INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

COMBINATORIAL INTERACTIONS CONTROLLING CARDIAC TRANSCRIPTION

By Daniel Durocher

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Medicine
Division of Experimental Medicine
McGill University
Montréal (Québec)

SUMMARY

The establishment and maintenance of the cardiac phenotype requires the activation of cardiac-specific as well as muscle-restricted genes in a tightly regulated spatial manner. This process is presumably governed by the combinatorial action of cell-restricted as well as ubiquitous transcriptional regulators. In order to unravel new cardiac transcriptional regulatory pathways and to characterize how they participate in the establishment of the cardiac transcriptional program, we used the atrial natriuretic factor (ANF) gene as a marker. This strategy led to the identification of a novel cardiac *cis*-element, the NKE, which is critical for the activity of the ANF promoter and which binds members of the cardiac NK2 family. The NKE is located in proximity to a critical GATA element and the phasing between these two elements is evolutionary conserved. Thus, we hypothesized that GATA-4 would functionally interact with Nkx2-5. Indeed, this work documents, for the first time, a molecular interaction between the essential cardiac proteins GATA-4 and Nkx2-5. Furthermore, this work provides a basis for GATA specificity *in vitro* and generates novel insights on how the cardiac genetic program is activated in early cardiogenesis. Finally, we report the cloning of a novel cardiac protein likely to have an important role in cardiogenesis, which was cloned in an effort to discover novel binding proteins that bind the critical CARE element required for high promoter activity in embryonic heart and postnatal atrium. This CARE-binding protein, CATF1, is a multifunctional helicase that defines a new family of DNA/RNA helicases. CATF1 is specifically expressed in the heart in neonates and is likely to be an activator of the ANF promoter. Collectively this work significantly furthers our knowledge on the establishment of cardiac- and chamber-specific transcription in the myocardium which is proposed here to occur via a combinatorial set of interactions among cardiac-restricted and ubiquitous transcription factors.

RÉSUMÉ

L'établissement et le maintient du phénotype cardiaque requiert l'activation concertée de gènes dont l'expression est restreinte au coeur et au muscle strié. Ce processus est présumément gouverné par l'action combinée de facteurs de transcription histo-spécifiques et ubiquitaires. C'est donc dans le but d'identifier de nouvelles voies de régulation de la transcription cardiaque et afin de caractériser leur participation dans l'établissement du programme génétique cardiaque que nous avons utilisé le promoteur du facteur natriurétique des oreillettes (ANF) comme marqueur. Cette stratégie a permis l'identification d'un nouvel élément en *cis* cardiaque, le NKE, dont la présence est critique pour l'activité du promoteur de l'ANF, et qui est lié par des membres cardiaques de la familles des homéoprotéines NK2. Le NKE est situé à proximité d'un site GATA et la phase entre ces deux éléments a été conservée au cours de l'évolution. A partir de ces observations, nous avons proposé que GATA-4 interagirait de façon fonctionnelle avec Nkx2-5. Ainsi, ce travail présente pour la première fois une interaction au niveau moléculaire entre les protéines cardiaques GATA-4 et Nkx2-5. De plus, ce travail établit une base moléculaire pour la spécificité entre les facteurs GATA, in vitro, et génère de nouvelles avancées sur la compréhension de l'établissement su programme génétique cardiaque. Finalement, nous rapportons le clonage d'une nouvelle protéine cardiaque qui joue probablement un rôle important dans la cardiogénèse. Cette protéine a été clonée dans un effort afin de découvrir de nouvelles protéines cardiaques liant l'élément CARE qui est requis pour l'activité maximal du promoteur ANF dans les myocytes embryonnaires et auriculaires adultes. Cette protéine liant le CARE, appelée CATF1, est une hélicase multifonctionelle définissant une nouvelle famille d'hélicases présentes dans les archea et eukaryotes. CATF1 est exprimé spécifiquement dans le coeur du rat nouveau-né et peu activer, en certaines circonstances, le promoteur ANF. Dans l'ensemble, ce travail approfondi de façon significatives nos connaissances en ce qui concerne la définition de la spécificité transcriptionelle histo-spécifique dans le muscle cardiaque. Nous proposons, que cette spécificité est acquise via des interactions combinée entre les facteurs de transcriptions restreints au coeur et ubiquitaires.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

1. INTRODUCTION

1.1. MECHANISMS OF GENE REGULATION

One of the fundamental problems of developmental biology is to understand how the zygote can form an extraordinary complex multicellular organism. The processes involved include cell division, cell differentiation, apoptosis and morphogenetic movements. All these processes, including homeostasis, require new cellular functions and hence differential expression of genes. These requirements underline the central role of gene expression regulation in developmental biology.

The central dogma of molecular biology defines the vectorial transfer of genetic information that leads to the synthesis of an active protein. The first step is transcription, the synthesis of mRNA from DNA. The mRNA is then matured by processes such as polyadenylation, capping, splicing and transport. The mature mRNA is finally translated into a protein. This section of the introduction will focus on these processes and their regulation with a particular emphasis on the control of transcription. It is noteworthy to add that multiple post-translational events occur to modify the newly translated protein but for the sake of brevity, this topic will not be discussed here.

1.1.1. Transcription regulation

Our genetic material is composed of approximately 100 000 nuclear genes and they are classified in three broad classes. The ribosomal RNA genes are transcribed by RNAPI and are called class I genes; protein-encoding genes (class II genes) are transcribed by RNAPII whereas tRNA genes are transcribed by RNAPIII. All three classes have common and distinct mechanisms but RNAPII has been much more studied since it is the differential expression of that gene class that generates the diversity observed during development and homeostasis.

1.1.1.1. Transcription initiation

Promoter regions of class II genes contain three types of sequences that are required for either basal or activated transcription. The TATA box and the Inr element represent the

two types of elements involved in basal transcription whereas the enhancer elements comprise *cis*-elements that activate or repress transcription (Zawel and Reinberg, 1993). A typical basal promoter is composed of either a TATA box or an Inr (or both). The TATA box (consensus TATAa/tAa/t) is usually located near positions -30 to -25 relative to the transcription start site and the Inr is a pyrimidine rich sequence of the consensus YYANt/aYY located near the initiation site (Weis and Reinberg, 1992; Javahery *et al.*, 1994). These sequences serve to correctly initiate transcription by recruiting basal transcription factors and the RNAPII enzyme to the promoter (Zawel and Reinberg, 1993). However, some exceptions exist where neither a TATA nor Inr motifs are present. They usually consist of housekeeping genes and they contain specialized elements that will not be discussed here (Ince and Scotto, 1995). Intense work over the last decade led to the identification of the general transcription factors (GTFs) involved in basal transcription and resulted in the understanding of how they interact with each other to initiate transcription.

The first component of the basic transcription apparatus is the RNA polymerase itself. Most of the work on RNAPII has been done in yeast but very high conservation between yeast and higher eukaryotic RNAPII is found. The RNAPII is biochemically defined as an 11 (in yeast) or 12 (in mammals) polypeptide enzyme with subunits ranging from 220 to 10 kDa (Young, 1991). So far, the RNAPII activity is not reconstituted from purified or recombinant subunits and it is still not clear if all the subunits are required for its enzymatic activity. One of the main features of the eukaryotic RNAPs is that, unlike the prokaryotic RNAP, they are devoid of specific DNA-binding activity although it is clear that the two largest subunits, which bear significant homologies with the β and β' subunits of the bacterial RNAP, are involved in nucleotide and non-specific DNA-binding. RNAPII harbors many catalytic activities such as DNA-directed and independent RNA synthesis, endoribonuclease activities and most probably helicase activity (Young, 1991).

The largest subunit of RNAPII contains an unusual C-terminal domain (CTD) which consists of multiple repeats of the consensus YSPTSPS. These repeats are found in variable number depending on the species. These numbers range from 26 in yeast to 52 in human (Young, 1991). These repeats are essential for cell viability since truncation of the CTD is lethal in yeast, *Drosophila* and mammalian cells (Edwards *et al.*, 1991; Nonet *et al.*,

1987; Allison *et al.*, 1988). The functions of these repeats are closely associated with the state of phosphorylation of the CTD. The hyperphosphorylated form (Pol IIo) is associated with elongating polymerase whereas the hypophosphorylated RNAPII (Pol IIa) is associated with the preinitiation complex (PIC) before promoter clearance (Laybourn and Dahmus, 1990; Cadena and Dahmus, 1987). This observation led to the hypothesis that phosphorylation of the CTD serves to dissociate the RNAPII from the PIC. Recently, by a combination of genetics (Thompson *et al.*, 1993) and biochemistry (Ossipow *et al.*, 1995), it has been demonstrated that a multiprotein complex called either the mediator or the srb complex, associates with the CTD to form a 30-50 subunit RNAPII holoenzyme [reviewed in (Koleske and Young, 1995)]. The holoenzyme will be discussed in greater detail in section 1.1.1.3. Moreover, after promoter clearance, the CTD serves as a docking site for heteronuclear RNA (hnRNA) maturation. Thus, the CTD is a multifunctional protein interface that links the RNAPII to either transcription initiation, elongation or hnRNA maturation (Steinmetz, 1997).

Unlike bacterial RNAP, no σ-like proteins are found in RNAPII (nor in the two other RNAPs). However, almost every class II promoter contains the so-called TATA box or Inr element. In 1984, chromatographic fractionation of *Drosophila* nuclear extracts identified the TFIID fraction as the TATA binding factor (Parker and Topol, 1984). The purification of TBP and the molecular cloning of its gene has been extremely frustrating and it was not until it was shown that yeast TFIID could substitute for human TFIID in *in vitro* transcription experiments (Cavallini *et al.*, 1988; Buratowski *et al.*, 1988) that the first GTF could be cloned. After the cloning of yeast TBP (Horikoshi *et al.*, 1989; Hahn *et al.*, 1989), human TBP was cloned in less than two years using degenerate PCR cloning (Kao *et al.*, 1990; Hoffman *et al.*, 1990). Primary structure analyses revealed that a major portion of TBP is evolutionarily conserved at a level of 80% and a portion of this core domain has a significant homology with a portion of prokaryotic σ factors that is involved in the recognition of the -10 element in prokaryotic promoters. Intriguingly, TBP is also found in the archea phylum, indicating that mechanisms of transcriptional control in this phylum might be conserved with those of eukaryotes (Rowlands *et al.*, 1994; Baumann *et al.*, 1995; Struhl, 1994). The N-terminal domain of TBP is very divergent among species whereas

some features are conserved in higher eukaryotes. Although yeast TFIID can substitute for human TFIID in biochemical assays *in vitro*, human TBP cannot complement a strain with a deleted yeast TBP (Gill and Tjian, 1991). The differences seem to lie in subtle differences in the conserved core domain since the N-terminus is totally dispensable for yeast viability (Gill and Tjian, 1991). The crystallographic structure of TBP has been obtained and TBP is a saddle-like protein binding the TATA element in the minor groove of DNA and the TBP-bound DNA is bent by an angle of 90° (Nikolov *et al.*, 1992). This bend is proposed to have an important structural implication for the stereo-specificity of PIC assembly.

The classical view of PIC assembly places the TFIID/TBP binding to DNA at the first step. This assumption comes from band shift-PIC reconstitution assays and by the fact that TBP is the major specific DNA-binding GTF. These band shift experiments established an orderly sequence of events leading to PIC assembly (Buratowski *et al.*, 1989). These assays take advantage of the fact that TBP is the only TATA-binding factor. Thus, upon addition of other GTFs, the mobility of the TBP-DNA complex will be reduced only if the added GTF binds to TBP-DNA or the TBP-GTF-DNA complexes (Buratowski *et al.*, 1989). In this way, it was found that TFIIA, a 3 polypeptide activity, stabilizes the binding of TBP on DNA. The importance of TFIIA for basal transcription remains controversial since it is dispensable for basal *in vitro* transcription but it seems that TFIIA's major role is to stabilize the TFIID-DNA complex (Kim *et al.*, 1994). Moreover, TFIIA is found to be loosely associated with TFIID in the absence of DNA and might be constitutively associated with TFIID *in vivo* (Reinberg *et al.*, 1987).

TFIIB, however, is absolutely required for *in vitro* transcription (Reinberg and Roeder, 1987). It is found in all eukaryotes and in archea (Qureshi *et al.*, 1995). It recognizes either the D-DNA (TFIID-DNA) or the DA-DNA complex, giving rise to the DB- or DAB-DNA complexes respectively (Buratowski *et al.*, 1989). TFIIB was purified to homogeneity and cloned via reverse genetics (Yamashita *et al.*, 1992). TFIIB is a 316 amino acid protein and its key role is to bridge the DA complex to the RNAPII complexed with other GTFs (Yamashita *et al.*, 1992). Binding of TFIIB on the DA-DNA complex is polar and this property serves in positioning the polymerase near the proper start site.

RNAPII is then recruited in the PIC via TFIIF, a tetramer composed of two dimers of the RAP30 and RAP74 subunits. TFIIF bears significant homologies with bacterial σ factors and targets RNAPII to the promoter via two complementary mechanisms that involve interactions with TFIIB and a TFIIF-dependent decrease in affinity of RNAPII for non-specific DNA.

After the formation of the DAB-PolF-DNA complex, PIC reconstitution experiments established that TFIIE enters the PIC (Buratowski *et al.*, 1989). TFIIE is necessary for *in vitro* transcription and is a tetramer of two subunits of 34 and 56 kDa (Peterson *et al.*, 1991). TFIIE might possess ATPase or kinase activity and has a proposed role in promoter melting since photocrosslinking experiments have located TFIIE to a region just upstream the initiation site. The binding of TFIIE on the DAB-polF complex is stabilized by the subsequent entry of TFIIH. TFIIH plays a crucial role in promoter melting and promoter clearance (Maxon *et al.*, 1994). TFIIH was purified to homogeneity and is composed of approximately 9 subunits. It has a cyclin kinase pair (cyclin H/MO15) that possess a kinase activity that mediates CTD phosphorylation (Roy *et al.*, 1994; Lu *et al.*, 1992). The CTD kinase activity of TFIIH is greatly stimulated by its incorporation in the PIC and by TFIIE in particular [reviewed in (Svejstrup *et al.*, 1996)]. Moreover, many subunits of TFIIH were shown, by somatic genetics, to be involved in nucleotide excision repair. The role of TFIIH helicase activity in transcription remains to be addressed since, using abortive initiation assays, Tjian's group showed that TFIIH is not needed for open complex formation but rather for promoter clearance (Maxon *et al.*, 1994).

The AdML promoter is a strong promoter that contains both an Inr and a TATA box. However, an increasingly large number of promoters are TATA-less and it is not clear that PIC assembly on these promoters reflects what is seen on the AdMLP. The mouse terminal deoxynucleotidyl transferase TATA-less promoter contains an Inr that was shown to be sufficient to promote specific initiation of transcription *in vitro*. Reconstitution experiments have shown that all GTFs are required for efficient initiation including TFIID and TFIIA (Pugh and Tjian, 1991). Unlike TATA-containing promoters, recombinant TBP cannot substitute for TFIID for basal transcription, implicating a TAF in the recognition of the Inr (Pugh and Tjian, 1991; Kaufmann and Smale, 1994; Burke and Kadonaga, 1996).

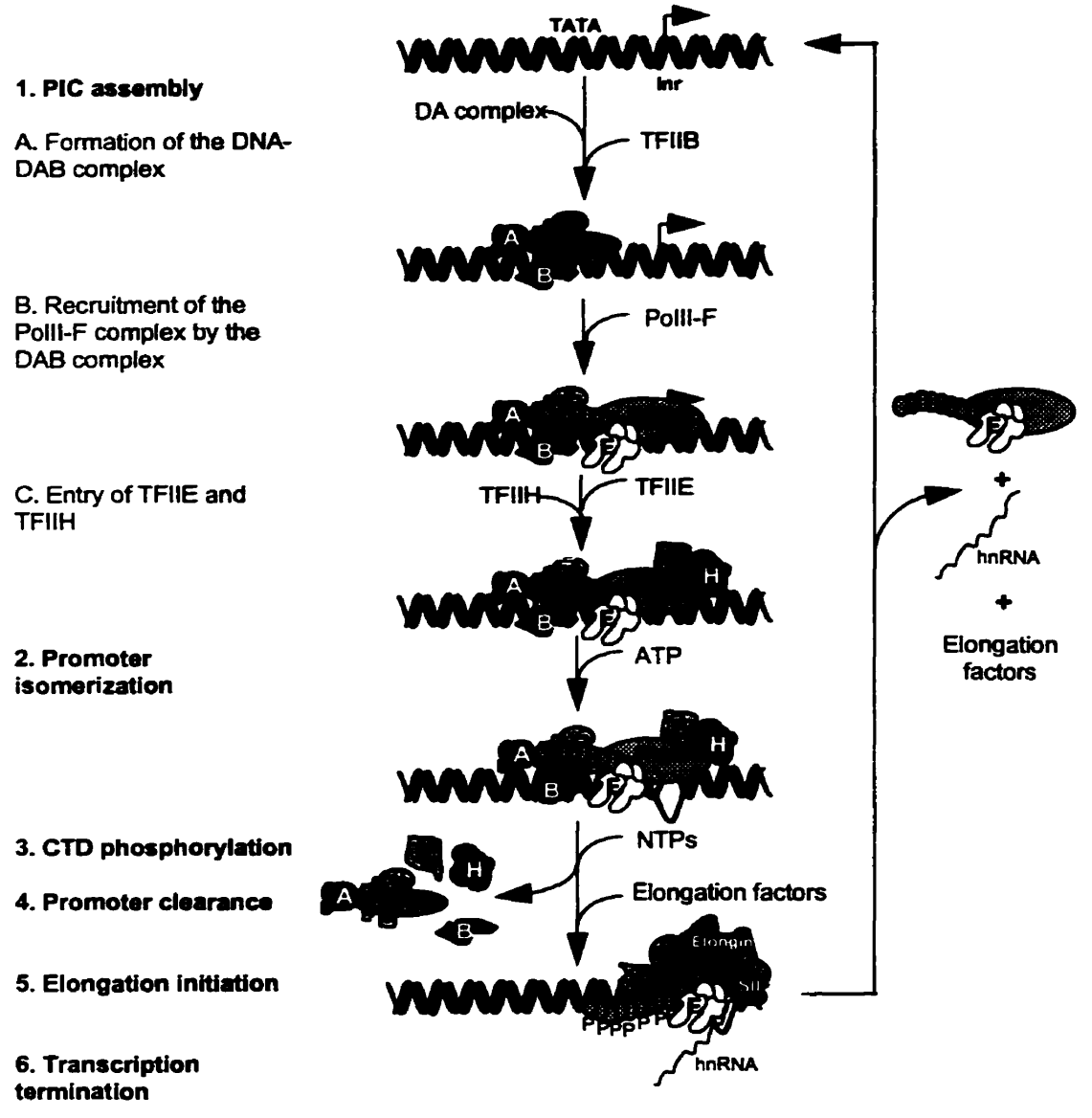


Figure 1.1.

The transcription cycle. The steps leading to initiation, elongation and termination of transcription are depicted with a particular emphasis on transcription initiation. Please note the state of phosphorylation of the CTD (the "tail" of RNAPII) throughout the cycle.

Reconstitution assays have demonstrated that a trimeric complex containing TBP, $TAF_{II}250$ and $TAF_{II}150$, was sufficient for effective initiation. In this complex, $TAF_{II}250$ is a bridge between TBP and $TAF_{II}150$ which is proposed to bind DNA and to recognize the Inr element (Verrijzer *et al.*, 1995). Thus, these data would implicate that alternative recruitment pathways to the promoter can be achieved without the binding of TBP to DNA and without the concomitant bend of DNA originally proposed to be required for TFIIB incorporation in the PIC. However, that step remains a sequence-specific DNA-binding recruitment of TFIID and hence does not affect the sequence of events leading to PIC assembly.

In the presence of nucleoside triphosphates, the assembled PIC rapidly melts DNA at the initiation site. This energy-mediated isomerization of the promoter converts the closed form of the promoter to the open form and this process is thought to be mediated by an helicase activity in PIC that is not TFIIH. The open complex is a very unstable state and, in the presence of dNTPs, transcription is rapidly initiated. The MO15 kinase of TFIIH is thought to phosphorylate at that point and leads to the dissociation of the PolBEFH complex from the DA-DNA complex (Lu *et al.*, 1992; Roy *et al.*, 1994). This process is known as promoter clearance and it is not known whether the DA-DNA complex remains attached to DNA for other rounds of initiation. Soon after promoter clearance, TFIIB, IIE and IIH are released from the elongating RNAPII (Kruger *et al.*, 1995).

1.1.1.2. Transcription elongation and termination.

In vivo, RNAP elongation rate is estimated at 1200-2000 nucleotides min^{-1}. However, elongation by a highly purified RNAPII is an inherently discontinuous process that proceeds optimally at the rate of 100-300 nucleotides min^{-1} and is frequently pausing or even arresting (Reines *et al.*, 1996). These observations underline the importance of elongation factors in the control of RNAPII processivity. Frequent cryptic pause sites are present on DNA templates (due to secondary structures) which force the polymerase to pause or arrest in the absence of elongation factors. The RNAPII reacts to these pause sites in two ways that led to the elaboration of the monotonic/discontinuous elongation model [reviewed in (Reines *et al.*, 1996)].

The classical view of transcription elongation depicts a polymerase that advances one nucleotide at a time, incorporating a ribonucleotide in the RNA molecule which is continuously extruded from the polymerase (monotonic mode). However, DNAse I footprinting on slowed polymerases revealed that the polymerase does not always elongate on a monotonic mode but rather advances many nucleotides at a time (discontinuous elongation or inchworm model) (Nudler *et al.*, 1994; Aso *et al.*, 1995a). This situation occurs when the polymerase meets a pause site and ceases to translocate even though it continues to transcribe. This situation is possible since the front end of the polymerase is physically different from the active site (Nudler *et al.*, 1994). Under these conditions, it was found that the polymerase is physically strained and resolves this situation either by escaping (provoking the leap observed by DNase I footprinting), arrest, or cleavage of the nascent RNA molecule by an SII-dependent endoribonucleotidic cleavage. This cleavage releases the strain and repositions the catalytic site of the polymerase in a non-strained conformation relative to the front end. Thus, without elongation factors, the RNAP would be very poorly processive and its overall rate of transcription could not accommodate the transcription of very large genes such as the dystrophin gene (2 Mb). Biochemical assays using a highly purified RNAP in limiting concentration of ribonucleotides were used to fractionate and purify most of the elongation factors. These factors are SII, P-TEFb, TFIIF, Elongin (SIII) and ELL.

Elongation factors enhance transcription at two levels: arrest prevention and pause suppression. The first mechanism, arrest prevention, increases the processivity of the enzyme and leads to the completion of an increased number of full-length transcripts. The general elongation factors SII and P-TEFb act at this level. SII (TFIIS) is an ubiquitously expressed monomer of 38 kDa (Aso *et al.*, 1995a). Mechanistic studies have shown that SII accelerates the overall rate of elongation but does not stimulate the catalytic activity of the RNAP. SII was found to expedite passage of the polymerase through a variety of impediments such as nucleoprotein complexes and intrinsic arrest sites (Reines *et al.*, 1992; Reines *et al.*, 1989; SivaRaman *et al.*, 1990). These arrest sites are found in many genes and are typically composed of at least two or more closely spaced stretches of A residues. As explained above in the discontinuous mode of elongation, SII binds to the RNAP and

stimulates the cleavage of nascent transcripts to release the conformational strain and enable the RNAP to resume transcription. SII is not the endoribonuclease but it rather stimulates the activity that seems to be intrinsic to the polymerase. *In vivo*, the importance of SII in transcription regulation is not so well established since inactivation of its yeast homologue (PPR2) is not lethal in normal conditions but rather confers sensitivity to the uracil analog 6-azauracil, a compound known to lower the pools of UTP and GTP (Archambault *et al.*, 1992). The mechanism by which P-TEFb also helps preventing arrest is, however, not understood. P-TEFb is a heterodimer of a 124 kDa and 43 kDa polypeptides. The p43 subunit has been shown to possess CTD kinase activity and its *modus operandi* might implicate the maintenance of the CTD in a phosphorylated state (Marshall and Price, 1995; Reines *et al.*, 1996).

Three elongation factors: Elongin (SIII), TFIIF, and ELL act in suppressing RNAP transient pausing. These proteins also bind the RNAPII but neither one, unlike SII, can promote nascent transcript cleavage. In all cases, they seem to act in keeping the 3' hydroxyl group of the nascent RNA near the catalytic site (Reines *et al.*, 1996; Aso *et al.*, 1996; Aso *et al.*, 1995a). Elongin is a heterotrimer of three subunits (A, B and C) of 110, 18 and 15 kDa respectively. Elongin A is the transcriptionally active protein of the complex and associates stably with the polymerase. Elongin B and C act as positive cofactors for Elongin A and can highly stimulate its transcriptional activation properties (Aso *et al.*, 1995b). Interestingly, recent genetic evidence has underlined the importance of the Elongin complex in cellular physiology. The von Hippel-Lindau (VHL) tumor suppressor gene product was found to compete with Elongin A for the binding of Elongin B and C. Thus, the VHL gene product is a negative regulator of transcriptional elongation by sequestering the positive cofactors of Elongin A (Duan *et al.*, 1995). In von-Hippel-Lindau disease, where the VHL gene product is not functional, apparition of neoplasia would then be explained, by an increased transcription of oncogenes which would lead to cell transformation (Krumm and Groudine, 1995). The importance of elongation in controlling cellular homeostasis is further exemplified by the ELL gene which encodes a monomeric protein of 80 kDa and which is often translocated in leukemias with the MLL gene (Shilatifard *et al.*, 1996). Although ELL and TFIIF suppress pausing, their mechanisms of

action are totally unknown.

