *et al.*, 2007).

The fact that some subunits of the BAF complex exhibit tissue-specific expression raised the question whether these factors contribute to tissue-specific gene regulation.

Baf60c is an isoform of the 60kDa BAF complex subunit and has been shown to be specifically expressed in the precardiac mesoderm, somites and midbrain (Debril *et al.*, 2004; Lickert *et al.*, 2004). RNA interference experiments in mouse and zebrafish embryos revealed an essential function for Baf60c in heart and skeletal muscle development. The heart defects were reflected by decreased expression of working myocardium markers Nppa, loss of

trabeculation markers Bmp10 and Irx3 and moreover by loss of outflow tract markers Fgf10, Pitx2 and Bmp4. Defects in the establishment of left-right asymmetry, namely looping of the heart, have also been reported (Takeuchi *et al.*, 2007). Additionally, it has been shown that Baf60c can enhance interactions between Brg1 and cardiac-specific DNA binding transcription factors Gata4, Nkx2.5 and Tbx5. The skeletal muscle defects were reflected by reduced Wnt11 expression.

Baf180 is a component of the PBAF complex and knock-out studies in mice showed that it plays a role in cardiac chamber maturation. Baf180 null embryos have very thin cardiac walls and reduced trabeculation, resembling the phenotype of retinoic acid receptor alpha (RXR*a*) knock-out mice (Sucov *et al.*, 1994; Wang *et al.*, 2004). Interestingly, Baf180 has been shown to potentiate the transcriptional activation of nuclear hormone receptors such as RXR*a*, VDR and PPAR*gamma* (Lemon *et al.*, 2001).

These results established that BAF and PBAF complexes play distinct roles in heart and skeletal muscle development.

Histone modifications in heart and skeletal muscle development

The aforementioned histone modifications have also been analyzed in the context of heart and skeletal muscle development. The role of histone acetylation has been of special interest, as it has been shown that histone lysine acetyltransferases (KATs) and histone deacetylases (HDACs) are important regulators of cardiac gene expression (Backs and Olson, 2006). In addition, HDACs have been suggested as a therapeutic target of cardiac hypertrophy (Song *et al.*, 2006; Trivedi *et al.*, 2007). Several DNA-binding transcription factors have been shown to depend on HDAC activity. For example, Bop, a SET and MYND domain protein essential for heart development, regulates transcription of the *Hand2* gene through an HDAC-dependent mechanism (Gottlieb *et al.*, 2002). Serum response factor (SRF), an important regulator of heart and skeletal muscle development (Li *et al.*, 2005; Parlakian *et al.*, 2004), is negatively regulated by interaction with HDAC4 (Davis *et al.*, 2003). More evidence for an interconnection of SRF and HDACs was provided by the finding that Homeodomain only protein (Hop) interacts with HDAC2, thus being recruited to SRF-dependent promoters (Barron *et al.*, 2005).

p300 is a transcriptional co-activator as well as an acetlytransferase (Asahara *et al.*, 2002; Chan and La Thangue, 2001) that interacts with cardiac transcription factors like Mef2d and is essential for proper ventricular trabeculation (Molinari *et al.*, 2004; Yao *et al.*, 1998). The Gata

family of transcription factors, which play an essential role in heart development, are also transcriptionally coactivated by p300 (Dai and Markham, 2001; Kakita *et al.*, 1999) .

A specific role for histone methylation is less well characterized. Nevertheless, it has been reported that knockdown of s*et and mynd domain-containing 1* (*smyd1)* in zebrafish embryos leads to defects in heart and skeletal muscle development due to disruption of myofiber maturation (Tan *et al.*, 2006). Smyd1 is a methyltransferase specifically expressed in skeletal and cardiac muscle (Hwang and Gottlieb, 1997). More evidence for a functional role of methyltransferases in muscle differentiation was provided by studies of Ezh2 (Caretti *et al.*, 2004). Ezh2 has been shown to be a repressor of cardiomyocyte differentiation. In light of this, methylation of lysine residues 9 and 27 of histone H3 might potentially be of interest, as these have been shown to exert a repressive function (Kouzarides, 2007). Moreover, as cardiomyocyte differentiation progresses, Ezh2 expression gradually decreases (Caretti *et al.*, 2004). Last, the methyltransferase Rae28 of the Polycomb-group has been shown to be important for sustained Nkx2.5 expression in the developing heart and lack of Rae28 leads to defects in heart looping morphogenesis (Shirai e*t al.*, 2002). A combinatorial analysis of the effect of histone methylation and acetlyation has been reported recently in two muscle cell lines, showing that not only the type of modification, but also their combination is essential for their regulatory potential, supporting the existence of a histone code (Fischer *et al.*, 2008).

Taken together, the role of chromatin structure and its impact on the expression of genes regulating heart and skeletal muscle development is beginning to emerge and knowledge about these mechanisms will be valuable for the application in stem cell research and disease prevention.

1.6 Purpose & Aims

Congenital heart defects are the most common birth defects in humans with an incidence of about 0.8 % among all live births. The development of the embryonic heart is tightly regulated by a gene program and perturbations to this program are, among confounding factors such as the environment, causative for congenital heart defects. Some progress has been made in the last decade to delineate to exact genetic mechanisms controlling heart development. Nevertheless, the understanding of heart formation is far from complete, hampering efforts of stem cell research to efficiently generate cardiomyocytes for heart repair.
The purpose of this study is the functional characterization of the transcription factor Dpf3, which had been identified in a screen for markers of congenital heart disease.

The molecular properties of Dpf3 will be analyzed by a combination of biochemical methods such as GST pulldown experiments. Tandem affinity purification of Dpf3 protein interaction partners combined with mass spectrometry will be performed in a collaborative effort within the workgroup (Martje Tönjes) and the mass spectrometry unit of the Max Planck Institute for Molecular Genetics. Chromatin Immunoprecipitation will be performed by Jenny Fischer and Christina Grimm within the group to identify direct transcriptional targets of Dpf3. The transcriptional regulation of the *Dpf3* gene will be characterized by *in vitro* and *in vivo* methods including reportergene assays in cell culture, gel shift assays and promoter analyses in transgenic mice. In cooperation with the group of Dr. Salim Abdelilah-Seyfried at the Max Delbrück Center for Molecualr Medicine (MDC), the role of Dpf3 during development will be addressed in zebrafish embryos using morpholino antisense oligonucleotide mediated knockdown combined with gene expression profiling to investigate underlying pathways affected.

The obtained results will allow for a functional annotation of Dpf3 during embryonic development and place it in the molecular hierarchy controlling heart development. Furthermore, the biochemical analyses of the molecular function of Dpf3 will add to the understanding of transcription factor function in general.