Termination of transcription is very poorly understood in vertebrates. This is perhaps explained by the fact that mRNA polyadenylation involves a cleavage of the hnRNA with the concomitant degradation of the uncapped 3' segment of the hnRNA. This segment is then unavailable for mapping studies. However, it seems that following termination, the CTD is rapidly dephosphorylated and elongation factors are released. The RNAP is then free to be recruited again in a new pre-initiation complex.

1.1.1.3. Mechanisms of gene activation

To adapt the cell to its extracellular environment and to orchestrate cell events such as cell division, DNA-repair, apoptosis, cell differentiation, etc; the cell has to express many genes differentially. This process is controlled by the combinatorial action of both repressors and activators of gene expression. This section will focus on the current knowledge of the mechanisms by which transcriptional activators augment transcription.

Transcriptional activators bind DNA on enhancer regions of genes in a sequence-specific fashion. They are modular by nature and their activities are usually transferable to heterologous proteins. The typical transcriptional activator possesses a DNA-binding domain, an activation domain and sometimes an oligomerization domain (Ptashne, 1988). Activation is thought to target the general transcription machinery either by a direct physical contact or via a co-activator (Ptashne, 1988). Although, some activators have been shown to physically interact with GTFs *in vitro*, the precise mechanism of transactivation is far from being completely deciphered. Recently, two important models have emerged which were originally based either on *in vitro* transcriptional assays or genetics. I will describe these models in addition to complementary paradigms that arose during the past five years

The biochemical characterization of the GTFs involved in the assembly of the PIC led to the hypothesis that upstream activators were regulating the assembly steps of the PIC. It has been proposed that either TFIID or TFIIB could be limiting factors in PIC assembly (Choy *et al.*, 1993; Choy and Green, 1993). After binding to their cognate binding sites, transcriptional activators would recruit the limiting GTF(s) or trigger conformational

changes in the PIC, enabling the entry of the subsequent GTF. This model implies that the rate of PIC assembly depends on very few defined critical steps. The model was also based on the fact that many activation domains could contact either TFIIB, TFIID or other GTFs (Choy *et al.*, 1993).

However, this model was challenged by experiments carried out by several laboratories who made the observation that TBP was not able by itself to sustain transcriptional activation *in vitro* whereas TFIID could (Dynlacht *et al.*, 1991; Zhou *et al.*, 1993). Thus, TFIID reconstituted from biochemically purified TAFs is able to support transcriptional activation (Chen *et al.*, 1994). These experiments generated a variation of the previous model and stated that transcriptional activation requires holo-TFIID (TBP+TAFs) although it was still acknowledged that some activators might target TFIIB or -IIA (Manuelidis, 1990; Stargell and Struhl, 1995; Choy *et al.*, 1993; Choy and Green, 1993; Roberts *et al.*, 1993). The TAFs, in this model, would serve as co-activators that would bridge TFIID to the upstream-bound proteins (Dynlacht *et al.*, 1991). TBP is associated with at least 8 polypeptides ranging in size from about 18 to 250 kDa (Verrijzer *et al.*, 1995). TFIID is not so well deciphered but it is clear that $TAF_{II}250$ plays a nucleating role in TFIID assembly since it is one of the few TAFs that physically contact TBP (Weinzierl *et al.*, 1993; Ruppert *et al.*, 1993). Moreover, multiple TAF-TAF interactions have been described, especially with $TAF_{II}250$, resulting in a highly networked protein complex which might explain the biochemical behavior of the complex during purification schemes. Interestingly, secondary structure analyses of multiple TAFs (namely $dTAF_{II}30\alpha$, $dTAF_{II}40$ and $dTAF_{II}60$) revealed that they are structurally related to core histone proteins, H2B, H3 and H4 respectively (Xie *et al.*, 1996). This interesting observation led to the hypothesis that TAFs might also modify promoter topology, a feature that can affect transcriptional initiation and activation (Xie *et al.*, 1996). Besides being a multifaceted target for upstream activators and being a modifier of DNA topology, TFIID possesses catalytic activities. Two enzymatic activities were associated with *Drosophila* or human $TAF_{II}250$, histone acetyl-transferase (HAT) (Mizzen *et al.*, 1996) and basal transcription factor kinase activities (Dikstein *et al.*, 1996). Histone acetylation is an important step in relieving the repressive effect of histones on transcription (this topic will

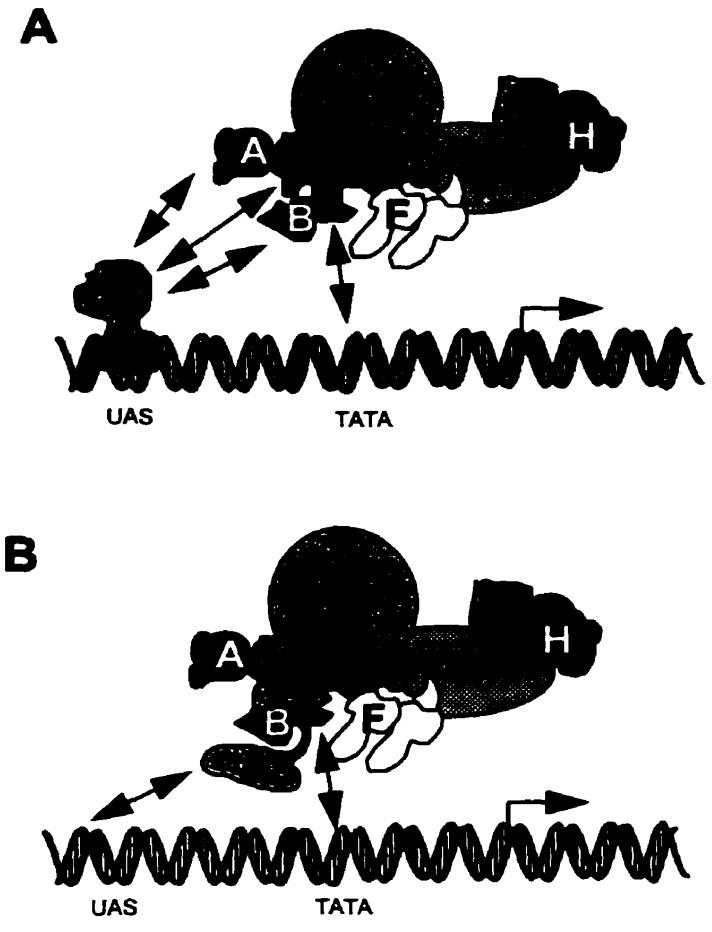


Figure 1.2.

(**A**) The holoenzyme recruitment model. The activator can either recruit a pre-assembled RNAPII holoenzyme containing all the components necessary for the accurate initiation of transcription through multiple contacts with different components of the holoenzyme or it can stabilize the binding of the holoenzymne to DNA. (**B**) Promoter bypass assay: a proof of the holoenzyme model. This assay consists of linking a DNA-binding domain to a component of the holoenzyme (in this case, TBP) and assay the transcriptional activity of the chimeric protein.

be discussed in section 1.1.2.1). In addition to HAT activity, $TAF_{II}250$ is a bipartite kinase which is able to phosphorylate itself or TFIIF (Dikstein *et al.*, 1996). TFIIF is an interesting target since it is constitutively associated with RNAPII and is involved in recruiting the RNAPII to the PIC. However, the role(s) of these activities in transcriptional activation remains to be addressed mainly because these activities are not present in yeast.

Genetic experiments in yeast were done to address the *in vivo* role of TAFs in transcription. Two approaches were used to inactivate TAFs, a temperature-sensitive mutant method (Walker *et al.*, 1996) and a double-shutoff method (Moqtaderi *et al.*, 1996), were used to create conditional mutants of TAFs since most of the TAFs are essential for cell viability. In both reports, the expression of the yeast homologue of $TAF_{II}250$ ($yTAF_{II}145$) was abrogated. Soon after the expression of $yTAF_{II}145$ was shut off, a rapid degradation of the remaining TAFs was observed probably because of TFIID disassembly (Walker *et al.*, 1996). According to the *in vitro* data, in these conditions, activated transcription was expected to be blocked. On the contrary, both basal and activated transcription of numerous genes was normal. Moreover, as controls, TBP, TFIIB or the large subunit of RNAPII knock-downs greatly affected transcription and were used to validate the experimental schemes (Walker *et al.*, 1996; Moqtaderi *et al.*, 1996). This obvious discordance with the *in vitro* paradigm suggested either that TAFs are not always necessary for activated transcription and might be necessary for only a subset of genes [for example, see (Sauer *et al.*, 1996)]. An alternative to these results is that TAFs are functionally redundant with other parts of the transcription complex machinery that were not present in *in vitro* studies (Moqtaderi *et al.*, 1996).

The latter explanation for the somewhat surprising results obtained by the knock-down of TAFs is further substantiated by the cloning of the mediator complex, another multi-protein complex associated with the basal machinery that is necessary for activated transcription *in vitro* and *in vivo*. The discovery of that complex by elegant *in vivo* and biochemical studies established a more complete model of transcriptional activation called the *holoenzyme recruitment model*.

The first evidences of an activity that is distinct from RNAPII and GTFs required for transcriptional activation came soon after the discovery that TFIID could not support

transcriptional activation alone. This observation was made when a potent transactivator (GAL4-VP16) was added to a transcription reaction which contained the core RNAPII, all the GTFs and an upstream-bound transactivator (in this case, the pyrimidine-rich binding factor). The outcome of this experiment was that GAL4-VP16, which had no cognate binding site on the template, reduced the extent of transactivation by the pyrimidine-rich binding factor. This result was also observed *in vivo* and this phenomenon was called « squelching » (Kelleher, 3d *et al.*, 1992; Flanagan *et al.*, 1991). The simplest interpretation of the squelching assay was there was a competition for a limiting factor in the reaction between the activation domain of GAL4-VP16 and the activation domain of the upstream-bound activator. Since the squelching effect could be reversed by a non-GTF or core RNAPII chromatographic fraction, this fraction was called the mediator of transcriptional activation (Flanagan *et al.*, 1991).

In parallel, the group of Richard Young took an alternative approach that led to the identification of the same multi-protein complex. It was shown in yeast and mammals that the CTD of RNAPII is necessary for transcriptional activation *in vitro* and *in vivo* (Gerber *et al.*, 1995). Moreover, the number of repeats of the CTD can dictate the extent of transcriptional activation observed (Gerber *et al.*, 1995). Using this property of the CTD, Gerber *et al.* (1995) designed a genetic screen in which they used a RNAPII that possessed a truncated CTD (with 11 heptapeptide repeats instead of 27) that is active enough to permit viability. However, the yeast strain containing the truncated CTD had severe growth defects and was cold-sensitive (Thompson *et al.*, 1993). The screen recovered dominant suppressors of the cold-sensitive phenotype, called srb for suppressor of RNA polymerase B truncations (Thompson *et al.*, 1993). Deletion of the entire coding region of SRB genes leads to either cell death (SRB4 and SRB6) or severe defects in cell growth (SRB2, SRB5) (Thompson *et al.*, 1993). Biochemical analyses revealed that SRB genes are part of a multi-protein complex of approximately 1.2 MDa containing the RNAPII and TBP (Thompson *et al.*, 1993). This complex, thereafter named the holoenzyme, is sufficient for both basal and activated transcription. The group of Kornberg, using antibodies against the SRBs, then confirmed that the mediator is the SRB complex and also found that several GTFs were part of the holoenzyme (TFIIF, TFIIH, TFIIB) (Kim *et al.*, 1994). There is still

controversy about the presence of TFIIE and TFIID in the holoenzyme but the work of Ossipow *et al.*, which reported the isolation of the human holoenzyme, support the notion that it contains TBP and TAFs, although at a lower stoichiometry. This is perhaps due to a lower affinity of TFIID for the holoenzyme and might explain why, in some purification schemes, TFIID was not present in the holoenzyme. Nevertheless, this important body of work dramatically changes our view on whether transcriptional activation acts at limiting steps of PIC or by recruiting a multisubunit complex, the RNAPII holoenzyme. The recruitment hypothesis also unifies the mechanism of eukaryotic transcriptional activation with prokaryotic activation since bacterial transcriptional activators act by recruiting the bacterial holoenzyme which contains both the σ subunit and the polymerase (Ptashne and Gann, 1997). The identification of the holoenzyme led to the direct testing of the recruitment model. Since recruitment involves a protein-protein interaction that is proposed to bridge the activation domain to a target in the holoenzyme, the most direct test is to look whether a DNA-binding domain physically linked to a component of the holoenzyme, can bypass the use of an activation domain. When the DNA-binding domains of either GAL4 or LexA were fused to almost all the components of the holoenzyme, potent transcriptional activation was observed (Ptashne and Gann, 1997; Klein and Struhl, 1994; Barberis *et al.*, 1995). Moreover, in a different study, using a mutant of GAL4 that has a cryptic high-affinity interaction domain with GAL11, a component of the mediator complex, it was shown that the strength of transcriptional activation is directly proportional to the strength of the interaction seen *in vitro*, another strong argument for the recruitment model (Wu *et al.*, 1996).

Furthermore, these results provide a molecular basis for synergy (where activation of two activators is superior to the sum of the individual activations). Multiple activators can make multiple contact points on the holoenzyme resulting in cooperative interaction and thus, enhanced recruitment. Although many issues regarding the holoenzyme remain unresolved so far (e.g. what is the proportion of RNAPII that is in the form of holoenzyme?) these major breakthroughs enable us to unify the mechanism of transcriptional activation. Although the model of the holoenzyme recruitment is the current paradigm, transcriptional activation in eukaryotes is modulated by other mechanisms that

act at other levels of transcription (elongation or chromatin remodeling).

The recruitment model does not look at the subsequent steps of transcription (complex formation, promoter clearance, transcription elongation and transcription termination). However, it is probable that upstream-bound activators affect these steps as well. Indeed, in prokaryotes, activators of the NTRC family function in increasing the catalysis of the open promoter complex (Geiduschek, 1992). In mammalian cells, it has already been shown that some viral activators (HIV Tat e.g.) were also affecting transcription by stimulating the rate of elongation. David Bentley's group found that DNA-bound cellular activators could also stimulate, in some cases, the rate of elongation (Yankulov *et al.*, 1994). These results imply that synergism is not only a question of cooperative recruitment of the holoenzyme, but can also be a consequence of the increased efficiency of multiple steps in the transcription cycle. (Blau *et al.*, 1996; Ptashne and Gann, 1997).

The classical definition of the vertebrate enhancer denotes an independence to both orientation and distance relative to start site, especially when multimerized. However, this definition does not account for another property of enhancers in their natural context which is the specificity of response to extracellular stimuli (Tjian and Maniatis, 1994). In fact, many enhancers defy the classical definition. In the cases of the mouse T cell receptor α (TCRα) gene enhancer or the virus-inducible enhancer of the interferon β (IFNβ) gene, it was found that none of the individual *cis*-elements composing the enhancer can recapitulate the biological tissue-specificity of the enhancers [reviewed in (Tjian and Maniatis, 1994; Thanos, 1996)]. Furthermore, the IFNβ enhancer requires an architectural component as exemplified by the necessity of HMG-I(Y) binding to the promoter for its proper activity (Thanos and Maniatis, 1995). HMGI(Y) does not possess activation domains (nor a HMG-box) and is rather an architectural DNA-binding protein that binds AT-rich sequences in the minor groove of the double helix and which induces a bend in DNA. In the case of the TCRα enhancer, an architectural requirement was found to be encoded in the major determinant for its tissue-specific activity of the enhancer, the LEF-1/TCF1α protein. LEF1 possesses an HMG-box which induces a sharp bend to DNA via its binding to the minor groove. Thus, the activity of some enhancers require a precise array and phasing of

cis-elements which results in a stereo-specific enhancer complex reminiscent of site-specific recombination [reviewed in (Grosschedl, 1995)]. An enhancer has both an architectural and an active component, and the final output is dependent on the combinatorial interactions within the complex. As previously mentioned, this mechanism has important implications for promoters that need to achieve a high level of specificity both during development and in response to specific stimuli.

The importance of promoter architecture underlines the importance of chromatin structure in gene regulation. It was recently shown that transcriptional activators could recruit to the promoter co-activators of the CBP/p300 family (Kwok *et al.*, 1994; Bannister and Kouzarides, 1995; Kamei *et al.*, 1996). These co-activators can act as another interface for protein-protein recruitment of the holoenzyme since they seem to be part of the holoenzyme (Nakajima *et al.*, 1997), but, perhaps more significantly, these co-activators mediate transcriptional activation via HAT activity which is both intrinsic to them or extrinsic (i.e. recruited) to these molecules. Core histone acetylation is thought to mediate nucleosome mobility, an important feature of active chromatin that will be discussed in section 1.1.2.1.

1.1.1.4. Transcriptional repression

The first proof of genetic control of gene expression, demonstrated by the seminal work of Monod and Jacob, underlined the importance of repressors in gene regulation. It is only much later that the wide action of activators was incorporated to the gene expression paradigm in bacteria. However, in eukaryotes, repression was long thought to play a marginal role (Johnson, 1995). This notion is now obsolete and we now appreciate that repression plays as important a role as activation in gene regulation. Repression involves the binding or the recruitment to the promoter of a repressor, a molecule that possess a structure similar to transcriptional activators but which harbor a repressive domain instead of an activating region (Herschbach and Johnson, 1993). These molecules produce a transient and reversible repressive effect on transcription. Furthermore, persistent repression leads to major chromatin condensation that excludes transcription factors from chromatin, a process called silencing or heterochromatinization. In prokaryotes, where

repressors have been extensively characterized, repressors act basically at all steps of transcription initiation: e.g. the λ repressor blocks the RNAP from binding the promoter; MerR prevents open complex formation and the Gal repressor blocks the synthesis of the first phosphodiester bond (Herschbach and Johnson, 1993). Eukaryotic repressors have been less studied than their activating counterparts, but, nevertheless, a classification of repressors has emerged and comprises three different and non-exclusive modes of repression: interference with a DNA-bound activator; active repression; and recruitment of chromatin modifiers [reviewed in (Johnson, 1995; Hanna-Rose and Hansen, 1996; Pazin and Kadonaga, 1997)].

The interference with a transcriptional activators can be of three types: misclocalization of the activator, competitive binding of its cognate DNA element, or quenching/masking of its activation domain (Herschbach and Johnson, 1993). Mislocalization of an activator by a repressor can be either a cellular mislocalization or inhibition of DNA-binding. The classical case of subcellular mislocalization is cytoplasmic sequestration. For example, the NF-κB transcription factor is associated with the inhibitor protein IκB which sequesters it in the cytosol (Baeuerle and Baltimore, 1988). Upon the activation of the appropriate signal transduction cascade, the IκB protein is phosphorylated and rapidly degraded (Ghosh and Baltimore, 1990). NFκB is then free to translocate in the nucleus and activate transcription. The inhibition of DNA-binding is often observed in the case where proteins unable to bind DNA as monomers, oligomerizes with a protein that does not possess a DNA-binding domain. This leads to the nucleoplasmic sequestration of the protein. This case is illustrated dimerization of Id with E2A. The Id protein possess an HLH domain but cannot bind DNA whereas E2A can only bind DNA in a dimeric form. The binding of Id to E2A results in an exclusion of E2A from DNA, as well as an exclusion of class B bHLH proteins such as MyoD from DNA (Benezra *et al.*, 1990).

Inhibition by competitive binding is by far the most simple mechanism of repression but seems to be the less common in higher eukaryotes. Repressors acting by competitive binding excludes an activator from its binding site. This mechanism has been elegantly described in *Drosophila* embryogenesis where the boundaries *of even-skipped* (*eve*) stripe 2 expression is controlled by two repressors, *giant* (*gt*) and *Krüppel* (*Kr*). *gt* controls the

anterior boundary of *eve* expression whereas *Kr* controls the posterior boundary (Small *et al.*, 1991; Stanojevic *et al.*, 1991; Frasch and Levine, 1987). They both have overlapping binding sites with nearly all the activators (*bicoid* (*bcd*) and *hunchback* (*hb*)) binding the *eve* stripe 2 enhancer. Moreover, *in vitro*, it was demonstrated that both *Kr* and *bcd* cannot occupy at the same time closely spaced binding sites (Small *et al.*, 1991). Therefore, for promoters regulated by competitive binding of repressors and activators, the final output depends on the affinity of the DNA-binding proteins for their cognate sites and their concentration. This mode of regulation can lead to complex expression patterns with simple proteins gradients of activators and repressors. It is perhaps why this mechanism is used in establishing the *eve* stripes in *Drosophila* embryogenesis. In vertebrates, enhancers are far more complex and a mode of action solely based on competitive DNA-binding would require a myriad of different repressors that compete with the wide array of DNA-bound activators. That might explain why this mechanism is less observed in vertebrates.

Interference with an activator can also be executed at another level, after the entry of the activator to the nucleus and its binding to the enhancer region. A repressor, bound or recruited at the promoter, can quench the transcriptional activation domain of the activator by either masking it or masking its transcriptional target. The paradigms for this situation are the GAL80 repressor in yeast which binds and masks the GAL4 activating region in the presence of glucose and, in mammalian cells, such interference is often observed between two activators that cancel each other's activation properties (Lohr *et al.*, 1995). For instance, the glucocorticoid receptor quenches the activity of the enhancer-bound AP-1 complexes on the collagenase promoter (Jonat *et al.*, 1990). Quenching requires activated transcription which is why it is often called short-range repression. In the case of the AP-1/GR interference, it has recently been shown that they share a common co-activator (CBP/p300) and GR might act by squelching out the AP-1 co-activators (Kamei *et al.*, 1996). The presence of a quenching mechanism enables to generate an precise transcriptional output especially where the promoter has to integrate a variety of extracellular stimuli, like in the case of the AP-1/GR interference.

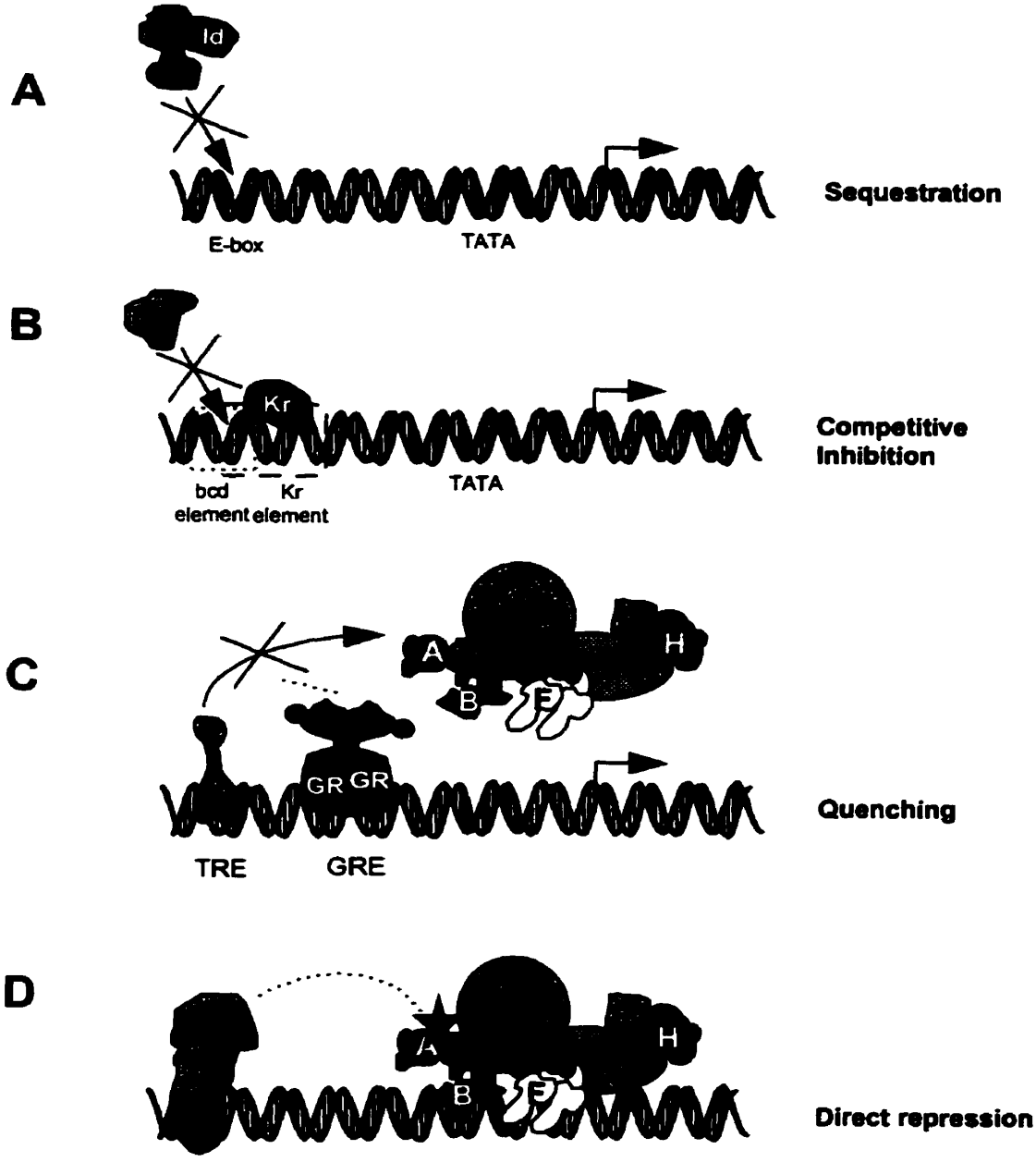
The second theme in repression is active repression (Hanna-Rose and Hansen, 1996). Active repression is defined by the ability of a repressor to interfere with basal

transcription. The mechanisms of active repression are still ill-defined but are starting to emerge. The MSX1 protein, the SSN6/TUP1 and the unliganded thyroid hormone receptor (TR) are paradigms for active repressors. *In vitro* studies have suggested that MSX1, an homeobox-containing repressor, interacts physically with TBP and TFIIB (Zhang *et al.*, 1996; Catron *et al.*, 1995). In the stepwise assembly model of the PIC, this result was suggesting that MSX1 blocks the assembly of the PIC but, in the light of the holoenzyme model, the MSX1 interaction with either TFIID or TFIIB could lead to a disassembly of the holoenzyme. Interestingly, this speculation is supported by the fact that MSX1 can repress transcription even when not bound to DNA. The disassembly of the holoenzyme would kinetically lead to a lower number of transcription cycles and thus to a decrease in basal transcription. Interestingly, the yeast SSN6/TUP1 repressor acts in another fashion to repress basal transcription. In a genetic study to find genes that are required for its repressive activity, it was found that the *SRB10* and *SRB11* genes were required for maximal levels of SSN6/TUP1 repression (Kuchin *et al.*, 1995). The gene products of SRB10 and SRB11 form a cyclin-kinase pair that is found in the mediator complex of the holoenzyme (Liao *et al.*, 1995). It is not clear yet if they are only a contact point for the repressor or if they actually modulate the activity of the kinase. If this is the case, it would be easy to imagine a mechanism in which the kinase activity of SRB10 would phosphorylate components of the RNAPII in order to reduce or inhibit steps of initiation.

Recently, an explosion of papers described how the unliganded nuclear receptors, the yeast Ume6 repressor or the Mad/Mxi heterodimer can actively repress transcription (Kadosh and Struhl, 1997; Zhang *et al.*, 1997; Laherty *et al.*, 1997; Nagy *et al.*, 1997). Although, the basics of chromatin structure and their role in transcription will be discussed in section 1.1.2.1, it was found that all these repressors were recruiting, via their repression domains, a multiprotein complex that contains Sin3A and the HDAC1 protein (Pazin and Kadonaga, 1997). All these repressors recruit the protein Sin3A directly (in the cases of Ume6 and Mad/Mxi (Kadosh and Struhl, 1997; Laherty *et al.*, 1997)) or indirectly (via SMRT and N-CoR, for the unliganded nuclear receptors (Pazin and Kadonaga, 1997)). Sin3A has already been found to be a co-repressor both in yeast and mammals. In this

Figure 1.3.

Mechanisms of repression. Examples of the four major modes of repression described in the text and based on (Herschbach and Johnson, 1993). In (**A**) the Id protein binds to class A bHLH proteins such as E2A and impairs their ability to bind DNA. (**B**) The bicoid (bcd) activator cannot bind its target site on the stripe 2 enhancer when the Kruppel repressor binds an overlapping DNA-binding sequence. (**C**) The ability of the AP-1 complex [Jun (J)-Fos (F)] to activate transcription is impaired by transcriptional interference from the glucocorticoid receptor bound to a nearby site on the collagenase promoter. (**D**) In the case of direct repression such as in MSX1-dependent repression, the effect is targeted directly at the PIC and includes either a holoenzyme disassembly or induction a conformational modification in the PIC.


A
Id
E-box
TATA
Sequestration
B
Kr
bcd element
Kr element
TATA
Competitive Inhibition
C
A
H
B
E
GR GR
TRE
GRE
Quenching
D
A
H
B
E
Direct repression

complex, the HDAC1 protein , a core histone deacetylase, was also found. Acetylation of histones is usually associated with active transcriptional regions and recruitment of the Sin3A/HDAC1 complex would locally affect chromatin structure, rendering them more compact and less accessible to both sequence-specific transcriptional activators and the holoenzyme. However, it does not seem that core histone deacetylation is the only mode of repression by Sin3A since deletion of the HDAC interaction domain of Sin3A still leads to transcriptional repression, and N-CoR contains two repression domains, one associated with Sin3A/HDAC1 interaction and another one for which the target is still unknown.

Altogether, this body of work indicates that repression is common in eukaryotes and that its mode of operation is diverse and acts at multiple levels. Again, although not reported yet, it would not be surprising to find "repressive synergism", in which two direct repressors having two different targets in the basal transcriptional machinery would have a cooperative repressive effect when both bound to DNA. For example, the unliganded TR would synergize negatively with MSX1 to shut off the activity of a promoter harboring both TRE an MSX binding sites. Interestingly, repressors, just like activators, do not seem to only have one mechanism of action: *Kr* can act as a short-range-repressor, via its interaction with the co-repressor KAP1, and can compete out some activators for DNA-binding; and Sin3A acts via deacetylase-dependent and independent mechanisms.

1.1.2. Epigenetic gene regulation.

The latter section on repression exemplifies the intricate interaction between transcription and chromatin structure, a new paradigm that definitely has been established during the past few years. It is increasingly difficult to separate gene expression and chromatin structure and this section will attempt to look at the implication of chromatin structure and DNA methylation in gene expression regulation.

1.1.2.1 Chromatin dynamics.

The basic unit of nuclear DNA organization in eukaryotes is the nucleosome. A typical nucleosome consists of 160-210 bp of DNA wrapped approximately twice around the histone octamer (one octamer contains two H2A-H2B dimers and two H3-H4 dimers)

(Croston and Kadonaga, 1993; Adams and Workman, 1993). The nucleosome represents only the first level of packaging and it is puzzling how nuclear processes can occur without grossly altering the nucleosomal structure (Adams and Workman, 1993). Moreover, it is now widely recognized that chromatin structure hinders the accessibility of transcription factors (sequence-specific or general) to DNA (Svaren and Horz, 1993). Interestingly, it was found that chromatin structure in transcriptionally active regions was dynamic and it was suggested that chromatin structure could be regulated. Hence, chromatin regulators would be *de facto* regulators of gene expression. A combination of powerful genetics and biochemistry is now starting to unravel these processes.

The active chromatin state is characterized by multiple modifications of the nucleosomal structure that leads to an increased mobility of the nucleosome and accessibility of the enhancer DNA for transcription factors as tested *in vivo* by nuclease hypersensitivity. Nuclease sensitivity is dynamic and usually correlates with the level of transcription at a particular locus. This increased sensitivity to nucleases is not generated by the loss of the core histones but rather reflects a loss of histone H1, an increased lysine acetylation of core histones and an increased lability of the H2A-H2B dimers (van Holde *et al.*, 1992). On the other hand, heterochromatinization is characterized by deacetylation of core histones and a reduced accessibility of DNA. Genetic studies in yeast and *Drosophila* laid the foundations for the role of chromatin in the regulation of gene expression [reviewed in (Moehrle and Paro, 1994; Rivier and Pillus, 1994)].

In yeast, two studies on the transcriptional regulation of either the HO endonuclease which is involved in mating type switching or the SUC2 gene, which is involved in metabolism of non-fermentable sugars have found mutations in the same class of genes involved in chromatin remodeling. These screens revealed either the SWI genes (Peterson and Herskowitz, 1992) (involved in controlling the transcription of the HO nuclease) or the SNF genes (involved in transcriptional control of the SUC2 gene) (Laurent *et al.*, 1991). The *swi1*, *swi2/snf2*, *swi3*, *snf5* and *snf6* strains have severe growth problems and have similar pleiotropic transcription defects for a number of diversely regulated genes including SUC2, HO endonuclease, GAL1, GAL10, INO1 (Carlson and Laurent, 1994). The SWI/SNF genes are in the same epistatic group, implying that they are part of the same

multiprotein complex (Peterson and Herskowitz, 1992). The first evidences that these genes were involved in chromatin reconfiguration came by genetic approaches. The first one, in an extragenic suppressor screen, two genes SIN1 and SIN2 were found to partially suppress the *swi/snf* phenotype (Hirschhorn *et al.*, 1992). Both SIN genes were found to be structural components of the chromatin as the SIN2 genes was found to encode histone H3 and SIN1, an homologue of HMG-1. Moreover, diminution of the dosage of the H2A-H2B dimer, partially suppressed the *swi/snf* phenotype (Hirschhorn *et al.*, 1992; Winston and Carlson, 1992). The partial suppression was perfectly correlated with the restoration of nuclease sensitivity at the SUC2 promoter, a strong indication that SWI/SNF genes are indeed involved in chromatin reconfiguration (Hirschhorn *et al.*, 1992). The last evidence came from *Drosophila* genetics in a screen to find dominant mutations that suppressed the *Polycomb(Pc)*-induced derepression of the HOM cluster (Tamkun *et al.*, 1992). Pc genes were already known to be involved in the formation of silenced chromatin in *Drosophila* (Kennison, 1995). This screen identified the Brahma (Brm) gene as a suppressor of the *Pc* phenotype. Molecular cloning of Brm revealed that its primary structure harbors significant homology with the SWI2/SNF2 gene (hereafter named SWI2), a helicase-like protein (Tamkun *et al.*, 1992). Studies in yeast have proposed that SWI2 was an energy-driven nucleosome reconfiguration motor since it is the only SWI/SNF gene so far that contains an ATP-binding site (Laurent *et al.*, 1993). Both SWI2 and Brm defines a new family of helicase-like proteins involved in nuclear processes. This family comprises members of all phyla which are involved in multiple cellular processes such as recombination (RAD54, ERCC6), basal transcription (MOT1) or chromatin reconfiguration (SWI2, STH1, BRM, ISWI) [reviewed in (Carlson and Laurent, 1994; Peterson and Tamkun, 1995; Kingston *et al.*, 1996)].

Biochemical analyses confirmed that the SWI/SNF proteins are part of a large ~1MDa multiprotein complex of approximately 11 polypeptides that is able to use energy to remodel nucleosomal DNA templates (Peterson *et al.*, 1994). The mechanism by which the SWI/SNF complex remodels chromatin has been proposed from *in vitro* studies to be an increase of lability of the H2A-H2B (Cote *et al.*, 1994). This activity is ATP-dependent and mutations in the ATPase motif of SWI2 inactivates this nucleosome remodeling

activity. Interestingly, parallel studies were carried out in yeast and *Drosophila* to purify ATP-dependent nucleosome remodeling complexes and two new distinct complexes were purified to near homogeneity, RSC (remodel the structure of chromatin) (Cairns *et al.*, 1996) and NURF (nucleosome remodeling factor) (Tsukiyama and Wu, 1995; Tsukiyama *et al.*, 1995). Molecular cloning of some components of both complexes revealed that they both contain a subset of homologues of SWI/SNF proteins and both contains a SWI2-like protein (ISWI for NURF, STH1 for RSC) reiterating the central role of these helicase-like proteins in chromatin remodeling. These complexes have distinct genetic roles so far and will most probably implicated in different nuclear processes. However, they have all three been found to facilitate DNA-binding of transcription factors to DNA. Recently, a provocative paper that has not received confirmation yet observed the SWI/SNF complex as being part of the holoenzyme (Wilson *et al.*, 1996). This data suggest that the SWI/SNF complex would be recruited to the promoter concomitantly of the recruitment of the holoenzyme and this model would elegantly answer the question of specificity of SWI/SNF action and expand the holoenzyme recruitment model to chromatin remodeling (Struhl, 1996).

As stated above, active chromatin is characterized by labile H2A-H2B dimers and a deficit of histone H1. This characteristic can be mechanically explained by the action of chromatin remodeling complexes. However, post-translational modification of histones, namely histone acetylation, is also an efficient marker of the state of chromatin (van Holde *et al.*, 1992). Acetylation of core histones is thought to trigger a conformational change in the nucleosome that increase its mobility. Alternative models have proposed that acetylation acts as a signal transduction signal, similar to phosphorylation but the precise role of histone acetylation remains to be further investigated. Acetylation is a dynamic process but a direct link between acetyl transferase and transcriptional regulation was not unraveled until it was demonstrated that the co-activators Gcn5p (Kuo *et al.*, 1996), in yeast, or CBP/p300 (Bannister and Kouzarides, 1996) (with or without P/CAF) and $TAF_{II}250$ (Mizzen *et al.*, 1996) in *Drosophila* and vertebrates, possessed histone acetylase activities. Thus, transcriptional activation domains are able to recruit another type of chromatin modifiers, the histone acetyltransferases to the promoter. Interestingly, HAT are

also networked with other targets of transcriptional activation via their association with multiprotein complexes. Thus, activators act by recruiting a holoenzyme that contains initiation factors and chromatin modifiers in order to establish an active chromatin state at the enhancer. These activities are in a fine balance with activities that are recruited by repressors which, in the long term can establish a chromatin structure that is so tightly packed that it results in an exclusion of DNA-binding proteins.

The maintenance of the active chromatin state is intuitively thought to be the action of competition between the nucleosome and the enhancer-bound activator and by the reiterated recruitment of the holoenzyme and chromatin remodeling activities. However, initiation of the repressed state of the chromatin is as less understood as the establishment of the long-term, inactive, chromatin state, heterochromatin. Studies on the silencing of the mating-type loci and telomeres in yeast as well as studies on developmental genes in *Drosophila* have established the molecular players that are involved in these phenomena. It is also noteworthy to add the recent results mentioned in section 1.1.1.4. in which transcriptional repressors also partially act by recruiting a histone deacetylase activities which, in light of the last paragraphs, decreases the mobility of nucleosomes and thus hinders transcriptional activation.

It is important to remember that the « basal » nucleosomal structure of chromatin is itself an transcription inhibitory activity. This was substantiated by studies in yeast which demonstrated that mutations in the histone genes or the architectural HMG proteins were able to derepress a number of genes (Clark-Adams *et al.*, 1988). Genetic studies in yeast also established that the basal repressive state of chromatin was further amplified by an increased condensation of chromatin (Barberis *et al.*, 1995). This process is thought to be triggered by RAP1, a telomere-binding protein and the origin recognition complex (ORC) (Loo and Rine, 1995). These two proteins bind specialized *cis*-elements called the silencers and recruit the SIR complex (silent information regulator) which is composed of the SIR2, SIR3, SIR4 proteins. This repressive complex is physically bound (via interactions with SIR3) to the N-termini of the core histones H3 and H4 establishing a « polymer-like », condensed, chromatin (Hecht *et al.*, 1995). However, although it seems that silencing at the telomeres and mating type loci are independent of transcriptional repressors,

heterochromatinization of active chromatin by persistent repression is not understood yet. In *Drosophila*, the Polycomb group of genes (Pc-G) were identified as repressors of the HOM (homeotic) cluster (Kennison, 1995). Genetic studies have implicated the Pc-G proteins in maintaining the repressive state instead of initiating it and have established a link between the repressor *Kr* and the initiation of repression in the *Abdominal-A* HOM gene (Shimell *et al.*, 1994). Biochemical studies on Pc-G proteins have established that, like the SIR complex, they form a multiprotein complex that is intimately associated with chromatin. Thus, the Pc-G proteins, like the SIR complex, is proposed to establish a long-term and heritable repression by forming a condensed chromatin state (Kennison, 1995). The importance of Pc-G genes has underscored the importance of chromatin regulation in *Drosophila* development. In mammals, Pc-like genes have been found to have major roles in embryonic and tumorigneic development a widely recognized role for them in silencing is starting to emerge (Schumacher and Magnuson, 1997).

1.1.2.2. Chromatin domains.

Chromatin structure influences the expression of genes at the level of the nucleosome but also at a higher level. It as been observed in the 1970's that the chromatin of histone-depleted metaphase nuclei have chromosomal loops of approximately 100 kb attached to the nuclear matrix via a region called the matrix attachment region (MAR) (Paulson and Laemmli, 1977; Mirkovitch *et al.*, 1984). It was found that these loops were enriched in active chromatin and it was then proposed that these structures might regulate gene expression by modifying chromatin topology or by insulating a transcription domain (Gasser and Laemmli, 1986). Experiments in *Drosophila* on MAR equivalents, the scs and scs' elements (specialized chromatin structure) rapidly confirmed that MARs were acting as insulators by creating an « autonomous » domain with defined boundaries (Cal and Levine, 1995). Moreover, MARs, when placed on each side of a transgene, can confer position independent and copy number-dependent expression of a transgene (Stief *et al.*, 1989). Usually, transgenes have an extremely variable expression that depends on their site of integration in the genome. The action of MARs would be to insulate the transgene from action of the surrounding chromatin. and this effect is seen either on heterochromatin or

euchromatin (Bode *et al.*, 1995). MARs were found to be nuclear matrix-bound DNA sequences that are characterized by the presence of AT-rich sequences (Mirkovitch *et al.*, 1984). These anchor points to the nuclear matrix are important to create the chromatin loops observed in histone-depleted metaphase chromosomes and to induce negative supercoiling, a topological modification that is known to positively regulate transcription. Interestingly, topoisomerase II is one of the major component of the nuclear matrix near MARs (Gasser *et al.*, 1986; Adachi *et al.*, 1989). These chromosomal domains are thought to create a scaffolding for nuclear processes such as replication, mRNA splicing and transcription. Interestingly, the positive effects of MARs on transgene expression requires the presence of a functional enhancer within the chromosomal domain (Forrester *et al.*, 1986). This requirement suggests that the establishment of an active chromatin domain is bipartite: it requires the creation of a physical domain (MAR-dependent) combined with the activation of chromatin within the chromosomal loop (enhancer-dependent). The creation of the active chromatin is probably accomplished via acetylases and SWI-like complexes that are recruited by the activation domains of enhancer-bound transcription factors.

MARs are not tissue-specific. However, the β-globin gene possess a composite element called the LCR (locus control region) which harbors characteristics of both a MAR and an enhancer. This composite element can confer tissue-specific, position-independent and copy number-dependent expression of a transgene (Grosveld *et al.*, 1987). The β-globin LCR was first identified in Dutch thalassemia patients having an heterozygous deletion which removes 100 kb upstream of the β-globin gene. This deletion left the β-globin gene intact, including all its known promoter elements (Orkin, 1990). This region upstream region is composed of four super-hypersensitive sites that are erythroid-specific. Sequence analyses of this region revealed multiple binding sites for ubiquitous and erythroid-specific transcription factor as well as functional MARs (Wood, 1996; Martin *et al.*, 1996). Therefore, it seems that LCRs work by creating a tissue-specific active chromosomal domain. Thus, higher order chromatin architecture, integrates the multiple levels of regulation of transcription to create a holo-complex that behaves in a specific spatial and temporal fashion.

1.1.2.3. DNA methylation.

Long-term and heritable repression is conferred by condensation of the chromatin structure. Methylation is also an epigenetic modification that affects negatively gene expression in a long-term and heritable fashion. In vertebrates, DNA methylation occurs on the cytosines of CpG dinucleotides to generate the 5-methylcytosine (m5C). m5C residues represent about 1% of the nucleotide bases in the vertebrate genome and around 60% of CpG dinucleotides are methylated (Bird, 1993). Two non-exclusive models are elaborated to explain the methylation-mediated transcriptional repression. First, CpG methylation dramatically decreases the affinity of some transcription factors for their recognition site (if they contain CpG dinucleotides) (Tate and Bird, 1993; Eden and Cedar, 1994). It was shown that the methylated CpG sequence of cAMP-response element (CRE) was inhibiting the binding of the CREB and thus prevent cAMP induction of a reporter gene (Iguchi-Ariga and Schaffner, 1989). However, in the case of the β-globin promoter, low-density of methylation is sufficient to repress its activity. This methylation does not have to be on specific residues, and this argues that abolition of transcription factor binding to DNA is not the sole mechanism by which methylation repress transcription (Murray and Grosveld, 1987; Boyes and Bird, 1992). The discovery of proteins (MeCP1 and MeCP2) that can specifically bind methylated CpG dinucleotides, suggested that methylation is a signal that triggers repression rather than being an intrinsic repressor (Meehan *et al.*, 1989). *In vitro* and *in vivo* transcription studies have demonstrated that MeCP1 was in fact a transcriptional repressor which binds m5C residues in a variety of contexts (Boyes and Bird, 1991). Methylation of cytosines is established in embryogenesis and is heritable by clonal transmission through the strong preference of the mammalian DNA (cytosine-5)-methyltransferase (MTase) for hemimethylated DNA (Bestor and Verdine, 1994). 98% of vertebrate DNA is methylated at a low density (1 m5C per 50-100 bp) whereas the remaining 2% are clustered in CpG islands that are located in the 5' region of about 56% of human genes. Since most CpG islands are methylation free, even were the genes are silenced, it was proposed that methylation is used to silence potential cryptic promoters in intergenic regions (Bird, 1993). These cryptic promoters could potentially affect the

expression of upstream and downstream genes. Moreover, methylation would also have the same role as in some invertebrates, which is to inactivate potentially dangerous exogenous sequences such as retroviral DNA and retrotransposons. Whether methylation is a process as dynamic as nucleosome acetylation is still unclear but the observation that some anti-oncogenes are hypermethylated in cancer cells reinforce the proposition that it is and it will be exciting to see how transcriptional pathways interact with the methylation apparatus.

1.1.3. Post-transcriptional gene regulation.

Regulation of gene expression is not solely dependent on gene transcription. Co- and post-transcriptional events enable the immature mRNA to be processed and transported in the cytosol where it will be translated. The modification of the mRNA is now known to be almost entirely co-transcriptional and involves 5' capping of the mRNA, removal of the introns by the spliceosome, cleavage of its 3' end followed by polyadenylation. Like transcription, all these steps (except capping perhaps) are regulated and can have profound effects on the final function of the protein. Moreover, specific regulatory points in these processes have an important impact during embryogenesis and post-natal life.

1.1.3.1. mRNA maturation.

Immediately after the catalysis of the first phosphodiester bond, the 5' end of the first nucleotide is capped by the nucleophilic attack of a molecule of GTP by the first nucleotide. This attack generates a 5'-5' triphosphate linkage and rapid methylation of the N-7 position of the guanine rapidly occurs. The capping of the nascent mRNA is required for efficient translation but also protects the pre-mRNA from exonuclease and phosphatase degradation (Lewis *et al.*, 1995). Capping is rapidly followed by pre-mRNA splicing which is defined by the excision of the intron sequences and ligation of the exons to form the mature mRNA. Nuclear splicing is performed in an orderly fashion which results in a stepwise excision of introns (although the sequence of intron excision seems to be empirical) and is carried out co-transcriptionally, in close association with the CTD of RNAPIIo (Mattaj, 1994; Steinmetz, 1997). Sequence analysis of 5' (donor) and 3' (acceptor) splice sites reveals consensus sequences that are required for their recognition.

The 5' splice site is characterized by the sequence AG↓GUAAGU (where ↓ represents the cut) whereas the 3' splice site is characterized by a pyrimidine tract followed by NCAG↓N (Stephens and Schneider, 1992). All splice sites are equivalent and, for example, a 5' splice site of the β-globin RNA can function with the 3' splice site of the SV40 T antigen transcript. These observations led to the question of how the splicing apparatus recognizes splice sites and why a 5' splice site is almost always fused to the adjacent acceptor site. *In vitro* splicing reactions were set up to study the mechanisms of splicing. Results from these studies established that splicing is initiated by the cleavage of the phosphodiester bond between the upstream exon and the 5' of the intron. This reaction is mediated by the nucleophilic attack of the 2' OH of the adenylate residue located within the intron at a sequence called the branch site (Padgett *et al.*, 1984; Ruskin *et al.*, 1984). The branch site is defined, in vertebrates, by the consensus YNYRAY. This nucleophilic attack results in a 2',5'-phosphodiester bond formed between the A nucleotide and the 5' terminus of the intron, and result in an intermediate structure called the lariat. The 3'-OH of the exon 1 can attack the phosphodiester bond between the 3' boundary of the intron and the first nucleotide of exon 2, resulting in the ligation of exons and the release of the lariat form of the intron.

Splicing is catalyzed in a molecular factory, the spliceosome, that can be isolated as a 50-60S sedimenting particle (Brody and Abelson, 1985). The spliceososme is a ribonucleoprotein (RNP) complex. Five ribonucleoproteins in addition to up to 40 proteins compose the spliceosome. Five small nuclear RNAs (snRNA) are present in the spliceosome: U1, U2, U4, U5, U6. Their primary sequence diverges between species but their secondary structure is rigorously conserved. They are all associated with proteins that are either associated with all the snRNPs (the eight Sm proteins) or specific to a particular snRNP [reviewed in (Kramer, 1996; Newman, 1994)]. Four snRNPs are associated with the spliceosome: U1 RNP, U2 RNP, U4/U6 RNP and U5 RNP. Both U4 and U6 snRNAs are part of the same RNP and they are associated to each other via extensive base pairing (Kramer, 1996). Biochemical studies and genetic studies in yeast have revealed that the snRNAs are required for splicing and they do not only act as a scaffold for proteins. Interestingly, the U1 snRNP possesses a sequence complementary to the consensus 5'

splice site (Zhuang and Weiner, 1986) and the U2 snRNA has also been described to bind the branch-site consensus via base-pairing (Pikielny *et al.*, 1986). However, in the case of the 3' splice site, it is not recognized by an RNA component of the spliceosome but rather by the pyrimidine tract-binding protein U2AF. U2AF is composed of two polypeptides of 35 and 65 kDa and it is $U2AF^{65}$ that recognizes pyrimidine tracts (Zamore *et al.*, 1992).

While 5' splice sites are well conserved, 3' splice sites have a poor sequence conservation. This poor conservation has an important role in regulating which acceptor site will be used (Chabot, 1996). The modification of the normal recognition of the 3' splice site can lead to exon skipping and hence, alternative splicing. Alternative splicing plays an important role in development and homeostasis since it can generate a variety of biologically diverse proteins from the same gene. For example, in the case of the CREM gene, alternative splicing can generate a transcriptional repressor by skipping an exon encoding an glutamine-rich activation domain (Laoide *et al.*, 1993). The definition of the 3' acceptor site is determined by the action of SR proteins which were first identified as essential proteins for the splicing reaction. SR proteins are characterized by a domain rich in serines (S) and arginines (R) but also by an RNA binding domain (Manley and Tacke, 1996). Their role in alternative splicing was suggested from biochemical studies where the addition of the SR protein ASF/SF2 favored the proximal splice site in a template containing two acceptor sites (Ge and Manley, 1990). Furthermore, in *Drosophila*, where sex is determined by a cascade of alternative splicing events, elegant genetic studies have determined that two SR-like proteins, transformer and transformer-2, favor a female-specific 3' splice site on the *doublesex* pre-mRNA (MacDougall *et al.*, 1995; Rio, 1993). This alternative splicing event depends on the binding of tra and tra-2 on a six 13 bp repeat present in the female-specific 3' exon (Hedley and Maniatis, 1991). Thus, alternative splicing depends on *cis*-elements that govern splice-site selection and seems to be the result of an interplay between spliceosomal proteins that favor either the proximal or the distal splice acceptors.

The first event leading to a mature 3' mRNA end is the recognition of the polyadenylation signal (AAUAAA) by the cleavage and polyadenylation specificity factor (CPSF) (Wahle and Keller, 1996). This factor is composed four polypeptides of 160, 100,

73 and 30 kDa and it is the largest subunit that possess the RNA binding activity (Bienroth *et al.*, 1991). The binding of CPSF is rather unstable and it is stabilized by the binding of the cleavage stimulatory factor (CStF), a multiprotein complex of 77, 64 and 50 kDa (Gilmartin and Nevins, 1989). After binding to RNA, the CSPF/CStF ternary complex recruits the poly(A) polymerase which is essential for cleavage. The nature of the endonuclease is still unknown but immediately after cleavage, the poly(A) polymerase synthesizes a stretch of approximately 200 adenosine residues (Wahle and Keller, 1996). The roles of polyadenylation are essentially found in transcription termination, mRNA transport, mRNA stability and stimulation of translation. When the mRNA is mature, it is ready to be translocated in the cytosol. This process does not rely only on passive diffusion and shuttling complexes were found to bind nuclear mRNAs. Thus, the nuclear post-transcriptional mechanisms are highly coordinated and regulated complexes that have an important impact on many cellular and developmental processes.

1.1.3.1. Translation control.

Translation is the process by which the ribosome reads the genetic code transported by the messenger RNA to translate it into a protein, the functional unit of most biological processes. Translation regulation is very important in development especially in the early embryogenesis where transcription is inhibited and where most of the zygotic protein synthesis depends on maternal mRNAs. Moreover, the discovery of a translational control of cell growth and transformation underscores the importance of this regulatory checkpoint.

Translation initiation requires the Met-tRNA$_f$ charged on the 40S ribosomes. This reaction requires the action of eIF2 bound to GTP which forms a ternary complex (the 43S preinitiation complex) formed of eIF2-GTP, Met-tRNA$_i$ and the 40S subunit (Hershey, 1991). The eIF2 protein (a trimeric protein containing an α,β,γ subunit) is released from the ternary complex prior to, or just after the assembly of the entry of the 60S subunit and leads to hydrolysis of GTP in GDP. eIF2 possess a much higher affinity (100 fold) for GDP than GTP and requires a GDP-exchange factor (GEF) which is eIF2B (Sonenberg, 1993). Such an exchange is inhibited when the α subunit of eIF2 is phosphorylated under stress conditions (heat shock, serum starvation or viral infection) and leads to a deadlocked

eIF2-GDP/eIF2B complex (Hershey, 1989). This phosphorylation can be accomplished by the double-stranded RNA protein kinase (PKR or p68 kinase) or by the heme control inhibitor (HCI) (Meurs *et al.*, 1990; Samuel, 1993). PKR is activated by dsRNA and HCI is activated in a variety of stress conditions. Thus, activation of PKR by a viral infection leads to an inhibition of GDP to GTP exchange and therefore to inhibition of cell growth by an inhibition of translation (Sonenberg, 1993).

The 43S preinitiation complex is targeted to the mRNA by the cap-bound eIF4F complex (Sachs *et al.*, 1997). eIF4F is composed primarily of eIF4E (the cap-binding protein) bound to eIF4G (or p220). eIF4F is bridged to the 43S subunit via its binding to eIF3 which also prevents the formation of the 80S ribosome in absence of a template. Then, the recruited 43S subunit with the additional action of RNA helicases such as eIF4A, can scan for the first AUG and form the 48S preinitiation complex (Pause *et al.*, 1994b). This sequence of events can be regulated in a number of ways by phosphorylation. An example of this is the insulin-mediated phosphorylation cascade in adipocytes, which leads to the phosphorylation of eIF4E. This phosphorylation enhances eIF4E cap-binding abilities and increases its affinity for eIF4G, resulting in a net increase in translation (Pause *et al.*, 1994a). On the other hand, eIF4E activity is counteracted by a family of proteins, the 4E-BPs (Pause *et al.*, 1994a). In their unphosphorylated form, the 4E-BPs can act as competitive inhibitors for the binding of eIF4E to eIF4G. Interestingly, 4E-BP1 was found to be identical as PHAS-I, one of the earliest substrate of the insulin-activated kinase cascades (Lin *et al.*, 1994). Thus, insulin enhances protein synthesis by decreasing the affinity of 4E-BP1 for eIF4E. After the formation of the 48S pre-initiation complex, the 60S subunit is recruited and forms the 80S initiation complex. Initiation elongation is triggered by elongation factors EF1α and EF1$\beta\gamma$ which are the eukaryotic counterparts of EF-Tu and EF-Ts (Hershey, 1991). The GTP-bound form of EF1α delivers the aminoacyl-tRNA in the A site of the ribosome. The peptidyl-transferase activity of the ribosome catalyses the transfer of the tRNA-bound amino acid (in the P site) to the A site, by forming a peptidic bond. The GTP-driven translocation of the newly formed aminoacyl-tRNA back to the P site is mediated by EF2 and termination is carried out by a single releasing factor, eRF (Hershey, 1991).

An interesting way to control translation is to control the stability of mRNA. Destabilization sequences are known to be localized in the 3' UTR region of many genes such as c-fos (Decker and Parker, 1995; Wilson and Treisman, 1988). These AU-rich sequences promote rapid deadenylation which leads to degradation (Shyu *et al.*, 1991). Interestingly, upon the addition of the appropriate signal, the mRNA is stabilized which leads its rapid translation, without the need of nuclear transcription. Another mRNA degradation translation control pathway is the nonsense-mediated mRNA decay (Jacobson and Peltz, 1996). This pathway is activated by the presence of a premature stop codon in mRNA sequences. These premature stop codons are encountered either when mutations (nonsense or frameshift) are present on the transcript or when mRNAs harbor a control element called upstream ORFs (uORFs). Both pathways trigger the activation of a translation control pathway that is mediated by the Upf1p/Upf2p/Upf3p complex in yeast (Ruiz-Echevarria *et al.*, 1996). Upf1 is a prototype for a new family of RNA/DNA helicases that also harbor a nucleic acid binding domain (Cui *et al.*, 1995; Weng *et al.*, 1996b). The Upf complex is localized in the cytoplasm and triggers the degradation of the aberrant transcript by an unknown transduction cascade (Czaplinski *et al.*, 1995). Although the role of this latter pathway is still undefined in development, it is likely to play an important role since uORFs are potent repressors of translation and a suppression of Upf1p action could lead to a rapid upregulation of uORF-containing transcripts.

Thus, post-transcriptional regulation plays an important role in regulating gene expression. The interplay of transcriptional and post-transcriptional regulatory events provides a precise spatio-temporal expression pattern for each gene and ensures that their end-products do not bear mutations, in order to provide a functional molecule.

1.2 CARDIOGENESIS.

Cardiac malformations account for more than 25% of congenital defects and lead to decrease in life expectancy and quality of life. For these reasons, the understanding of cardiac embryology is a medical issue of paramount importance. Cardiogenesis is a complex embryonic developmental program which requires the interaction of the precardiac mesoderm with all its surrounding tissues. These complex interactions give rise to the first

functional organ of the embryo. Moreover, the primitive heart has to reorganize itself through a complex series of morphogenetic movements and transformations to generate in the mature four-chambered heart. These movements cannot interfere with heart function since the heart is already pumping blood in the embryo. Thus, the heart is an interesting model organ to study embryogenesis both for its importance in physiology but also for its complex patterns of interaction with its surrounding tissues.

1.2.1. Early embryogenesis

1.2.1.1. Pre-implantation development

Fertilization of the oocyte by the spermatocyte triggers the embryonic developmental program. In mammals, fertilization occurs in the oviduct and, following sperm entry, a rapid membrane depolarization of the zona pellucida (the egg membrane) is triggered to prevent polyspermy. The male pronuclear envelope rapidly vesiculates and break down to expose the haploid genome to the egg cytoplasm. A rapid chromatin decondensation occurs and the two haploid genomes unite to form the diploid zygote (Gilbert, 1994). Fertilization is followed by cleavage, a process in which rapid mitotic divisions occur whereby the large volume of the egg cytoplasm is divided into numerous, smaller cells called the blastomeres. Mammalian mitotic divisions are particularly slow compared to other animals and these cell divisions are asymmetric. Moreover, unlike other animals where maternal transcripts are sufficient for cleavage up to the 16- or 32-cell stages, mammalian cleavage requires RNA transcription starting at the 2-cell stage (Gilbert, 1994). At the 8-cell stage, the mammalian embryo undergoes a dramatic morphological change called compaction. At this stage, blastomeres are loosely attached together and can be readily distinguished under a microscope. However, at this point, they maximize their contact points and form a ball of cells where the individual blastomeres are hardly distinguishable. The outside cells of the « ball » stabilize the compacted structure by forming tight junctions whereas cells located inside, form gap junctions which enable the passage of small molecules and ions. Compaction seems to require a major adhesion molecule, E-cadherin, since incubation of 8-cell stage embryo with antibodies against E-

cadherin totally blocks this event (Peyrieras *et al.*, 1983).

After compaction, cells of the embryo divide to form the 16-cell morula. The morula is composed of 1-2 internal cells surrounded by the remaining 14-15 cells (Barlow *et al.*, 1972). Most of the external cells will give rise to trophoblasts (or trophoectoderm). These cells do not produce embryonic structures as such but will rather form extra-embryonic structures such as the chorion and the embryonic portion of the placenta (Pedersen *et al.*, 1986; Peyrieras *et al.*, 1983). The trophoblasts secrete a fluid, the blastocoel, by a process called cavitation, into the morula to create an internal cavity. The inner cells of the morula generate the inner cell mass (ICM) which will give rise to both embryonic and extra-embryonic structures. The organization of the ICM cells and trophoblasts generates a structure called the blastocyst, and the specification of the morula cells into trophoblast and ICM cells constitutes the first definitive differentiation event in the embryo (Pedersen *et al.*, 1986; Peyrieras *et al.*, 1983). At this point, the embryo has migrated up to the uterus and must hatch i.e. exit from the zona pellucida in order to bind the uterine wall. Thus, by 5 dpc in mice, implantation of the embryo in the uterus is achieved (Gilbert, 1994).

1.2.1.2. Gastrulation

It is during gastrulation that the primordial body plan is established. This extremely important process generates the three germ layers: the outer ectoderm, the inner endoderm and the interstitial mesoderm through highly integrated cell movements. As a result, cells are brought into new positions, enabling them to create new cell to cell interactions. Although extremely different between species, the morphological changes occurring during gastrulation result in the combination of a few fundamental movements which include epiboly (movement of epithelial sheets to enclose the deeper layers of the embryo), invagination (the infolding of a region of cells), ingression (migration of individual cells into the interior of the embryo) and delamination (the splitting of a cellular sheet into parallel sheets) (Gilbert, 1994). For conciseness, this section will only look at gastrulation in mammals. However, it is noteworthy to stress that, in mammals, the cellular movements during gastrulation are very similar to those of birds which have been extensively studied.

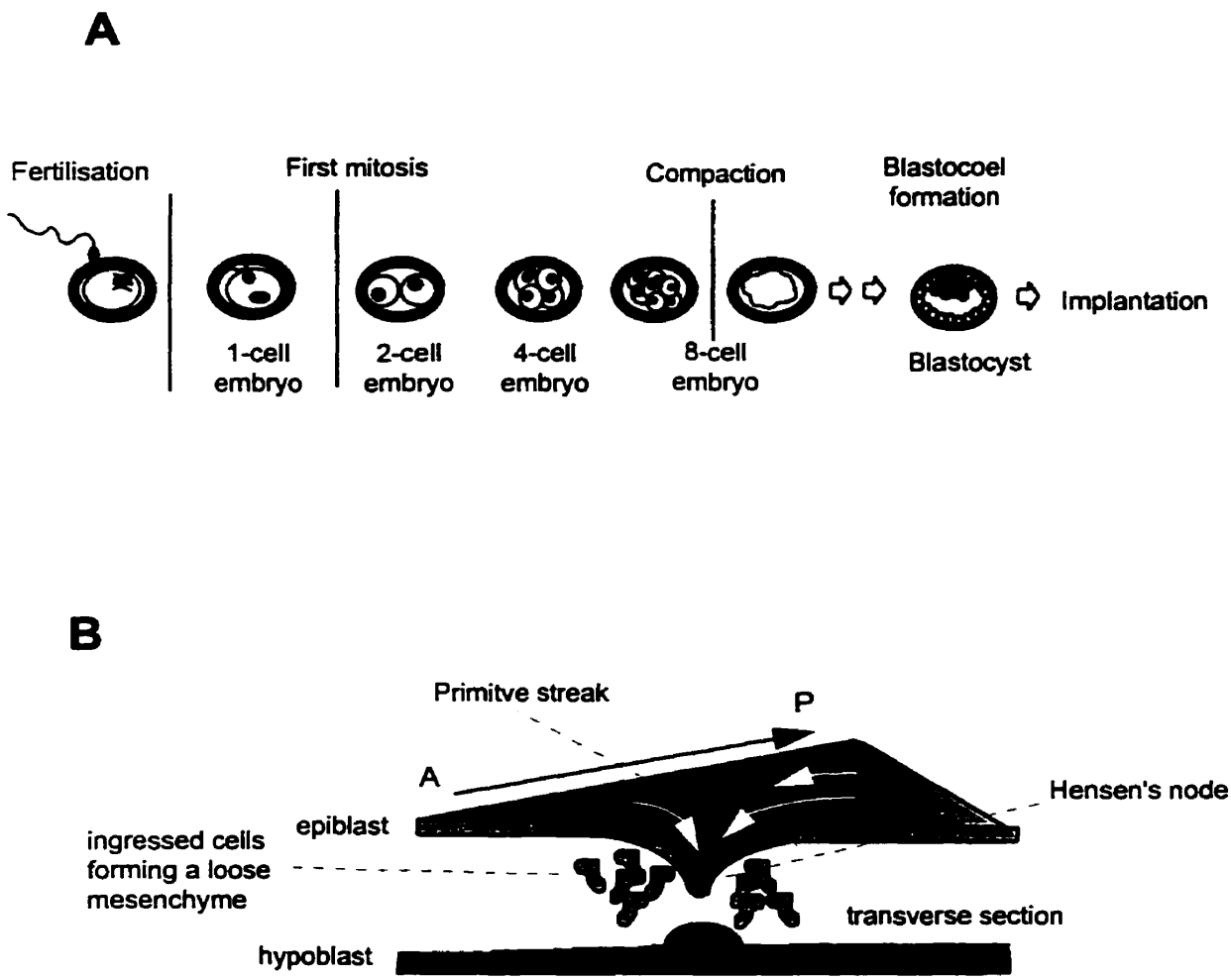


Figure 1.4.

(**A**) Pre-implantation development. After fertilization and cleavage, the embryo undergoes compaction and cavitation to form the blastocyst. The thick membrane around the embryo is the zona pellucida from which the embryos must hatch to perform implantation in the uterine wall. (**B**) Diagram of gastrulation. The primitive streak is depicted as an infolding of the epiblast and the white arrows illustrate the cell movements during gastrulation. The ingressing cells which form a loosely connected mesenchyme in the blastocoel are also depicted.

Mammalian and avian gastrulation can be divided in two steps, the formation of the epiblast and hypoblast cell layers and migration of the epiblastic cells through a primitive streak. The segregation of cells to form the hypoblast is initiated within the ICM and results in the formation of the lining of the blastocoel. Lineage tracing experiments revealed that the mammalian hypoblast, as in birds, does not contribute to the formation of embryonic structures (Rosenquist, 1966; Lawson *et al.*, 1991). The epiblast consists of the remaining cells of the ICM. The epiblastic cells are further specified to become cells that will contribute to the formation of the embryo and lining of the amnion. These latter cells secrete the amniotic fluid, in the amniotic cavity, which protects the embryo from desiccation.

The major embryological structure of the gastrula is the primitive streak. The streak establishes the antero-posterior axis of the embryo and is analogous to the amphibian blastopore. The streak consists of a thickening of the epiblast at the posterior region of the embryo which is caused by an ingression of mesodermal cells from the epiblast into the blastocoel. As the thickening narrows, the streak moves anteriorly to form the definitive primitive streak which is composed, at its anterior-most region, of a regional thickening called the Hensen's node, a structure similar to the amphibian dorsal lip of the blastopore. The formation of the primitive streak is rapidly followed by migration of cells through the primitive streak into the blastocoel. The individual migrating cells entering the blastocoel cavity form a loosely connected mesenchyme and the first stream of cells that migrate through the primitive streak are those destined to form the endoderm (Rosenquist, 1972). These cells migrate anteriorly and displace the hypoblastic cells that are pushed laterally. The second stream of infiltrating epiblastic cells spread throughout the blastoceol between the hypoblast and the epiblast and give rise to the mesodermal structures of the embryo (Smith *et al.*, 1994; Yuan *et al.*, 1995; Garcia-Martinez and Schoenwolf, 1993).

As illustrated in Fig. 1.4B, the primitive streak is not a static structure and regresses posteriorly, with the node. As the node moves, the notochord assembles itself from the chordamesodermal cells. At the end of the migration of the node to the anal region, the remainder of the epiblastic cells are destined to become presumptive ectodermal cells (Lawson *et al.*, 1991).

1.2.2. Cardiogenesis

After the formation of the three germ layers, organogenesis, the creation of organs, can proceed. In the case of cardiogenesis, recent data have shown that most of the commitment to the cardiac fate occurs within the cardiogenic field, a region of the anterior lateral plate mesoderm, and the gastrulation process is proposed to be required only to locate the precardiac mesoderm close to its inducing tissues (Tam *et al.*, 1997). These important results indicate that the process of ingression through the primitive streak, is not required for cardiac commitment but rather to generate the heart field and to perhaps polarize the heart field. Thus, it seems that the formation of the cardiogenic field is the determining factor in the initiation of cardiac cell differentiation.

Table 1.1. Milestones of cardiac development in different species [from (Fishman and Chien, 1997) and references within].

	mouse	chick	human	frog	zebrafish
Migration of cells from the epiblast	7 dpc	HH4	15-16 days	stage 10	5.5 hpf
First assembly of myocardial plate	7 dpc	HH5	18 days	stage 13	13 hpf
Generation of the tubular heart	8 dpc	HH9	22 days	stage 28	19 hpf
Beginning of contractions	8.5 dpc	HH10	23 days	stage 33	22 hpf
Looping	8.5 dpc	HH11	23 days	stage 33-36	33 hpf
Cushion formation	9 dpc	HH17	28 days	stage 41	48 hpf

1.2.2.1. The cardiogenic field

An organogenic field is defined as an area of the embryo that is competent to form an organ and reflects its developmental potential. It possesses a border where not all the cells within the field will differentiate in the prospective fate. Rather, there is a gradient of developmental potency within a field. The cells within the field are more receptive to the appropriate inductive signals than the cells at the periphery. Moreover, the field is plastic and can tolerate the ablation or the addition of cells without perturbing the final outcome of differentiation (Jacobson and Sater, 1988). The cardiogenic field obeys these rules and is

defined as a region responsive to cardiac inducing signals (Sater and Jacobson, 1990). Furthermore, the recent work of Patrick Tam and colleagues illustrates that the field can accommodate epiblastic cells that have not yet ingressed through the primitive streak and induce them to commit to the cardiac fate (Tam *et al.*, 1997).

The epiblastic cells destined to become cardiac cells are located on both sides of the primitive streak near the Hensen's node (Garcia-Martinez and Schoenwolf, 1993; Forman *et al.*, 1995). After ingression, these cells migrate rostrolaterally to form the lateral plate mesoderm which lies on each side of the paraxial mesoderm (Garcia-Martinez and Schoenwolf, 1993). The lateral plate is itself segmented in somatic mesoderm (which will form the pleura and the peritoneum) and splanchnic mesoderm. The heart field is localized within the anterior part of the splanchnic mesoderm and is thus dependent on the segmentation of the mesoderm. Studies in *Drosophila* have provided significant insights about the genetic control of precardiac mesoderm segmentation. In these studies, it was found the segmentation of the mesoderm in precardiac and visceral mesoderm is genetically dependent on the *Tinman* gene, a NK2-class homeoprotein (Bodmer, 1993; Azpiazu and Frasch, 1993). Fly embryos harboring a *tinman* (*tin*) mutation fail to segment their mesoderm in visceral and cardiac mesoderm without affecting the formation of the somatic (muscle) mesoderm. *Tinman* expression is first pan-mesodermal, under the control of the mesoderm determination factor *Twist*, a bHLH gene and then gets regionalized in the most dorsal part of the mesoderm by the action of an inductive activity localized in the overlaying ectoderm (Azpiazu and Frasch, 1993; Bodmer, 1993). Thorough genetic analyses revealed that the expression of *tinman* was dependent on the action of a TGF-β homologue, decapentaplegic (*dpp*) which is normally secreted in the dorsal ectoderm (Frasch, 1995; Staehling-Hampton *et al.*, 1994). *dpp* mutant embryos fail to regionalize *tinman* expression which is not anymore detected after the completion of mesoderm migration. This failure of maintaining *tin* expression in the dorsal mesoderm results in the absence of visceral and cardiac mesoderm specification. Furthermore, ectopic expression of *dpp*, either in the ectoderm or the mesoderm, is sufficient to induce ectopic expression of *tinman* and *bagpipe*, a *tinman*-dependent gene required for visceral mesoderm formation (Frasch, 1995; Staehling-Hampton *et al.*, 1994). Thus, *dpp* is an ectodermal signal required

for cardiac mesoderm formation. The importance of the mesoderm/dorsal ectoderm interaction is further exemplified by the characterization of the *heartless* mutation in *Drosophila* which targets the fly homologue of the FGF receptor-2. *Heartless* mutants have defects in mesodermal migration that disrupts *dpp*-mediated segmentation; ectopic expression of *dpp* in *heartless* mutants restores the formation of cardiac and visceral mesoderm (Beiman *et al.*, 1996; Gisselbrecht *et al.*, 1996). However, *dpp* only restricts the cardiac and visceral mesoderm from the somatic mesoderm and it is likely that other signals further specifies the cardiac mesoderm from the visceral mesoderm (Frasch, 1995). This possibility is substantiated by the observation that ectopic *dpp* expression results in the appearance of ectopic visceral, rather than cardiac, mesoderm. A candidate gene for such an activity is the *wingless* gene product (Wu *et al.*, 1995) which is a secreted, cysteine-rich signaling molecule that is expressed in the overlying ectoderm. *Wg*, as well as its downstream effectors, *armadillo* and *disheveled*, is required for proper cardiac mesoderm specification and its overexpression generate an increased number of cardiac precursors (Baylies *et al.*, 1995; Wu *et al.*, 1995; Park *et al.*, 1996; Azpiazu *et al.*, 1996). However, ectopic expression of *wg* does not result in ectopic tin expression but rather an enlarged expression domain within the mesoderm. Thus, in *Drosophila*, the precardiac mesoderm is specified by the combinatorial action of *dpp* and *wingless*.

In vertebrates, it appears that the early events leading to heart field formation might be conserved with those of flies. Numerous studies have established that the formation of the heart field is regulated by the neighboring endoderm and ectoderm [reviewed in (Litvin *et al.*, 1992; Fishman and Chien, 1997)]. Moreover, the discovery of the Nkx2-5 gene a vertebrate homologue of *tinman* that is expressed in precardiac mesoderm and which is required for heart formation established a molecular basis for this proposed conservation [(Lints *et al.*, 1993; Lyons *et al.*, 1995) and section 1.3.2.2 for further details]. Furthermore, the closest *dpp* homologues, BMP-2 and BMP-4 are expressed in the ectoderm and endoderm lying adjacent to the cardiac mesoderm and have been implicated in dorso-ventral patterning of the embryo. In an elegant study, Thomas Schultheiss and colleagues reported that BMP-2 and BMP-4 can induce robust expression of the early precardiac mesodermal markers, GATA-4 and Nkx2-5 in naive anterior paraxial mesoderm which is

not fated to become cardiac mesoderm. Moreover, the BMP antagonist, noggin, can inhibit cardiac myogenesis, indicating that ventralizing agents such as BMP-2 and -4 are necessary components of heart field specification in vertebrates. This study is supported by the targeted disruption of the BMP-2 gene in mice which have major cardiac defects (Zhang and Bradley, 1996). Interestingly, overexpression of Nkx2-5 in *Xenopus* and zebrafish leads to an expansion of the cardiogenic field and ventralization of the embryo, a result strikingly similar to the addition of BMP proteins to mesodermal explants (Chen and Fishman, 1996; Cleaver *et al.*, 1996). Interestingly, the action of BMP as an inducer of the cardiac fate is only accomplished in the anterior mesoderm, suggesting that the antero-posterior segmentation of the mesoderm is also an important feature of heart field formation (Schultheiss *et al.*, 1997). This segmentation is perhaps accomplished by members of the FGF family since FGF-1, -2 and -4 are expressed in the anterior endoderm and the combination of FGF-4 and BMP-2 is able to induce cardiac myogenesis in posterior mesoderm explants (Lough *et al.*, 1996). This result indicates that FGF-4 is able to induce posterior mesoderm to adopt an anterior fate, and thus be responsive to BMP proteins. However, although it is clear that FGFs are able to induce cardiac myogenesis, it is not clear if they are the cardiac-inducing activity located in the anterior endoderm. The superimposition of the patterns of expression of the FGFs in the anterior endoderm and those of the BMPs create a bilateral crescent located at the lateral-anterior margins of the embryo that is superimposable with the pre-cardiac mesoderm. The importance of wingless homologues in vertebrate heart development has not been addressed yet but a number of Wnt genes are expressed in the developing heart. The Wnt2 gene is expressed in the heart field at around 7.5 dpc, a time consistent with a role in precardiac mesoderm segmentation but its targeted disruption does not induce cardiac defects (Monkley *et al.*, 1996). Genetic redundancy among this gene family might explain the lack of cardiac phenotype and the knockout of downstream effectors of the Wnt signal might be a better approach to unravel a role for Wnt signaling in cardiac development.

Collectively, the current body of evidence enables us to propose a model for heart field formation, largely based on the model of Schultheiss *et al.* (1997) (Fig. 1.5C). Heart field formation would be the result of a combinatorial action of BMP proteins that are

expressed at the lateral margins of the embryo and a signal from the anterior endoderm, probably a member of the FGF family. The superimposition of these signals would result in heart cell commitment and differentiation. The importance of the Wnt transduction cascade remains to be addressed but it might also play a ventralizing role. The fact that Nkx2-5 is induced upon BMP treatment leads to the conclusion that the dpp-*tinman* pathway is conserved in vertebrates, an observation that would enable researchers to use the power of *Drosophila* genetics to unravel the early steps of cardiac mesoderm formation.

1.2.2.2. Myocyte lineage specification

The formation of the heart field is accompanied by cell movements that will lead to the formation of the tubular heart. At stage 8 in chick (see Table 1.1), neurulation occurs and the epithelial sheet forming the splanchnic mesoderm is brought at the midline where it fuses to form the tubular heart at stage 10 (Fishman and Chien, 1997). Just prior fusion, rhythmically beating cells can be observed, the hallmark of myocardial cell differentiation (Montgomery *et al.*, 1994). Terminal cardiac cell differentiation is defined by the ability of the cell to beat and to express a set of phenotypical markers such as the natriuretic hormones and the extracellular matrix proteins. This differential, cardiac-specific, expression is controlled by a set of transcriptional regulators that will be discussed in greater details in section 1.3. The formation of the primitive heart tube is also the hallmark of myocyte lineage specification where myocytes further differentiate in atrial, ventricular and cardiac conduction system myocytes.

The demarcation between the atrial and ventricular lineages can be observed with probes specific to compartment-specific myosins as soon as stage 8 (Bisaha and Bader, 1991; Yutzey and Bader, 1995). At this point, the precardiac mesoderm is still an epithelium and very few cell mixing is observed. Fate mapping experiments revealed that the anterioposterior identity of the future heart cells (atria being the most posterior compartment, followed by the ventricle and the bulbus cordis) matches the pattern of ingression (Stainier *et al.*, 1993). These observations suggest that ingression might play a role in the definition of the anteroposterior identity of myocardial cells. Ingression occurs in an antero-posterior fashion and it is possible that either the timing of ingression or the

position relative to the node determines the myocardial cell fate along the anteroposterior axis. Experiments using the morphogen retinoic acid (RA) in zebrafish (Stainier and Fishman, 1992) or chicken (Yutzey *et al.*, 1994) revealed that it was possible to modify the antero-posterior patterning of the heart tube. Treatments of gastrulating embryos with RA perturbs the formation of the heart tube and results in a dose-dependent posteriorization of the tube. The expression domains of the atrial-specific markers AMHC (Yutzey *et al.*, 1994) or mAb S46 (Stainier and Fishman, 1992) are expanded suggesting that these heart tubes would form larger atria. These results suggest that the heart field is polarized and plastic up until the fusion of the bilateral primordia (Yutzey *et al.*, 1994). Furthermore, transplantation of precardiac mesoderm in another region of the field will result in the adoption of the fate of the host region. Thus, it does not seem that the atrial and ventricular cells arise form a common population of myoblast precursors but rather from distinct regions of a polarized cardiogenic field. The source of polarization of the field remains undefined but signals from the node are likely to be involved. Interestingly, many experiments looking at the inductive signals emanating from either the node or the endoderm such as those looking at BMPs, activins and FGFs, never looked at the expression of chamber-specific markers. It would be interesting to repeat these experiments using chamber-specific markers in order to possibly identify atrial or ventricular-specific signals. Finally, the large-scale genetic screens in zebrafish identified several mutants defective in atrio-ventricular specification (Stainier *et al.*, 1996; Chen *et al.*, 1996). Interestingly they are all characterized by a deletion of the ventricular chamber. These mutants establish a genetic basis for the specification of myocytes and the molecular cloning of these genes will greatly help our understanding of myocyte lineage specification.

1.2.2.3. Differentiation of the endocardial and Purkinje fibers

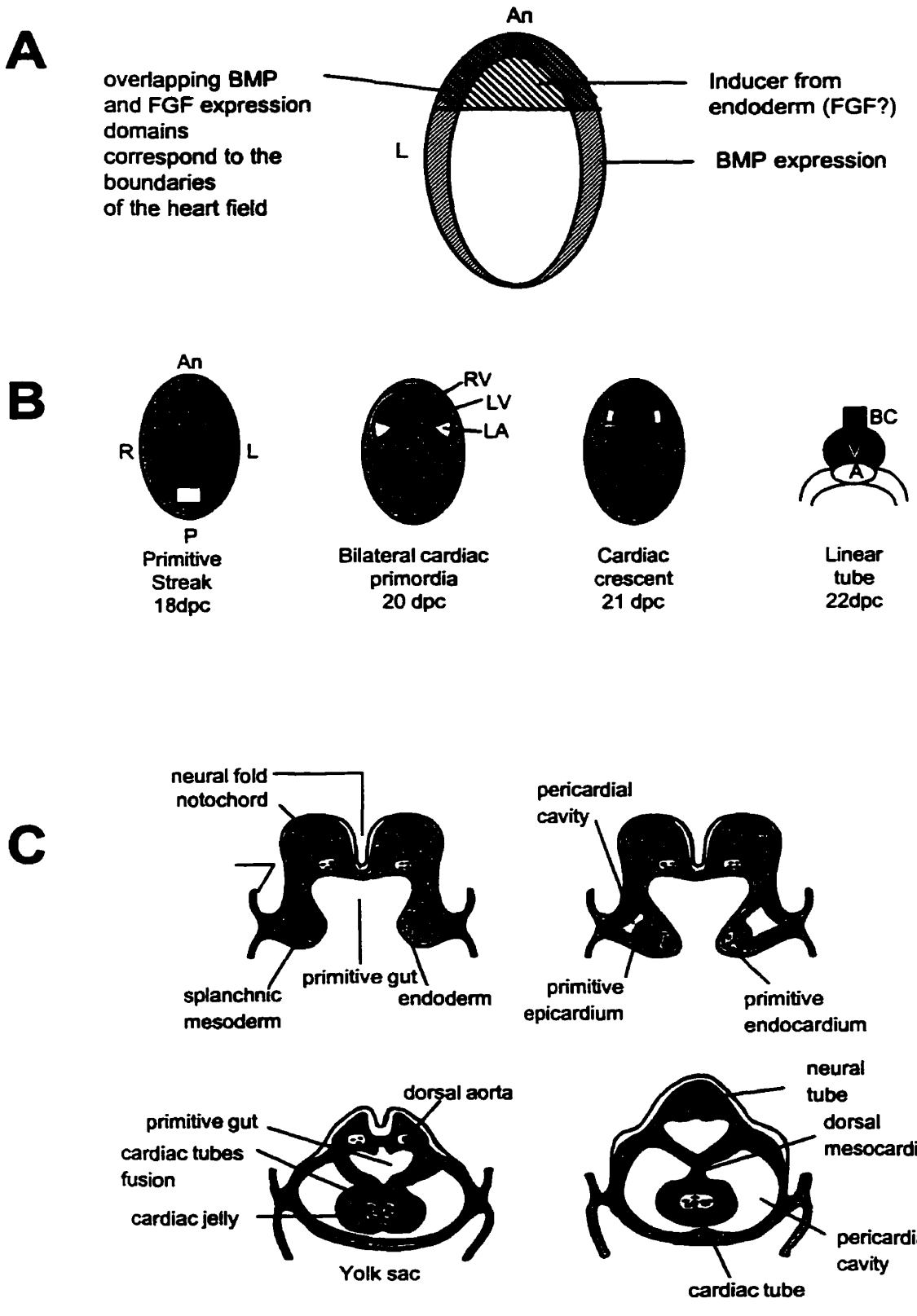
At the heart tube stage, the embryonic blood circulation begins. The high blood pressure of higher chordates would be detrimental for cardiomyocytes without the presence of a specialized endothelial lining, the endocardium. The origins of endocardial cells are still debated but the consensus seems to tend to a heart field origin. Fate-mapping studies in zebrafish located the endocardial progenitors sequestered within the heart field (Lee *et*

al., 1994). However, it remains to be answered if they really represent a subdivision of the field or represent a group of mesodermal cells that migrated to this location. Zebrafish mutant for *cloche*, have a total absence of endocardial cells and have severe defects in hematopoiesis (Stainier *et al.*, 1996). This is consistent with the previous suggestion that endocardial, endothelial and hematopoietic cells originate from a common precursor, the hemangioblast localized in the ventral region of the embryo (Stainier *et al.*, 1996). However, these progenitor cells are not restricted to the cardiovascular fate and still have developmental plasticity. This is perhaps why some cells harbor, within the field, markers for both myocardial and endocardial cell types (Eisenberg and Bader, 1995). Thus, pre-endocardial cells are sequestered at the ventral-most region of the field and are incorporated in the tubular heart by the action of the bilateral cardiac primordia during fusion at the midline.

The contractions of the early heart tube are peristaltic but nevertheless highly coordinated to push the blood from the sinus venosus to the outflow tract. These contractions are initiated by a pacemaker localized on each sides of the inflow tract and get later restricted to the left side of the inflow tract (de Jong *et al.*, 1992). The mature cardiac conduction system (CCS) is composed of a central element composed of the sinoatrial and atrioventricular nodes and the atrioventricular bundle. The action potentials are generated form the sinoatrial node and propagate through the atria and initiate their contraction. The action potential is not transmitted to the ventricular cells but rather to the atrioventricular node and then, propagated rapidly to the atrioventricular bundle. The action potential is finally transmitted to the ventricular muscle through the network of Purkinje fibers (Icardo and Manasek, 1996). The differentiation of the components of the CCS remain poorly understood largely because of the lack of knowledge of their cellular origins. The Purkinje fibers can be first by their high levels of the gap junction protein connexin 42 (connexin 40 in mice) (Gourdie *et al.*, 1993; Delorme *et al.*, 1995). Their differentiation is closely related, in a spatial and temporal manner, with the development of coronary arteries (Gourdie *et al.*, 1995). Retroviral-mediated cell tagging by the group of Mikawa revealed that ventricular myocytes tagged as early as the heart tube stage can form clones of cells differentiating in Purkinje fibers, unraveling a myocyte origin of the peripheral element of

Figure 1.5.

(A) Cardiac mesoderm induction. Schematic representation of cardiac induction. This model is based on the model of Schultheiss *et al.* (1997) except that the anterior endoderm cardiac-inducing activity which they termed X is described here as an FGF-related molecule based on a recent study (Lough *et al.*, 1996). The overlapping of the BMP and FGF expression domain corresponds to the heart crescent. **(B)** Diagram of cell movements occurring in early human cardiac development. Please note that the anteroposterior identity of the future cardiac compartments is established at the late primitive streak stage. Abbreviations= *R*: right; *L*: left; *P*: Posterior; *An*: Anterior; *RV*: future right ventricle; *LV*: future left ventricle; *LA*: future left atrium; *BC*: bulbus cordis. **(C)** Formation of the primitive heart tube showing the anatomical modifications leading to the formation of the primitive heart tube.


A
An
overlapping BMP
and FGF expression
domains
correspond to the
boundaries
of the heart field
L
Inducer from
endoderm (FGF?)
BMP expression
B
An
R
L
P
Primitive
Streak
18dpc
RV
LV
LA
Bilateral cardiac
primordia
20 dpc
Cardiac
crescent
21 dpc
BC
V
A
Linear
tube
22dpc
C
neural fold
notochord
splanchnic
mesoderm
primitive gut
endoderm
pericardial
cavity
primitive
epicardium
primitive
endocardium
dorsal aorta
primitive gut
cardiac tubes
fusion
cardiac jelly
Yolk sac
neural
tube
dorsal
mesocardium
pericardial
cavity
cardiac tube

the CCS (Gourdie *et al.*, 1995). Purkinje fibers express myocyte markers as well as neural markers, which had previously led to the suggestion that they might originate from neural crest migration. However, the tagging experiments were done before the migration of the cardiac neural crest cells, thus excluding a neural crest origin for the Purkinje fibers (Mikawa and Fischman, 1996).

1.2.3. Cardiac morphogenesis

The primitive heart tube has to undergo major morphogenetic events to yield the mature, four-chambered heart. The range of morphogenetic events comprise cardiac looping which is initiated after the definition of the left-right body axis, septation and valvulogenesis. The four-chambered heart will then be able to start the pulmonary circulation, which is not readily needed in the embryo. The maturation of the inflow and outflow tracts will not be discussed here but are extremely important as well since defects in these processes often lead to congenital malformations.

1.2.3.1. Left-right asymmetry

The looping of the heart to the right is the first observable left-right (LR) asymmetric event of embryogenesis. Moreover, the sidedness of cardiac looping and visceral asymmetry is the same among all vertebrates, implying a conserved mechanism for LR axis definition. It is thought that LR axis is defined subsequent to, and relative to the other two major axis of the embryo, the antero-posterior and dorso-ventral axis and it has been shown that alteration of the dorsoventral axis would lead to a randomization of the LR axis. Laterality can be sometimes reversed in mice or humans, a situation called *situs inversus*. In this condition, the heart is looped to the left and the spleen, stomach and the trilobed lung are located on the right side of the individual (Yost, 1995). Sometimes, the sidedness of the thoracoabdominal organs is less clear and leads to a situation called heterotaxy (Yost, 1995). These conditions have a genetic basis since a few, non-linked, recessive mutations leading either to the development of situs inversus or heterotaxia have been found in humans and mice. In mice, where they have been well characterized, these mutations occur either in the *iv* gene which lead to heterotaxia, or at the *inv* locus which

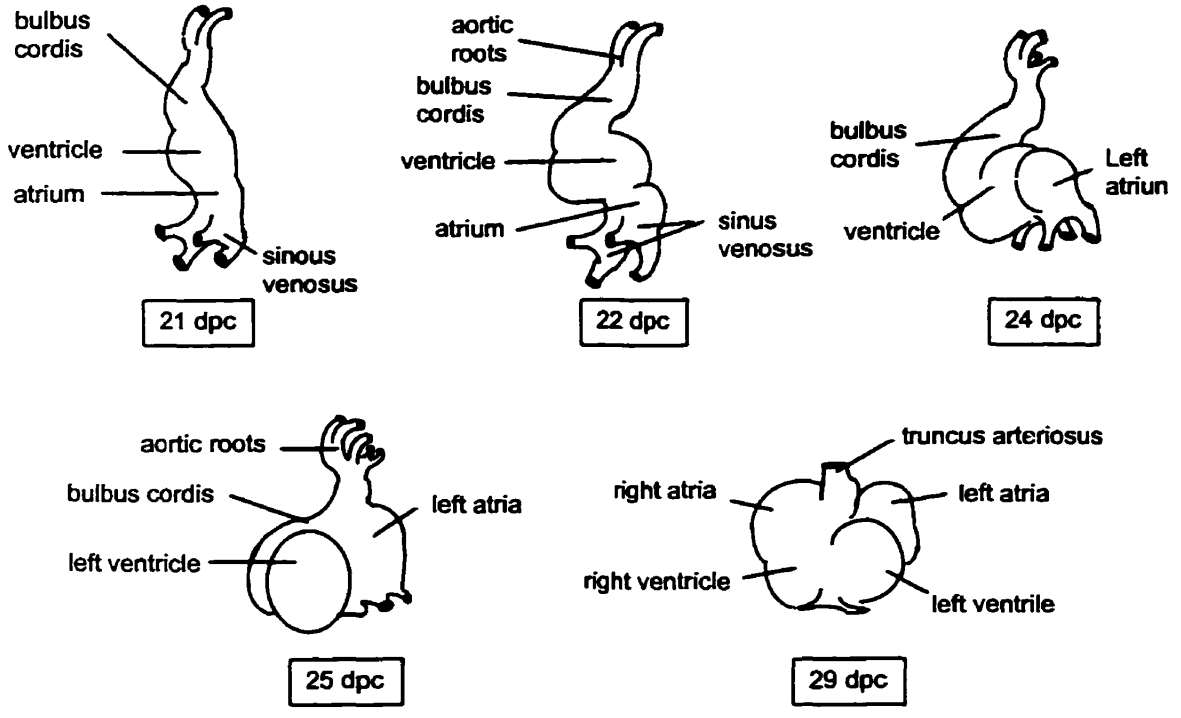


Figure 1.6.

Development of the four-chambered, mature heart from the primitive heart. The indicated days correspond to those of human gestation. The sequence of events leading to heart tube formation is the same in mice and in humans. Please look at Table 1.1. for the equivalent embryonic stages in the two species.

result in an almost fully penetrant reversal of the LR body axis (Yokoyama *et al.*, 1993; Brueckner *et al.*, 1989). These laterality defects have clinical importance since they are often associated with pathological conditions related to cardiac malformations (Kartenager and an X-linked syndrome in humans) or liver dysfunction (*inv* in mice) (Yost, 1995). However, none of the genes mutated in these mice models or genetic diseases have been cloned so far.

Recently, intense efforts have been directed toward the characterization of asymmetrically expressed genes in both chicken and mice. The rationale for these searches were that the first events leading to LR asymmetry has to involve asymmetric expression of signaling molecules and effectors. Moreover, since definition of the LR body axis seems to involve components of the DV or AP axis, and that the definition of the LR asymmetry has to occur prior cardiac looping, the group of Cliff Tabin carefully examined, by whole-mount *in situs*, the expression of a number of developmental regulator during the primitive streak stage, up to the first few somites stage. They found that a few genes were transiently and asymmetrically expressed on either side of the embryo. The first asymmetrically expressed gene is, in the chick, the activin receptor-IIa (cActR-IIa) which is transiently expressed on the right side of the primitive streak. This expression is proposed to activate an activin-mediated suppression of the sonic hedgehog (shh) signaling on the right side. Thus, shh expression on the left side is regulated by activin suppression on the right. Shh turns on the asymmetric expression of TGFβ-related genes, nodal and lefty, which, in turn, might repress the expression of the putative transcriptional repressor, snail-related (SnR) (Meno *et al.*, 1996; Collignon *et al.*, 1996; Isaac *et al.*, 1997). Treatments of chick embryo with antisense oligonucleotides to the cSnR gene results in randomization of cardiac looping (Isaac *et al.*, 1997). The asymmetric expression of these genes is downstream of the *inv* an *iv* genes since their expression is either randomized in *iv* mice or on the opposite side in *inv* mice (Lowe *et al.*, 1996; Meno *et al.*, 1996). However, how these laterality signals can control embryonic turning or cardiac looping is still unknown.

1.2.3.2. Cardiac looping

Establishment of LR body axis is a prerequisite for cardiac looping.

Morphologically, cardiac looping is initiated by the bending of the ventral side of the heart tube. The whole tube then turns toward the right side of the embryo, resulting in an S-shaped tube. The future right ventricle is located to the right and the future left ventricle and atria then start to migrate rostrally. How laterality signals affect looping is still unclear. Differential apoptosis or mitotic divisions have been proposed to explain cardiac looping but these mechanisms do not seem to occur at that stage of cardiac morphogenesis (Icardo and Manasek, 1996).

Studies carried out in *Xenopus* revealed that perturbation of the extracellular matrix can result in randomization of cardiac looping. Injection of Arg-Gly-Asp peptides, which disrupts fibronectin fibrils, resulted in heterotaxia (Yost, 1992). These experiments are concordant with models of cardiac looping involving the extracellular matrix proposed originally by Wolpert and Manasek [(Icardo and Manasek, 1996) and references therein]. Briefly, in these models, the epithelial nature of the heart is extremely important to respond mechanically to tensions generated by either the ectoderm-derived ECM, or the cardiac jelly. Prior to looping, the heart is composed of a relatively thin and compact myocardial wall separated form the endocardium by a thick zone of extracellular matrix, the cardiac jelly. The cardiac jelly is synthesized by both the myocytes and the cardiac fibroblasts, and it is composed mainly of glycosaminoglycans, glycoproteins, collagens and proteoglycans. The mechanical forces resulting from the growth of the extracellular matrix would result in the observed deformations of the cardiac tube. The asymmetrical deformation of the tube would be the result of a differential compliance of the cardiac epithelium (Manasek *et al.*, 1972; Johnson *et al.*, 1974; Icardo and Manasek, 1996). Recently, a component of the extracellular matrix, flectin, has been asymmetrically localized on the left side of the early looping heart (Tsuda *et al.*, 1996), suggesting that components of the cardiac extracellular matrix might be effectors of cardiac looping.

1.2.3.3. Post-looping morphogenesis

After looping, the heart is still not the four-chambered pump. It has to undergo major external and internal changes. At this point, many factors are thought to be involved such as differential growth and apoptosis, cell migration and perhaps hemodynamic factors.

External changes comprise mainly reorganization of the bulbus cordis, modifications in the ventricular shape and an expansion of the primitive atrium toward the left side. In addition, a constriction occurs between the ventricle and the atrium, forming the atrioventricular canal. Finally, the sinus venosus is also reorganized to be incorporated to the dorsal wall of the atrium (Icardo and Manasek, 1996).

The first major internal transformation is a thickening of the ventricular wall initiated by mitotic divisions of the myocytes. This thickening is followed by an eptihelio-mesenchymal transformation event. The myocytes migrate vertically in the cardiac jelly, toward the endocardium, resulting in an oligoclonal cone-shaped formation that will form the primitive trabeculae (Mikawa *et al.*, 1992b; Mikawa *et al.*, 1992a). Following the model of Mikawa and Fischman (Mikawa and Fischman, 1996), these cone-shaped ingrowths are the morphogenetic unit of the heart and are involved in ventricular wall growth, septal wall genesis and primary trabeculation. The trabecular network is extremely important for cardiac physiology. Its absence would lead to severe dysfunctions in contractility since the heart would be less compliant and too compact. Moreover, it maximizes the surface of contact with blood, especially in the thick ventricular wall. Interestingly, the fate of the trabeculae varies: they either remain as is or coalesce to be incorporated in the ventricular wall. There are still very few molecules or signaling pathways known to be involved in the process of trabeculation. An interesting candidate is the FGF signaling pathway since FGF acts as an endogenous mitogen. Expression of a dominant-negative FGF receptor results in suppression of myocyte proliferation *in vivo*, in chicken heart development (Mima *et al.*, 1995). In addition, adhesion molecules such as N-cadherin are also required for the regulation of myocyte migration, probably for the epithelio-mesenchymal transformation. Finally, molecules that facilitate the migration of myocytes in the cardiac jelly such as hyaluronidases are upregulated and are likely to be required during trabeculation (Icardo and Manasek, 1996).

Septation occurs independently in four regions of the primitive heart: the atrium, the atrioventricular canal, the bulbus cordis and the ventricle. Septa either form from the coalescence of the primordium trabeculae in the case of the interventricular septum; the outgrowth of myocardial tissues (interatrial septum and septum secundum); or the fusion of

cushion tissues (atrioventricular canal septum and bulbar septation) (Icardo and Manasek, 1996). Valvulogenesis starts by the swelling of the cardiac jelly at two specific sites, the atrioventricular canal and the outflow tract (Eisenberg and Markwald, 1995). This swelling of the cardiac jelly is caused by a specific synthesis of ECM proteins by the myocardial cells of these two regions. The regionalized synthesis of ECM is proposed to be due to the patterning of the heart tube since modification of either the antero-posterior axis by retinoic acid or the LR axis (in *iv* or *inv* mice) leads to valvular defects (Yost, 1995). Consistent, with this hypothesis is the AV canal- and outflow tract-specific expression of two molecules involved in patterning, msx-2 and BMP-4 (Yost, 1995). Following their formation, cardiac cushions get cellularized by a mesenchymal invasion of endocardial cells. The dissection of the molecular events leading to the formation of the definitive, cellularized cushion is greatly helped by the development of an *in vitro* model based on the ability of the epithelial endocardium to undergo mesenchymal transformation on a collagen gel (Brown *et al.*, 1996). Using this assay, it was shown that only the endocardial cells from the AV canal and outflow tract were able to undergo transformation and that transformation was triggered by a signal from the AV canal or outflow tract myocardium. Moreover, it was possible to demonstrate the requirement of the ES complex as well as TGFβ signaling (Brown *et al.*, 1996). In addition to these molecules, down-regulation of the adhesion molecule N-CAM and the up-regulation of tenascin is required for the migration of the mesenchymal cells within the cushion (Eisenberg and Markwald, 1995). The final step in valvuloseptal development is the differentiation of the mesenchyme in embryonic valves (and septa). The regulatory proteins involved in this last step are not known yet but it was observed that two fibrillar proteins: fibrillin and fibulin were specifically expressed at this point. Examination of the regulatory networks controlling their expression in terminal valvuloseptal differentiation might be a useful tool to decipher this important step of cardiac morphogenesis.

1.2.3.4. Genetic analysis of cardiac morphogenesis

In past few years, the large-scale genetic screens in zebrafish and the generation of targeted mutations in the mouse have identified genes that are required for proper cardiac

Table 1.1. Genetic control of cardiac morphogenesis

Process	Gene	Model organism	References
Heart tube formation	GATA-4	mouse	(Kuo *et al.*, 1997;
	bonnie and clyde	zebrafish	Molkentin *et al.*, 1997)
	miles apart	*ibid*	(Stainier *et al.*, 1996)
Chamber generation	*pandora*	zebrafish	(Stainier *et al.*, 1996)
	lonely atrium	*ibid*	(Chen *et al.*, 1996)
Axial positioning	*heart and soul*	zebrafish	(Chen *et al.*, 1996)
Looping/laterality	Nkx2-5	mouse	(Lyons *et al.*, 1995)
	MEF2C	*ibid*	(Lin *et al.*, 1997)
	dHAND	*ibid*	(Srivastava *et al.*, 1997)
	overlooped	zebrafish	(Chen *et al.*, 1996)
Valve/septum formation	RXRα	mouse	(Sucov *et al.*, 1994)
	NT-3	*ibid*	(Donovan *et al.*, 1996)
	PDGFα	*ibid*	(Schatteman *et al.*, 1995)
	VCAM-1	*ibid*	(Kwee *et al.*, 1995)
	jekyll	zebrafish	(Stainier *et al.*, 1996)
Concentric growth	N-Myc	mouse	(Charron *et al.*, 1992)
	WT-1	*ibid*	(Kreidberg *et al.*, 193)
	bARK-1	*ibid*	(Jaber *et al.*, 1996)
	valentine	zebrafish	(Stainier *et al.*, 1996)
	santa	*ibid*	(Stainier *et al.*, 1996)
	heart of glass	*ibid*	(Stainier *et al.*, 1996)
Trabeculation	neuregulin	mouse	(Meyer and Birchmeier, 1995)
	ERB-2	*ibid*	(Lee *et al.*, 1995)
	ERB-4	*ibid*	(Gassmann *et al.*, 1995)
	RXRα	*ibid*	(Sucov *et al.*, 1994)
Aortic arch/sac	dHAND	mouse	(Srivastava *et al.*, 1997)
development	HoxA3	*ibid*	(Chisaka and Capecchi,
	Splotch	*ibid*	1991)
	Endothelin-1	*ibid*	(Epstein *et al.*, 1991)
	Connexin 43	*ibid*	(Kurihara *et al.*, 1995)
	NT-3	*ibid*	(Reaume *et al.*, 1995)
			(Donovan *et al.*, 1996)

morphogenesis. Interestingly, all these genes affect aspects of cardiogenesis posterior to the differentiation of the cardiac myocyte. These observations are consistent with a model in which the formation of the cardiac cell does not depend on a single gene but is rather the result of a combination of redundant gene networks. Nevertheless, the genetic dissection of cardiac morphogenesis will be an extremely useful too to understand the mechanisms underlying heart development and also to identify potential disease-causing genes that, when mutated, result in congenital cardiac malformations.

The zebrafish is an ideal model to study cardiac development. It is a vertebrate with chambers and valves (although it is not four-chambered). The embryo is transparent and the heart is accessible and visible at all stages of cardiac development. Moreover, unlike in mouse or chicken, the heart does not have to be functional until very late in its development. These features, combined to the possibility of generating large-scale genetics screens and the ability to raise a large numbers of fish that are at the same developmental stage, make the zebrafish model an extremely useful one (Stainier and Fishman, 1994).

Although spontaneous mutants have arisen in the zebrafish population that have some deficiencies in cardiac development. The most noticeable example being the *cloche* mutant where no endocardium is formed (Stainier *et al.*, 1995), the recent completion of the large-scale screen in zebrafish revealed interesting cardiovascular mutants (Driever *et al.*, 1996; Stainier *et al.*, 1996; Chen *et al.*, 1996; Haffter *et al.*, 1996). Mutants affecting heart morphogenesis can be classified in six large groups (see Table 1.2): 1) Mutations affecting fusion of the primordia at the midline such as *miles apart*. In these mutants, two hearts with their endocardium form on either side of the embryo. 2) Mutations affecting ventricle formation such as *lonely atrium* and *pandora*. These mutants does not produce a ventricle and have larger atria, establishing a genetic basis for chamber formation. 3) An axial positioning defect is observed in the *heart and soul* mutant where the ventricle develops inside the atrial chamber. 4) Looping and laterality mutants such as in the *overlooped* mutants that have the atrium and ventricle placed at 90 degrees from each other or in the *curled-up* mutant which results in randomization of the orientation of the cardiac looping. 5) Valvular deficiencies are found in the *Jekyll* mutant where no valves are formed. 6) Finally, cardiac growth defects characterize the class of mutants that include *valentine*,

santa and *heart of glass*. The hearts of these mutants are enlarged due to a thinner myocardium. Additionally, they do not form valves, and it is proposed that, in this class of mutants, the defects arise from a problem in endocardium-myocardium communications (Fishman and Chien, 1997). Thus, the genetic screens in zebrafish already provide an interesting set of gene networks involved in different aspects of cardiogenesis. The molecular cloning of these genes and the understanding of how they interact with each other will give heart developmental biology a major leap in the next decade. However, since zebrafish genomics is still in its infancy, the positional cloning of genes involved in cardiovascular development will be tedious and long. This last characteristic is the major drawback for the zebrafish model.

In mice, very few spontaneous or mutagen-induced mutations have been found to affect cardiac development, most probably because a functional heart is required early in development. However, several targeted gene mutations produce a cardiac phenotype. Cardiac development in mice is very similar to that of human and processes such as trabeculation and septation are not present in zebrafish, making the mouse a necessary genetic model to study cardiac development. The genetic approach to cardiovascular development in mice is mainly based on reverse genetics where candidate genes are knocked-out and their phenotypes analyzed. Since the heart is the first functional organ in the embryo and since its functionality is required for viability, a number of lethal mutations have been found to affect heart development [reviewed in (Rossant, 1996; Fishman and Chien, 1997; Grépin *et al.*, 1995a)].

The classes of mutations found in mice are similar to those of zebrafish except that mutations affecting trabeculation and septal development can be added. This apparent genetic conservation indicates that mammals and fish perhaps share a common genetic program. Interestingly, we can further divide the mutants found to affect cardiac development in the mouse in three functional classes of genes: adhesion and ECM molecules, signaling proteins and transcription factors. This observation reinforces the notion that multiple cellular activities operate in cardiac development. However, it is too soon to reconstitute these pathways and epistatic studies will have to be performed.

1.2.4. Perinatal and post-natal development

The development of the heart does not stop after the formation of the mature four-chambered heart around 15 dpc in the mouse. The heart grows in size, mainly through mitotic divisions. However, following birth, the mitotic index of the heart decreases sharply and a genetic reprogramming of the heart occurs. These major events have serious consequences on the normal and pathological physiology of the heart.

1.2.4.1. Perinatal gene switching

The cardiac physiology of the embryo is different from the physiology of the post-natal heart. The electrophysiology as well as the compliance is different in the post-natal heart in response to a different hemodynamic load. These changes in the physiological properties of the heart are caused by a genetic reprogramming of the heart. This reorganization of the gene expression program is concomitant with the fall of the mitotic index (Manasek, 1968; Claycomb, 1992; Chen and Solursh, 1995) and includes myosin isoform switching, extinction of the expression of the *ANF* gene in ventricles and the apparition of muscle creatine kinase expression in the atria (Lyons, 1994). The molecular pathways involved in this process are still ill-defined but seem to involve a combination of positive and negative transcriptional regulatory pathways.

1.2.4.2. Hypertrophic growth

It is often stated that a hallmark of terminal differentiation is the concomitant exit from the cell cycle. This exit from the proliferative state is necessary for differentiation since forced proliferation inhibits or interferes with the differentiation program. Interestingly, cardiac muscle cells undergo differentiation in the presence of proliferation during embryogenesis (Puelles and Rubenstein, 1993). However, shortly after birth, the mitotic index of cardiomyocytes decreases sharply enabling heart cells to complete their terminal development (Chen and Solursh, 1995). Thus, in cardiac muscle, there is a dichotomy between proliferation and differentiation in embryogenesis and a link (similar to skeletal muscle) between the exit from the cell cycle and terminal development in perinatal development. After birth, the heart mass increases mainly by the increase of myocyte

cytoplasmic content, a process called hypertrophic growth. At a cellular level, hypertrophic growth can be triggered by treatment of cardiac cells with mitogenic signals such as growth factors (Morgan and Baker, 1991), or following a mechanical stress that releases hypertrophic cytokines such as catecholamines, angiotensin II and perhaps cardiotrophin (Simpson, 1985; Sadoshima *et al.*, 1993; Pennica *et al.*, 1995). Interestingly, incubation of cardiac cells with either mitogenic signals (Claycomb, 1983) or forced expression of the adenoviral protein E1a is able to trigger entry of myocytes to S-phase although the cardiomyocyte remains refractory to mitosis and cytokinesis (Lee *et al.*, 1996; Soudais *et al.*, 1994). At this time, little is known about this total block to enter M phase. Heterokaryon experiments have not yet been done to demonstrate if this phenomenon was recessive (the absence of a mitosis factor) or dominant (the presence of a cardiac inhibitor of mitosis). Nevertheless, it is possible to interfere with the perinatal withdrawal of the cell cycle when a strong mitogenic signal, such as the SV40 T antigen is consistently expressed prior to the decline of the mitotic index (Field, 1988; Katz *et al.*, 1992).

Uncontrolled hypertrophic growth leads to a pathological condition called cardiac hypertrophy (Simpson, 1985; Izumo *et al.*, 1988). In pathological conditions such as those following a myocardial infarct or hypertension, the heart responds to the mechanical stress by an hypertrophic growth. This response has immediate beneficial effects by increasing the strength of the contraction but diminishes the compliance and, in the long-term, leads to a further increase in mechanical stress. At the molecular level, a hallmark of cardiac hypertrophy is the ventricular re-expression of fetal genes such as ANF and β-MHC. This genetic reprogramming is responsible for the modifications of cardiac contractility and compliance that accompany cardiac hypertrophy. Interestingly, this genetic reprogramming enables the molecular dissection of the transcriptional pathways controlling fetal gene re-expression, and the identification of transcriptional regulators governing this process will provide interesting targets for therapeutics.

1.3. GENETIC CONTROL OF CARDIOGENESIS

The establishment and the maintenance of the cardiac phenotype requires the activation of cardiac-specific as well as muscle-restricted genes in a tightly regulated and

spatial manner. The great majority of cardiac genes studied so far, besides the natriuretic peptide genes, are co-expressed in the adult or embryonic skeletal muscles. This observation led to the hypothesis that common transcriptional pathways were shared between the myocardium and the skeletal muscle. For this reason, many groups undertook the study of transcriptional regulation of contractile protein genes in myoblastic cell lines. Their hypothesis was that a large subgroup of regulatory pathways would be shared between the myocardium and the skeletal muscle. Thus, *in vitro,* analyses of the promoters of genes encoding the muscle creatine kinase (MCK), actins, myosins and troponins enabled the identification of regulatory elements specifically active in myotubes or myoblasts [see (Olson, 1992) for example]. However, gene-targeted disruption of the myogenic bHLH proteins indicated that cardiac and skeletal muscles possess different transcriptional regulatory pathways (Hasty *et al.,* 1993; Rudnicki *et al.,* 1992; Braun *et al.,* 1992). Moreover, a growing body of work based on analyses of promoters in transgenic mice and in primary cardiac cells clearly establishes that distinct regulatory mechanisms are involved in the regulation of transcription in the myocardium (Lyons, 1996).

1.3.1. Cardiac cis-elements.

During the past decade, the development of neonatal or embryonic cardiomyocyte primary cultures from rats or chicken, as well as the utilization of transgenic mice enabled the identification of several cardiac-restricted *cis*-acting elements. Most of these motifs are shared by more than one cardiac promoter or enhancer and by more than one class of cardiac genes (e.g. contractile protein genes and natriuretic peptide genes). These cardiac *cis*-elements cannot, by any means, explain the dynamic patterns of expression observed in the cardiac muscle. Therefore, many cardiac elements remain to be discovered and more importantly, they are likely to interact with each other and with ubiquitously active elements in order to generate plastic and dynamic transcriptional activities.

1.3.1.1. The GATA element.

The existence of GATA elements was first discovered from studies on erythroid promoters. These elements are defined by the consensus (A/T)GATA(A/G), or WGATAR

and are critical for the expression of numerous erythroid genes as exemplified by an elegant genetic studies in which a punctual mutation in a GATA element was found to abolish the erythroid-specific expression of the Duffy gene (Tournamille *et al.*, 1995). Moreover, GATA elements have been found to be important for the tissue-specific expression of numerous genes expressed in a variety of cell types including the endothelium (Wilson *et al.*, 1990), megakaryocytes (Martin *et al.*, 1990), neurons (Lawson *et al.*, 1996), lymphocytes (Joulin *et al.*, 1991) and mast cells (Martin *et al.*, 1990). GATA elements are often disposed in tandem, palindromic or overlapping arrangements suggesting that they are influenced by their promoter context. For example, they are found in LCR regions of the globin clusters where, in association with MARs and other erythroid-specific elements, they play a critical role in establishing erythroid-specific hypersensitive sites (Fong and Emerson, 1992). In addition, GATA elements also exist as a specialized TATA box of numerous genes such as β-globin and BNP (Fong and Emerson, 1992; Grépin *et al.*, 1994). The exact role for these -30 GATA sites is still debated but they are required for enhancer-dependent transcription, at least in the case of the chicken β-globin gene. However, it is clear that GATA-binding proteins cannot act as an alternative to TFIID since TBP binds these elements and the binding of TBP and GATA-binding proteins is mutually exclusive (Fong and Emerson, 1992; Aird *et al.*, 1994). Finally, GATA motifs are often present in composite elements where their activity is modified by the presence of overlapping motifs. For example, in the IL-2 receptor-α promoter, a GATA motif is required for an IL-2 response element. This motif is embedded within HMG(I)Y and STAT-5 binding sites and the occupancy of the site by GATA-binding proteins requires STAT-5 activation (John *et al.*, 1996). Thus, GATA motifs play versatile and diverse roles in transcriptional regulation via their functional interactions with neighboring *cis*-elements.

The first important improvement in our knowledge of cardiac-specific transcription came with the studies on the transcriptional regulation of the cardiac natriuretic peptides genes. Two natriuretic peptides are synthesized by cardiac muscle, the atrial natriuretic factor (ANF) and the B-type natriuretic peptide (BNP). These peptides are secreted and act via specific receptors on many target tissues in order to increase natriuresis, diuresis and vasodilatation [reviewed in (Rosenzweig and Seidman, 1991)]. The first gene of this

TABLE 1.2. Cardiac *cis*-elements and their cognate binding proteins. Please see text for references.

Cis-element	Consensus	Binding protein(s)	Promoters
GATA	WGATAR	GATA-4	ANF, BNP, cTpC, cTpI, α-
		GATA-5	MHC, MLC3, α-CA
		GATA-6	
CARE	GCTGG	?	ANF, β-MHC, cTpC
NKE	TNAAGTG	Nkx2-5	ANF, BNP, MLC-2, cTNT, β-
		Nkx2-7	MHC,
		Nkx2-8	
CArG	CC(A/T)6GG	SRF	ANF, α-CA, α-MHC, skA, β-
		C2BP	MHC,
MEF2	(C/T)T(A/T)$_4$ATA(A/G)	MEF2A	MCK, Desmin, α-MHC, MLC-
		MEF2B	2, MLC1/3
		MEF2C	
		MEF2D	
M-CAT	CATTCCT	TEF1	BNP, cTpC, cTpI, cTNT,α-
			MHC, β-MHC
E-box	CANNTG	USF	MLC-2, α-MHC, αB-crystallin,
		eHAND	α-CA, MCK, MLC-2
		dHAND	

family of cardiac hormones, ANF, was cloned less than 15 years ago (Nemer *et al.*, 1984). Interestingly, recent studies have demonstrated that the ANF and BNP genes are localized within a 50 kb fragment on human chromosome 1, suggesting the presence of a cardiac natriuretic peptide locus (Arden *et al.*, 1995). Contrary to other cardiac genes, the expression of the ANF and BNP genes is restricted to cardiocytes and is not detectable in embryonic skeletal muscle (Dagnino *et al.*, 1991). Thus, these genes represent interesting markers to identify the molecular mechanisms regulating cardiac-specific transcription. Interestingly, transgenic experiments demonstrated that the first 500 bp of the rat ANF promoter were sufficient to reconstitute the spatial and temporal specificity of ANF (Seidman *et al.*, 1991; Field, 1988; Beaulieu *et al.*, 1984), a feature that greatly facilitates the study of this promoter since the major *cis*-elements required for its activity are confined in a relatively short sequence. In the case of the BNP promoter, transgenic studies established that the first 114 bp of the promoter are sufficient for high transgene expression in the heart (S. Bhalla and M. Nemer, unpublished results) and analysis of the promoter region of BNP revealed a structural organization reminiscent of erythroid promoters since GATA, NF-E2-like and CACC box motifs were present. This organization is totally conserved among many species, including the human, canine and rodent BNP genes and mutagenesis of this site underlined the crucial role of the GATA elements for BNP promoter activity (Grépin *et al.*, 1994). Interestingly, the GATA elements of the BNP promoter are organized in tandem, on the same side of the double-helix which is reminiscent of the GATA motifs organization in non-cardiac promoters. This tandem only binds one GATA-binding protein as tested in EMSA, but binds GATA proteins more efficiently than single sites (unpublished results). These GATA elements are active in cardiac myocytes where they do not display chamber-specific activity. In addition, the BNP promoter harbors a third GATA element at -30, for which a role in BNP promoter activity has not been assessed yet. However, there is the interesting possibility that since the BNP and ANF genes are organized in a locus, the -30 GATA site might be required, as for the β-globin gene, to establish a chromatin domain. Many important GATA elements have then been found in cardiac genes such as in ANF (Durocher *et al.*, 1996), cardiac troponin C (cTpC) (Ip *et al.*, 1994), cardiac troponin I (cTpI) (Murphy *et al.*, 1997), MLC3

(McGrew *et al.*, 1996), α-cardiac actin (R.J. Schwartz, personal communication) and α-MHC promoters (Molkentin *et al.*, 1994). Thus, the cardiac GATA elements are found in different classes of cardiac promoters and their widespread distribution implies that they play an important role in establishing or maintaining cardiac gene expression.

1.3.1.2. The CARE element.

The use of primary cultures of atrial or ventricular myocytes from late embryonic (18 dpc) or neonatal rats (day 1 or 5) enabled the identification of regulatory elements that are required for basal cardiac transcription, such as GATA elements. Interestingly, this system can be used to study transcription specific to different stages of development. The ANF gene is an excellent model to study chamber-specific transcription since it undergoes a perinatal decline of expression in the ventricle. The ANF gene is expressed constitutively in the atria but its expression in ventricles is regulated during development. In particular, its ventricular expression is considerably diminished starting the first postnatal week. In adulthood, ANF expression is almost considered atrial-specific (Argentin *et al.*, 1994). Interestingly, ANF mRNA or ANF promoter activity in cardiac myocytes derived from neonatal (1 day-old) or 4 day-old rats mimics the situation seen in the animal. Thus, it was possible to identify positive and negative elements involved in the atrial-specific activity of the ANF promoter.

In the embryonic myocardium (atria and ventricle), transfection studies established that the ANF promoter depends largely on the CARE (cardiac regulatory element) motif [formerly called MS or PE2 (McBride *et al.*, 1993)]. This element was first identified as a strong cardiac footprint over the ANF promoter between nucleotides -382 and -355. Deletion of this element revealed that it is required for high activity of the ANF promoter in 18 dpc cardiocyte cultures, and in atrial myocytes from post-natal cultures. Interestingly, its activity in ventricles declines with time in post-natal development and strikingly follows the decline of ANF expression in the ventricle. This element is able to confer strong transcriptional activation (10 to 50 fold) when oligomerized in three copies upstream of an heterologous promoter. The oligomerized CARE site is active in embryonic myocytes and postnatal atrial myocytes but not in ventricular myocytes of 5-day old neonates.

Mutagenesis studies revealed that the CARE element is a palindromic sequence containing the half site 5'-GCTGG-3'. This palindrome is conserved between species although the spacing is different in human and rodents. The differences in spacing does not seem to affect activity since the human CARE site is as active as the rat sequence in rat cardiocytes. The presence of the GCTGG box is detected in a variety of cardiac promoters such as the cTpC gene promoter where it is present in a direct repeat instead of a palindrome. EMSAs on whole-heart nuclear extracts revealed that a cardiac-specific complex binds the palindromic sequence. The abundance of this complex declines in the ventricle after birth, a result consistent with the activity of the element. Moreover, the banding pattern observed on gel shifts is consistent with the hypothesis that CAX proteins (CARE complex proteins) bind the CARE element as homo- or heterodimers. The identity of the nuclear protein(s) binding this element remains unknown but its identification will generate novel and significant insights for chamber specification in the myocardium and embryonic transcription. Thus, the characterization of this element suggests that at least one part of the atrial-specificity resides in the differential expression (or activity) of positive *trans*-regulators.

1.3.1.3. The CArG box.

Studies on the activity of ANF promoter elements in stage-specific cardiomyocyte cultures also led to the identification of a perfectly conserved element among the rat and human genes, the ANF CArG motif, that possesses an activity which is the mirror image of the CARE element in that it is not active in embryonic or post-natal atrial myocytes, but rather in 4 day-old ventricular myocytes (Argentin *et al.*, 1994). CArG motifs are found in many muscle promoters, such as MLC1/3 enhancer (Ernst *et al.*, 1991), cardiac actin (Mohun *et al.*, 1989; Taylor, 1991; Sartorelli *et al.*, 1990; Pari *et al.*, 1991) skeletal α-actin (Lee *et al.*, 1992), MCK (Vincent *et al.*, 1993) and MHCs (Thompson *et al.*, 1991), and are defined by the consensus $CC(A/T)_6GG$. They were first identified as muscle-specific elements but their close sequence homology with the SRE as well as the identification of the major CArG binding proteins (MAPF1 and MAPF2) as SRF and YY1 proteins, which are expressed ubiquitously, suggested that CArG boxes might not provide tissue-specific

activity (Gualberto *et al.*, 1992; Mohun *et al.*, 1991).

Interestingly, the work on the ANF CArG established that neither SRF or YY1 were the transcription factors responsible for its cardiac activity. Instead, binding studies identified a complex (called C2) that is distinct from SRF and YY1 (Boxer *et al.*, 1989; Argentin *et al.*, 1994). Moreover, this C2 complex can be observed in cardiac nuclear extracts from Xenopus and rodents although it is not clear yet if the C2 complex and the C2-like complex observed in *Xenopus* are identical even though they have similar mobility (Taylor, 1991; Argentin *et al.*, 1994). The ventricular-specific activity of the ANF CArG box is dependent on C2 binding since a discriminating mutation abolishing C2 formation (but not SRF or YY1 binding), abolished the cardiac-specific activity of the element (Argentin *et al.*, 1994). The identity of the proteins composing the C2 complex is not yet known but the characterization of the ANF CArG will certainly facilitate the cloning of this cardiac-enriched CArG-binding factor.

1.3.1.4. The NK2 response element (NKE).

The ANF promoter depends on approximately the first 150 bp for high cardiac activity in primary cardiomyocytes. In this region, apart from the GATA element that was discussed above, three other conserved elements or motifs are found: the AT-rich sequence, the *phenylephrine response element* (PERE) and a third one, located between these two, called the NKE (NK2 response element). Both the PERE and AT-rich sequence do not contribute significantly to ANF basal promoter activity. Deletion analyses revealed that the NKE, however, is essential for ANF promoter activity in cardiomyocytes [the identification of the NKE will be described in greater detail in chapter 2] (Durocher *et al.*, 1996). Interestingly, its cardiac activity is modulated in postnatal development where it does not contribute to promoter activity in ventricular myocytes of four day-old rats. This result is reminiscent of the activity of the CARE element but unlike CARE, this ventricular decline of activity does not depend on the element itself since it is still active, when placed upstream of a minimal promoter, in ventricular myocytes of four day-old rats. Rather, this chamber-specific activity is the consequence of a repressor activity located between the GATA element and the NKE (Durocher *et al.*, 1996). As will be discussed below, the NKE

binds members of the NK2 family of homeoproteins (comprising Nkx2-5) and its composition (GCAAGTGACAGAATGG) slightly diverges (bold residues) from the canonical NKE (TNAAGTG) or a preferentially selected TAAT core motif (CATAATTN) (2744). These two different boxes (underlined nucleotides) can each bind Nkx2-5 at a lower affinity than the consensus NKE. Interestingly, the NKE is found in many cardiac promoters such as BNP, MLC-2, β-MHC (Durocher *et al.*, 1996) and cTNT. Furthermore, in *Drosophila*, a consensus NKE was found to be required for the cardiac expression of the transcription factor D-Mef2 (Gajewski *et al.*, 1997). Thus, given its evolutionarily conserved importance, it is likely that the NKE, like the GATA motif, plays a widespread and versatile role in cardiac transcription.

1.3.1.5. The MEF2 element.

The MEF2 site was first identified during studies on skeletal muscle transcription. The MEF2 element is composed of an AT-rich sequence of the consensus (C/T)T(A/T)$_4$ATA(A/G), which is strikingly similar, although distinct, to the CArG motif. Deletion or mutation of this sequence decreases the activity of the MLC-2 (Navankasattusas *et al.*, 1992; Zhu *et al.*, 1991), α-MHC (Molkentin and Markham, 1993), MCK (Gossett *et al.*, 1989), MLC3 (McGrew *et al.*, 1996) and desmin promoters both in cardiac and skeletal muscle cells. Interestingly, it seems that MEF2 elements collaborate with cell-restricted elements to generate either skeletal or cardiac muscle activity. In skeletal muscle cells, MEF2 elements often collaborate with muscle specific E-boxes to generate strong skeletal muscle promoter activity (Molkentin *et al.*, 1995; Naidu *et al.*, 1995; Kaushal *et al.*, 1994). In cardiac muscle cells, studies on the transcriptional regulation of the MLC-2 promoter led to the identification of a composite 28 bp element (called HF-1) that is required and sufficient to restrict the expression of a transgene to the ventricle (Zhu *et al.*, 1991). Mutation analyses of this composite site revealed that two major positive *cis*-acting sequences compose HF-1 (Navankasattusas *et al.*, 1992). The first sub-element, HF-1a binds the ubiquitous protein YB-1 whereas the HF-1b sub-element binds specifically members of the MEF2 family. It seems that the cardiac specificity is dictated by the HF-1b moiety in collaboration with the putative cardiac-restricted cofactor protein (p30) which is

physically associated with YB-1. This proposed combinatorial interaction might give rise to the ventricular pattern of expression seen in transgenic mice. Nevertheless, the composition of the HF-1 element strongly suggests that MEF2 and AT-rich elements play important roles in cardiac transcription and it seems that *cis*-acting sequence interaction is a common theme in MEF2 site-dependent transcription.

1.3.1.6. The M-CAT element.

Promoter analyses of genes expressed in skeletal muscle revealed a sequence motif called M-CAT (muscle-CAT heptamer) required *in vitro* for high promoter activity both in cardiac and skeletal muscles. This element is characterized by the consensus CATTCCT and has been found on multiple promoters such as cTpC, cTNT, BNP, α-MHC, skeletal α-actin and β-MHC (Parmacek *et al.*, 1992; Karns *et al.*, 1995; Thompson *et al.*, 1991; Mar and Ordahl, 1988; Molkentin and Markham, 1994; Gupta *et al.*, 1994). However, *in vivo* studies on the α-MHC promoter in transgenic mice revealed that the mutation of both M-CAT motifs was not sufficient to abolish the transgene activity (Subramaniam *et al.*, 1991). Rather the generation of a triple mutation disrupting both M-CAT motifs plus a presumptive SP1 site abolished transgene activity. Since the single SP1 binding site mutation does not affect β-MHC promoter activity *in vivo*, it was proposed that M-CAT elements work in a combinatorial fashion. This situation is likely to occur in other promoters such as in the α-MHC gene or the BNP promoter where both M-CAT sites are in close relationship with either an E-box (Gupta *et al.*, 1994) or a GATA element (Thuerauf and Glembotski, 1997) respectively. The requirement of other *cis*-elements for restricting M-CAT activity to muscle cell types is supported by mutational analyses of the flanking sequences surrounding the M-CAT motif. This showed that motifs such as an imperfect NKE/E-box restrict the expression of the M-CAT in striated muscle cells. Thus, it seems that the M-CAT element harbors the properties of a number of cardiac elements described so far i.e. that they act mainly in collaboration with other regulatory pathways.

1.3.1.7. The E-box.

The discovery of the MyoD family and its importance in skeletal myogenesis

sparked a lot of interest for the study of bHLH proteins in the heart. However, the targeted disruption of several myogenic regulators did not generate cardiac abnormalities and heterokaryon experiments discarded the possibility that a dominant transcriptional regulator was acting like MyoD in cardiac cells (Evans *et al.*, 1994). bHLH proteins recognize a sequence motif on DNA called the E-box. The E-box is defined by the consensus CANNTG and the identity of the N residues depend on the composition of the bHLH dimer. Interestingly, several *in vitro* and *in vivo* studies have pinpointed the importance of E-boxes in cardiac transcription. For instance, the MLC-2 (Navankasattusas *et al.*, 1994), α-MHC (Molkentin *et al.*, 1993), αB-crystallin (Gopal-Srivastava *et al.*, 1995), α-cardiac actin (Moss *et al.*, 1994) and MCK (Shield *et al.*, 1996) gene promoters harbor important cardiac E-boxes. Most of the work focused on transfection assays or direct injection in cardiac muscle cells. However, the work of the Hauschka laboratory assessed the role of the MCK E-boxes in transgenic mice and provided convincing evidence for an *in vivo* role of E-boxes in cardiac transcription (Shield *et al.*, 1996). In this study, the three E-boxes present on the promoter were mutated collectively and several independent transgenic animals were generated. To their surprise, the abolition of three E-boxes had only mild effects in body and limb muscle expression of the transgene whereas it had a drastic effect on the cardiac expression (Shield *et al.*, 1996). This work suggest that E-boxes participate in cardiac-specific transcription. However, it is still unclear if the proteins binding to these boxes are bHLH proteins. Several cardiac proteins could have overlapping binding sites that could be mistaken for E-boxes. This is the case for Nkx2-5 which harbors the E-box CAAGTG in its high affinity binding site (Chen and Schwartz, 1995; Durocher *et al.*, 1996).

1.3.2. Cardiac transcription factors.

As reviewed in the previous sections, cardiogenesis is a complex developmental process that requires commitment of splanchnic mesodermal cells to the cardiac lineage, activation of the cardiac genetic program, and a complex morphological elaboration of the primitive heart tube into the four-chambered, mature heart. The establishment and the maintenance of the cardiac phenotype requires the activation of cardiac-specific as well as

muscle-restricted gene expression. This process is presumably governed by the combinatorial action of cell-restricted as well as ubiquitous transcriptional regulators. The identification of such cell-restricted regulators and the understanding of their interactions is of paramount importance in the study of cardiac development. Currently, very few cardiac transcription factors have been found to be enriched in the myocardium. This section is dedicated to their structural composition, expression patterns and transcriptional activities.

1.3.2.1. The GATA family

The vertebrate GATA family comprises six members (GATA-1 to -6) characterized by their DNA-binding domain, which is organized in two non-canonical zinc-fingers of the consensus: Cys-x-Asn-Cys-$(x)_{17}$-Cys-Asn-x-Cys (Evans and Felsenfeld, 1989; Tsai *et al.*, 1989). This signature is an ancient protein domain that is present in gene regulators from yeast, *Neurospora crassa*, *Drosophila*, and *C. elegans*. The Zn-finger domain of GATA proteins is highly conserved across species and mutation as well as deletion analyses has revealed that the C-terminal Zn-finger of GATA proteins is responsible for the recognition of the GATA element (Martin *et al.*, 1990; Yang and Evans, 1992; Omichinski *et al.*, 1993b). The exact role of the N-terminal finger is still debated but seems to be required either to stabilize the binding of the protein to DNA (Yang and Evans, 1992), to increase its DNA-binding specificity (Trainor *et al.*, 1996; Whyatt *et al.*, 1993) or to serve as a protein-protein interaction interface (Weiss *et al.*, 1997). Outside the zinc-finger domain, much less sequence conservation is observed although, for each member, the N- and C-terminal flanking domains are more conserved across orthologues than across homologues within the same species. The high level of homology within the DNA-binding domain explains the very modest differences in DNA-binding capacities among the GATA family; even when extensive site selection was used (Ko and Engel, 1993; Merika and Orkin, 1993; Whyatt *et al.*, 1993).

This very modest variation of the optimal target sites for GATA proteins is explained by the resolution of the structure of the 59 aa C-terminal Zn-finger of the chicken GATA-1 protein complexed with its binding site (Omichinski *et al.*, 1993a). The C-terminal Zn-finger (and most probably the N-terminal Zn-finger as well) is arranged in two

subdomains: the protein core (aa 2-51) which consists of the zinc coordination site; and the C-terminal tail (aa 52-59). The protein core is composed mainly of two short, irregular antiparrallel β sheets (aa 6-27) followed by a short α-helix (aa 28-38) and a long loop (aa 39-51) (Omichinski *et al.*, 1993a). The core domain is structured around the zinc atom which is tetrahedrally coordinated to the sulfur atom of Cys^{7}, Cys^{10}, Cys^{28} and Cys^{31} (Omichinski *et al.*, 1993a). This structure enables the side chain residues of the alpha helix, and the loop connecting the first 2 β-strands, to make specific contacts with the bases in the major groove. Interestingly, there is an almost absolute conservation of the residues involved in specific DNA-binding among all GATA members and when a substitution is observed, it is a conservative one (Omichinski *et al.*, 1993a). Thus, it is not surprising to observe a near identical, optimal binding site for all GATA proteins.

Despite their almost identical binding sites, the expression patterns of the GATA proteins suggest that they possess unique functions. GATA proteins can be divided in two subgroups based on tissue distribution and sequence homology. The first subgroup comprises GATA-1, -2 and -3 which are predominantly expressed in hematopoietic cells. GATA-1 is restricted to the erythroid mast cell and megakaryotic lineage; GATA-2 transcripts are predominantly found in hematopoietic stem cells and GATA-3 mRNA is mainly found in T lymphocytes. However, all three « hematopoietic » GATA transcripts can be detected in a variety of other tissues such as the gonads (GATA-1), the embryonic brain and endothelial cells (GATA-2) and central and peripheral nervous system as well as the kidney and the adrenal gland (GATA-3) [reviewed in (Simon, 1995)]. The conservation of the expression patterns across species and the higher degree of conservation among orthologues suggest that each GATA factor has conserved properties and roles during development. In order to understand the unique roles of hematopoietic GATA proteins in development targeted disruption of their respective genes was carried out. The gene knockout of the *gata1* allele on the X chromosome in male (XY) ES cells enabled to study, without germ line transmission (Pevny *et al.*, 1991), the contribution of GATA-1-deficient cells to the developing tissues. All non-hematopoietic tissues showed an important contribution of $GATA^{-}$ cells whereas GATA-1^{-} cells failed to populate the differentiated red blood cell population and thus showed no chimerism (Pevny *et al.*, 1991). This

demonstrated that GATA-1 is required for erythropoiesis. *In vitro*, analysis of hematopoietic differentiation revealed that yolk sac (primitive) erythropoiesis is initiated but the erythroid progenitors die by apoptosis at the proerythroblast stage (Weiss and Orkin, 1995; Pevny *et al.*, 1995). The upregulation of GATA-2 expression is seen in early erythroid cells, which might perhaps explain why the erythroid progenitors can initiate erythropoiesis (Weiss *et al.*, 1994). Interestingly, in the mast cell and megakaryocytic lineages, where both GATA-2 and GATA-1 are expressed (Linask and Lash, 1993; Pevny *et al.*, 1995), the upregulation of GATA-2 expression is sufficient to produce mature mast cells and platelets (Pevny *et al.*, 1995; Fujiwara *et al.*, 1996). The recent transmission to the germ line of the disrupted *gata1* allele (Fujiwara *et al.*, 1996) confirmed that the major, non-redundant, role of GATA-1 is in definitive erythropoiesis. For GATA-2, the gene disruption experiment confirmed the earlier proposition (Briegel *et al.*, 1993) that GATA-2 plays an important role in the proliferation of stem cell or early hematopoietic progenitors, as a severe deficit in the number of hematopoietic progenitors was observed (Tsai *et al.*, 1994). Finally, disruption of the GATA-3 alleles leads to severe abnormalities in tissues where GATA-3 is normally expressed during embryogenesis such as in the central nervous system and during fetal liver hematopoiesis and T cell formation (Pandolfi *et al.*, 1995; Ting *et al.*, 1996). Collectively, these studies on the hematopoietic GATA proteins indicate that they each play unique and non-redundant roles during development.

However, how this uniqueness is acquired at the molecular level remains to be answered. One hypothesis would be that specificity among GATA proteins is acquired by differential interactions with nuclear proteins. Indeed, GATA-1, -2 and -3 have been shown to interact with a variety of proteins. GATA-1, -2 or -3 interact with other Zn-finger proteins such as the estrogen receptor (Blobel *et al.*, 1995a), the Kruppel-related protein EKLF and SP1 (Merika and Orkin, 1995) as well as the leucine zipper proteins Fos and Jun (Kawana *et al.*, 1995) and the LIM protein rbtn2 (Osada *et al.*, 1995). Recently, an hematopoietic cofactor for GATA-1, FOG (friend of GATA) has been cloned (Tsang *et al.*, 1997). FOG is a multi-Zn-finger protein that binds specifically to the N-terminal Zn-finger of GATA-1 -2 and perhaps -3. FOG acts as a transcriptional cofactor of GATA proteins and potentiates erythroid differentiation *in vitro*. However, because it interacts equally well

with both GATA-1 and -3, FOG cannot explain the specificity among GATA members and specificity cofactors for hematopoietic GATAs remain to be identified.

GATA-1, -2 and -3 are not expressed in the heart. The identification of the GATA motif as an important determinant of both ANF and BNP promoter activities in cultured cardiocytes led to the possibility that additional GATA factors were present in cardiac cells. Moreover, gel shift analyses confirmed, at that time, that a non-hematopoietic GATA factor was present in cardiomyocytes. A low-stringency screen of either whole-embryo or cardiomyocyte cDNA libraries with the zinc finger of GATA-1 led to the identification of the GATA-4 protein (Arceci *et al.*, 1993; Grépin *et al.*, 1994). GATA-4 is a 443 amino acid protein that has a primary structure that is very similar to that of other GATA proteins. GATA-4 possesses two class V Zn-fingers which bear a high degree of homology with those of the hematopoietic GATAs. Northern blot and *in situ* hybridization analyses revealed that GATA-4 expression is restricted to the heart for most of embryogenesis and is later expressed in the developing gonads (Heikinheimo *et al.*, 1994; Grépin *et al.*, 1994; Arceci *et al.*, 1993). GATA-4 can be detected in the precardiac mesoderm as early as 7.0-7.5 dpc at the late primitive streak stage where it precedes the expression of the earliest cardiac differentiation markers such as contractile protein genes and natriuretic peptide genes. In the precardiac mesoderm, GATA-4 expression is confined in the cardiogenic crescents, on each side of the embryo. This expression pattern is very similar to that of Nkx2-5 and coincides with the heart-forming region in mouse, chicken and *Xenopus* (Laverriere *et al.*, 1994; Jiang and Evans, 1996; Heikinheimo *et al.*, 1994). At later stages, GATA-4 expression is detected throughout the myocardium and endocardium where it remains expressed up to adulthood (Heikinheimo *et al.*, 1994; Morrisey *et al.*, 1997b; Kuo *et al.*, 1997). Thus, GATA-4 is an early marker of the cardiac cell fate and might represent an early regulator of cardiogenesis.

Transfection studies in non cardiac cells established that GATA-4 is a potent transactivator of GATA-containing promoters. In fact, numerous cardiac [ANF (Durocher *et al.*, 1997a), BNP (Grépin *et al.*, 1994), cTpC (Ip *et al.*, 1994), cTpI (Murphy *et al.*, 1997), α-MHC (Molkentin *et al.*, 1994)] and non-cardiac [PDGF-α receptor (Wang and Song, 1996), IL-5 (Yamagata *et al.*, 1995) and J6 (Bielinska and Wilson, 1995)] promoters

are transactivated *in vitro* by GATA-4. Structure-function studies established that GATA-4 possesses at least three transcriptional activation domains. Two of these are located in the N-terminal region whereas at least one additional transactivation domain is located in the C-terminus of the protein (Durocher *et al.*, 1997a; Morrisey *et al.*, 1997a). Interestingly, these activation domains do not seem to function equivalently in all cell types. For instance, the N-terminal activation domains are not required in P19 cells whereas it is absolutely required in HeLa cells (unpublished observation). This observation suggests that GATA-4 requires transcriptional cofactors. Thus, since the cardiac promoters which are targets for GATA-4 are part of every aspect of cardiac physiology and differentiation, GATA-4 is likely to be a master regulator of cardiac transcription and development.

Two other members of the GATA family have been found to be expressed in the heart: GATA-5 (Morrisey *et al.*, 1997b) and GATA-6 (Morrisey *et al.*, 1996). GATA-5 and GATA-6 have very significant homologies with GATA-4, inside (87 to 83%) and outside the Zn-finger region (41 to 52%). They are more closely related to each other than to GATA-1, -2, -3 and thus constitute the second subgroup of the GATA family. Expression studies in *Xenopus*, chick and mouse have been done and despite significant species-specific differences in their expression patterns, these reports support the notion that GATA-4, -5 and -6 represent a family of closely related transcription factors involved in heart and gut development (Laverriere *et al.*, 1994; Jiang and Evans, 1996; Morrisey *et al.*, 1996; Morrisey *et al.*, 1997b). GATA-6 is expressed in the precardiac mesoderm at the late primitive streak stage (Morrisey *et al.*, 1996; Gove *et al.*, 1997; Jiang and Evans, 1996) but is also expressed in other tissues such as the vascular smooth muscle of the large arteries, the developing gastro-intestinal system and the urogenital ridge. On the other hand, GATA-5 is also expressed in the precardiac mesoderm at the late primitive streak stage as well as in gastro-intestinal structures later on. However, at midgestation (12.5 dpc) GATA-5 transcripts are detected mainly in the atria where they seem to be restricted to the endocardium. Soon after (16.5 dpc), GATA-5 expression is not detected in the developing heart. Thus, the « cardiac » GATA members have overlapping but distinct spatial and temporal expression patterns during cardiogenesis. Like GATA-4, GATA-5 and -6 are transcriptional activators of GATA-containing promoters (Morrisey *et al.*, 1996; Morrisey

et al., 1997b). Interestingly, the activation domains of GATA-6 seem to be similar to those of GATA-4. Thus, the modest differences in the expression patterns and the activation properties of cardiac GATA proteins imply that they are likely to have mostly redundant roles during cardiogenesis.

In order to assess the role of GATA proteins in cardiogenesis, gene inactivation and gain-of-function studies were carried out *in vitro and in vivo. In vitro,* GATA-4 expression was knocked-down by an antisense strategy in P19 teratocarcinoma cells which provide a cellular model of inducible cardiac differentiation. In GATA-4$^-$ lines, DMSO-treated aggregates could not achieve terminal cardiac differentiation, and massive apoptosis was observed (Grépin *et al.*, 1995b; Grépin *et al.*, 1997). This phenotype was not a consequence of a pleiotropic defect of differentiation since neuronal and muscular (skeletal) differentiation pathways were unaffected. Moreover, GATA-4 is not required for mesodermal differentiation since expression of mesodermal markers such as Brachyury were detected and persisted in GATA-4$^-$ P19 lines. Furthermore, cardiac differentiation was initiated since early markers of the cardiac fate such as Nkx2-5 and MEF2C were expressed. Recently, two GATA-4 gene knockouts have been generated in mice and the results obtained confirm that GATA-4 is essential for cardiogenesis (Kuo *et al.*, 1997; Molkentin *et al.*, 1997). Mice lacking GATA-4 die in utero as a consequence of the failure of GATA-4$^{-/-}$ mice to form a linear heart tube. However, the cardiac defect does not seem to be cell-autonomous since chimeric mice containing GATA-4$^{-/-}$ cells have normal hearts and these hearts are populated by GATA-4$^{-/-}$ cells (Kuo *et al.*, 1997). The *cardia bifida* phenotype is explained by the absence of ventral closure of the embryo resulting in the absence of morphogenetic movements required for the fusion of the bilateral primordia. Thus, the cardiac phenotype is not caused by a lack of terminal differentiation as predicted from the studies in P19 cells. This major difference implies that GATA-4 function in terminal differentiation *in vivo* might be compensated by other proteins and/or signals in the mouse. One candidate for this compensation is GATA-6 (and GATA-5 to a lesser extent). GATA-6 transcript levels in GATA-4$^{-/-}$ mice are massively upregulated (Kuo *et al.*, 1997; Molkentin *et al.*, 1997), but not in the P19 model (Bronchain and Nemer, unpublished results) suggesting that GATA-6 is partially redundant with GATA-4.

Complementary gain-of-function studies revealed another aspect of GATA factor function during differentiation. Experiments carried out in P19 cells, where GATA-4 was stably overexpressed, revealed that GATA-4 markedly enhanced cardiogenesis as evidenced by the earlier appearance of beating cardiac cells. Moreover, GATA-4 can even promote cardiac cell differentiation without the addition of an inducer prior to aggregation (Grépin *et al.*, 1997). These results which are remarkably similar to those obtained in *Xenopus* and zebrafish when Nkx2-5 (see below) was overexpressed (Cleaver *et al.*, 1996; Chen and Fishman, 1996), suggest that GATA-4 can potentiate cardiogenesis by augmenting the potency of the cardiogenic field. Furthermore, in *Xenopus laevis*, gain-of-function studies revealed that GATA-6 could also be a regulator of the proliferation of the cardiogenic field (Gove *et al.*, 1997). Injection of GATA-6 mRNA in gastrulating embryos resulted in transient block of cardiac differentiation. This block led to an expansion of the cardiogenic field and upon the decrease of GATA-6 mRNA levels, the precardiac mesoderm escaped the block to differentiation and a larger number of cardiomyocytes resumed differentiation to generate an enlarged heart. Interestingly, this phenotype is reminiscent of the phenotype observed when GATA-2 is expressed in an erythroid cell line (Briegel *et al.*, 1993), and suggests that GATA-6 might be a regulator of precardiac mesoderm proliferation. Collectively, these genetic manipulations demonstrate the crucial roles of GATA factors in cardiac transcription and development but do not explain the non-redundant functions of these two factors during cardiogenesis.

1.3.2.2. NK2 homeoproteins

Homeodomain-containing proteins are transcriptional regulators that exert important roles in development, especially in determining cell fate and positional identity. The homeodomain is typically a 60 aa domain which is shared by more than 350 genes and is present, so far, in all metazoans (Gehring *et al.*, 1994a). It was first discovered in genes causing homeotic transformation in *Drosophila* and was shown to be either a DNA- or RNA-binding domain. The general signature of the homeodomain comprises almost invariant residues at position 5 (Arg), 16 (Leu), 20 (Phe), 48 (Trp), 49 (Phe), 51 (Asn) and 53 (Arg) whereas another 10 aa are conserved in more than 80% of the homeodomains and

12 additional positions share two aa in more than 80% of cases (e.g. Arg or Lys at position 2) (Gehring *et al.*, 1994a; Scott *et al.*, 1989). The majority of homeodomains recognize DNA sequences containing a 5'-TAAT-3' core motif (Gehring *et al.*, 1994b) and the structural basis of such similar target sites is explained by the resolution of the tertiary structure of homeodomains complexed to DNA. So far, structures of DNA-bound homeodomain has been solved for the MATa1/MATα2 (Li *et al.*, 1995), paired (Halaban *et al.*, 1987), engrailed (Kissinger *et al.*, 1990), vnd/NK2 (Gruschus *et al.*, 1997) and the Antp (Otting *et al.*, 1990) protein. Interestingly, their structure is strikingly similar and consists of three helical regions folded in a tightly globular domain. Helix II is preceded by a flexible N-terminal arm and separated from helix III by a loop (see Fig. 1.8). This motif is similar to the helix-turn-helix motif found in prokaryotic gene regulators and implies that this structural domain has been widely conserved during evolution. The recognition helix (helix III/IV) rests in the major groove of the DNA recognition site and is connected to the helix-turn-helix motif by a loop and the bulk of the specific interactions between the homeodomain and DNA are made by helix III/IV, although additional specific contacts are made in the minor groove by the N-terminal arm (Otting *et al.*, 1990; Gruschus *et al.*, 1997).

Genetic studies in *Drosophila* and mice established that most of the homeoproteins have either totally different or only partially redundant functions even if coexpressed in the same cell (Corsetti *et al.*, 1992; Dessain *et al.*, 1992). This result might be somewhat surprising given the almost identical DNA-binding-specificity for most of the homeoproteins. Thus, in order to explain this paradox, it has been postulated for some time that the specificity of homeodomain action is mediated via cofactors. Indeed, recently, a mechanism for imparting specificity among HOX class homeoproteins has been characterized. This mechanism relies on the requirement of cofactors that selectively augment the DNA-binding specificity of HOX proteins by dimerization. This was first demonstrated in *Drosophila* by a combination of elegant genetics and biochemistry. In these studies, the homeoprotein extradenticle (EXD), an « atypical » class homeodomain, was shown to be a selectivity cofactor for the HOX proteins Ultrabithorax (UBX) and labial (LAB) (Van Dijk and Murre, 1994; Chan *et al.*, 1994). EXD extends the DNA-binding site

of HOX proteins by physically binding to them via the YPWM motif that is present in a subset of HOX proteins. This motif, also called the hexapeptide, is required but not sufficient for the binding of EXD to HOX proteins, and specific residues located in the N-terminal arm of the homeodomain are also required for the interaction. Furthermore, this interaction modifies the conformation of the HOX proteins and enables it to bind DNA more efficiently (Chan *et al.*, 1996). This interaction has been conserved throughout evolution and the vertebrate homologues of *Drosophila* UBX, LAB (HOX) and EXD (PBX) act as cofactors to regulate transcription (Phelan *et al.*, 1995; Knoepfler and Kamps, 1995; Peltenburg and Murre, 1997). Interestingly, two recent reports established that the *fushi-tarazu* (Ftz) homeoprotein uses transcriptional cofactors in a fashion similar to HOX proteins. Using a combination of genetic and molecular approaches, the Ftz-F1 Zn-finger protein was shown to physically and genetically interact with Ftz in order to increase its DNA-binding activity (Yu *et al.*, 1997; Guichet *et al.*, 1997). Thus, homeodomains use a variety of protein-protein contacts to modulate their transcriptional activities and to generate distinct protein complexes having unique targets.

The importance of homeobox-containing genes in cardiogenesis was first established by the molecular cloning of the cardiogenic gene *tinman (msh-2, NK-4). Tinman* is part of the rapidly growing family of NK2 homeobox genes. NK2 homeoproteins have been found in practically all metazoans examined, such as planarians, leech, *C. elegans*, *Drosophila*, and vertebrates [reviewed in (Harvey, 1996)]. However, they have not yet been observed in any unicellular organisms, implying that NK2 proteins might be involved in morphogenetic or organogenetic processes. In flies, three members of the family have been cloned and characterized. The NK-2 protein (also known as vnd) is required for the differentiation of CNS neurons whereas *bap* (NK-3) and *tin* (NK-4) are involved in mesoderm segmentation (Bodmer, 1993; Azpiazu and Frasch, 1993; Jimenez *et al.*, 1995). In vertebrates, the NK2 family currently comprises eight members so far: Nkx2-1/TTF-1 (Guazzi *et al.*, 1990), Nkx2-2 (Price *et al.*, 1992), Nkx2-3 (Buchberger *et al.*, 1996), Nkx2-5 (Lints *et al.*, 1993), Nkx2-6 (Lints *et al.*, 1993), Nkx2-7 (Lee *et al.*, 1996), Nkx2-8 (McKnight *et al.*, 1988), and Nkx3-1/Xbap (Newman *et al.*, 1997). They are all expressed in a tissue-restricted fashion where they are thought to participate in

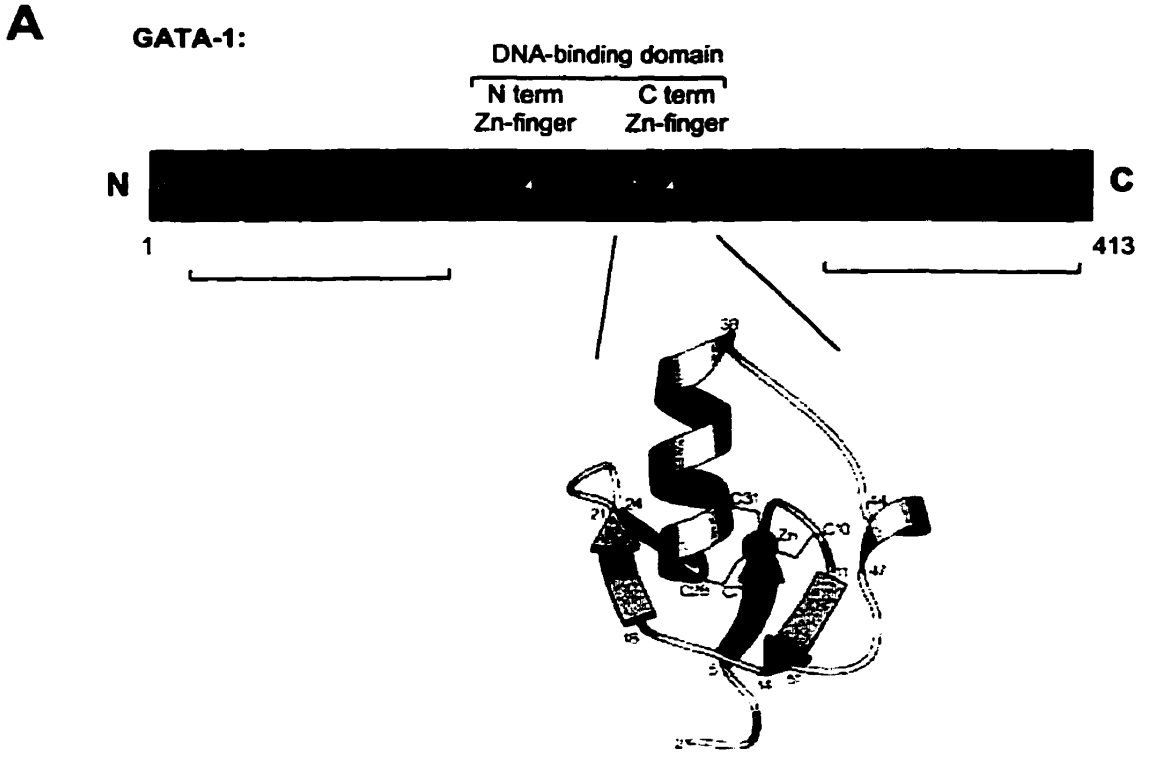
organogenetic processes. The NK2 homeodomain is characterized by the presence of a tyrosine at position 54. Structural studies have revealed that residue 54 plays a role in DNA recognition of either synthetic (Chen and Schwartz, 1995; Tsao *et al.*, 1995; Gruschus *et al.*, 1997) or natural (Durocher *et al.*, 1996; Gajewski and Thompson, 1996; Gajewski *et al.*, 1997) target sites. These studies established that the NK2 recognition site differs from the usual TAAT core motif and is characterized by the consensus TNAAGTG. Interestingly, mutation of Tyr54 into a Met residue, which is found in Antp at position 54, is sufficient to convert the NK2 binding site into an Antp binding site (Damante *et al.*, 1996). Thus, Tyr54 is a sequence and functional determinant for the NK2 family. The presence of Tyr54 in the homeodomain is not the sole determinant of the NK2 class. NK2 proteins can be classified in three classes depending on the presence of two conserved domains outside the homeodomain (Harvey, 1996). Most of the NK2 homeoproteins contain both the conserved eh1 domain as well as the NK2-SD motif and are referred to as Type I. Type II NK2 homeoproteins are defined by the presence of only the eh1 domain (such as *Tinman*) and Type III represents NK2 homeoproteins, usually from more primitive species such as CEH-22 from *C. elegans* or Dth-1 and -2 from flatworms, that contain neither of the 2 motifs. The eh1 domain was first referred as the TN domain but recent work established that this domain is also conserved in a variety of homeoproteins such as in engrailed, MSX and NK1 families. This domain probably mediates protein-protein interactions and was defined as a repression domain in engrailed (Wang *et al.*, 1996). Despite its characterization in engrailed, it is still unclear if this domain plays a similar role in NK2 proteins. However, the NK2-SD was detected thus far only in vertebrates NK2 proteins. This domain possesses an hydrophobic core with a valine or an isoleucine at every two positions [consensus: (V/I)x(V/I)xVLVR] and it is thought to be a protein-protein interface.

In the developing myocardium, four Type I NK2 proteins (Nkx2-3, Nkx2-5, Nkx2-7 and Nkx2-8) are expressed in the cardiogenic field, before the fusion of the cardiac primordia. They constitute a so-called cardiac NK2 family (Tonissen *et al.*, 1994; Lints *et al.*, 1993; Evans *et al.*, 1995; Chen and Fishman, 1996; Lee *et al.*, 1996).

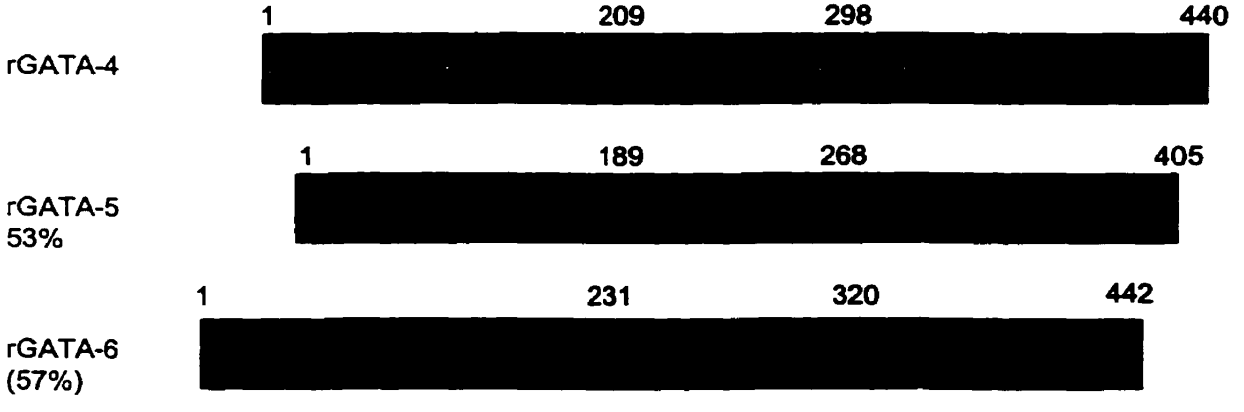
Figure 1.7.

(**A**) Schematic representation of a GATA protein (in this case GATA-1). A typical vertebrate GATA protein is composed of a DNA-binding domain consisting of two zinc-finger motifs. The N- and C-terminal domains harbor activation domains. The 3-dimensional structure of the GATA-1 C-terminal Zn-finger was taken from (Omichinski *et al.*, 1993a). (**B**) The cardiac GATA subfamily. Homologies between GATA-4, -5 and -6 were calculated with the GAP software of the GCG package.

A


GATA-1:
DNA-binding domain
N term
Zn-finger
C term
Zn-finger
N
C
1
413

B


1
209
298
440
rGATA-4
1
189
268
405
rGATA-5
53%
1
231
320
442
rGATA-6
(57%)

Type I NK2 homeoprotein:

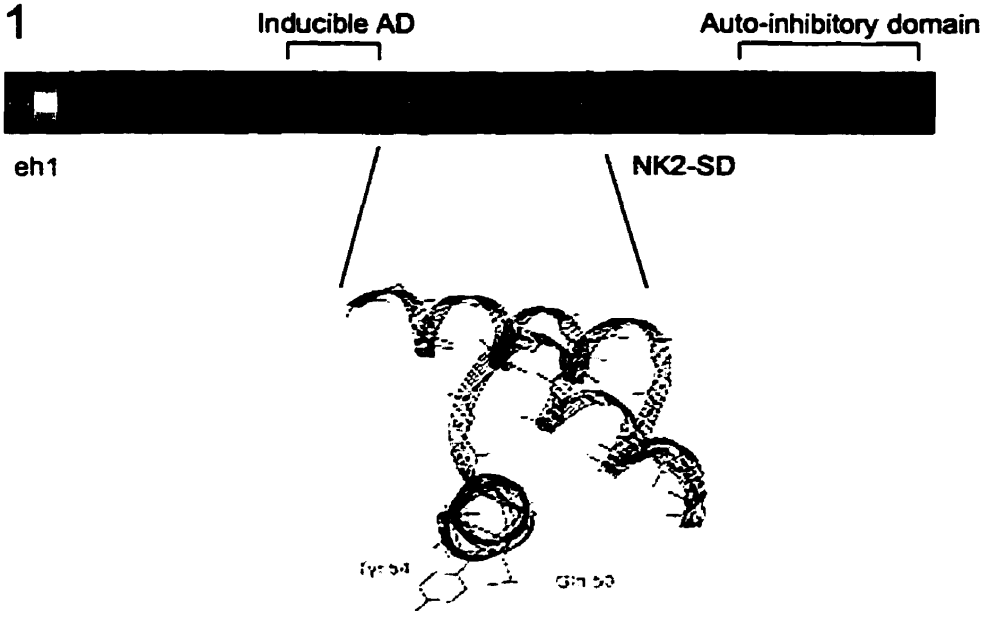


FIGURE 1.8

Schematic representation of a type I NK2 protein. A typical Type I NK2 homeoprotein contains a NK2-type homeodomain characterized by a tyrosine at position 54 and two protein motifs, the eh1 domain and the NK2-SD domains that are likely to participate in protein-protein interactions. The activation domains are located in N-terminal of the HD whereas an auto-inhibitory domain is located in the C-terminal. The 3-dimensional model of the NK2 homeodomain was taken from (Damante *et al.*, 1996).

They have distinct temporal and spatial expression patterns during cardiac development and are also expressed in other embryonic structures such as the branchial arches, the pharyngeal endoderm (Lints *et al.*, 1993; Lee *et al.*, 1996). Furthermore, some species to species differences in expression patterns can be observed. For instance, *Xenopus* and avian Nkx2-3 have been found to be expressed in the heart (Evans *et al.*, 1995; Cleaver *et al.*, 1996) whereas murine Nkx2-3 is not detected in the developing myocardium (Pabst *et al.*, 1997). These overlapping but distinct expression patterns suggest that cardiac NK2 proteins have partially redundant functions in cardiogenesis. In mammals, Nkx2-5 is thus far the only member characterized. It is a modest transcriptional activator that activates promoters harboring NKE sites, such as the ANF promoter (Chapter 3) and structure-function analyses of this protein suggest that it may also behave as a repressor depending on the promoter environment, post-translational modifications or protein-protein interactions (Chen and Schwartz, 1995). Additionally, in the case of the α-cardiac actin promoter, Nkx2-5 can be recruited to the SRE-bound SRF through protein-protein interactions (Chen and Schwartz, 1996). Thus, the expression pattern and the transcriptional activities of Nkx2-5 suggest that it might be an important regulator of the cardiac developmental and gene expression programs.

The importance of Nkx2-5 in cardiogenesis was assessed by inactivation of its alleles (Lyons *et al.*, 1995). Nkx2-5 null mice die at day 9.5 dpc due to both myogenic and morphogenetic defects. This phenotype is strikingly different from what is observed in *Drosophila* when the *tin* gene is mutated (an absence of myocardial differentiation) and it is thought to be the result of genetic redundancy among cardiac NK2 members. Nevertheless, Nkx2-5 gene disruption underlines the unique functions of Nkx2-5 in cardiogenesis. The major cardiac defect observed is an arrest at the looping morphogenesis stage. At around 9.0 dpc, mutant hearts rarely develop beyond the heart tube stage. Trabeculation and formation of the endocardial cushions were also completely (Lyons *et al.*, 1995). However, it is important to underline, as stated in a recent review by the author of the Nkx2-5 knockout, that heart morphogenesis might depend on its normal function and that the defects observed might depend on a hidden myogenic defect (Harvey, 1996). This is exemplified by the Scl/tal mutation where a heart morphogenetic defect is observed as

consequence of a deficit in embryonic blood circulation (Robb *et al.*, 1995). However, a recent report from Richard Harvey's group reinforced the possibility that the looping defect observed in Nkx2-5$^{-/-}$ embryos might be a true heart morphogenetic defect. In this study, the expression of two bHLH proteins, dHAND and eHAND, which are also thought to play a role in heart looping [see below and (Srivastava *et al.*, 1995; Srivastava *et al.*, 1997)] were carefully studied. It was found that eHAND expression was almost exclusively left-sided in the looping heart. Interestingly, in Nkx2-5$^{-/-}$ mice, this left-sided expression was abrogated whereas it remained unaffected in the other domains of eHAND expression (Biben and Harvey, 1997). Moreover, this extinction of eHAND expression in the future left ventricle and atrium did not occur in Scl$^{-/-}$ mice indicating that this defect is specific for Nkx2-5. Interestingly, Nkx2-5 expression is not itself asymmetric, which suggests that it is its transcriptional activity that is modified by laterality signals. Interestingly, in the Nkx2-5 gene knockout, only a subset of cardiac genes are mildly or severely affected in Nkx2-5-deficient mice and these comprise eHAND (Biben and Harvey, 1997), MLC-2 (Lyons *et al.*, 1995) and ANF (R.P. Harvey, personal communication). ANF expression is completely shut off in Nkx2-5-deficient embryos, which suggests that ANF is a direct target of Nkx2-5.). Since the ANF promoter is thus far the sole cardiac promoter known to be directly bound and regulated by Nkx2-5, it represents a unique tool to study Nkx2-5-dependent transcription.

Gain-of-function experiments demonstrated that Nkx2-5 was not sufficient to induce cardiogenesis although its overexpression in gastrulating zebrafish and *Xenopus* embryos resulted in ventralization and myocardial hyperplasia (Chen and Fishman, 1996; Cleaver *et al.*, 1996; Fu and Izumo, 1995). These results are also similar to those of the GATA-4 gain-of-function experiment in P19 cells and suggest that Nkx2-5 is also a regulator of cardiogenic field. Thus, since GATA-4 and Nkx2-5 can only accelerate cardiogenesis in committed cells one can interpret that both need at least another signal (or cofactor) from the committed cells to exert their respective roles during cardiogenesis. Interestingly, recent transcriptional studies in *Drosophila* proposed that *Tinman* could also be using a cofactor in order to regulate the cardiac transcriptional activity of the D-Mef2 promoter (Gajewski *et al.*, 1997). Therefore, the extremely similar phenotypes observed in GATA-4

or Nkx2-5 gain-of-function studies, in addition to the proposition that cofactors are needed for NK2 proteins, led our laboratory to postulate that GATA-4 and Nkx2-5 could be mutual transcriptional cofactors.

1.3.2.3. MADS box proteins

The major proteins that bind the MEF2 and CArG elements (MEF2 and SRF) are part of the MADS box family. The MADS box is a 56 aa motif that is found within the DNA-binding domain of numerous eukaryotic activators. The name MADS box derives from the four founding members of the family: MCM1, Agamous, Deficiens and SRF (Shore and Sharrocks, 1995). Sequence analyses reveal that the N-terminus of the MADS domain is predominantly composed of hydrophilic residues with a high proportion of basic aa whereas the C-terminal region is characterized by the presence of hydrophobic residues. Members of the MADS box family have important roles in both developmental and homeostatic transcriptional regulation. The Agamous and Deficiens genes are critical for flower development and the MCM1 protein is necessary for pheromone response in yeast (Shore and Sharrocks, 1995). The SRF gene was first identified as the serum response factor, a protein required for growth factor induction of promoters such as c-fos and for the activity of some CArG boxes present on cardiac promoters (Norman *et al.*, 1988; Boxer *et al.*, 1989). On SREs, SRF requires physical association with accessory proteins to fully mediate its transcriptional activation. Indeed, the growth factor induction on the c-fos promoter requires, in addition to SRF, the presence of an ETS family protein, elk/TCF. This protein binds the CArG motif with SRF to form a ternary complex that is the primary target for signal transduction pathways (Marais *et al.*, 1993). However, in the case of muscle CArG boxes, it is not clear if TCF also bind these boxes since the elk1 binding site is not always found. Interestingly, SRF has been described to interact with the muscle-enriched homeodomain proteins Mhox (Phox1) (Grueneberg *et al.*, 1992) and Nkx2-5 (Gajewski and Thompson, 1996). In both cases, the MADS domain of SRF interacts physically with the homeodomain of either Nkx2-5 and might represent the muscle equivalent of the SRF/TCF interaction. Thus, this interaction would explain the muscle-restricted activity of the CArG box. Furthermore, in a recent report that awaits

confirmation, the apparent muscle-specificity of CArG boxes could be explained by an enrichment of SRF expression in differentiating cardiac and skeletal muscle (Chen *et al.*, 1996). Nonetheless, the role of SRF in cardiac transcription since some reports identified cardiac and muscle-specific CArG-binding complexes that are distinct for SRF (Argentin *et al.*, 1994; Boxer *et al.*, 1989). Moreover, SRF binding to the ANF CArG box is not required for its cardiac activity since a mutation that eliminates SRF binding does not affect cardiac activity of the element (Argentin *et al.*, 1994). Finally, mutation or deletion of the *Drosophila* SRF homologue (pruned or DSRF) does not affect muscle function and all its biological function seems to require the Drosophila elk1 rather than other ancillary proteins (Guillemin *et al.*, 1996). Thus, further work such as the SRF knockout in mouse will be required to understand the role of SRF in cardiogenesis and to understand the relevance of the Nkx2-5/SRF interaction.

The MEF2 proteins (also known as RSRF) are also members of the MADS box family. The MEF2 gene products were identified in screens to find either proteins similar to SRF (Pollock and Treisman, 1991) or for their ability to interact with the MEF2 element (Yu *et al.*, 1992; Gossett *et al.*, 1989). Additional members of the MEF2 subfamily were cloned by low-stringency screening using the MADS domain of MEF2A and four different genes are now cloned [MEF2A (Yu *et al.*, 1992; Pollock and Treisman, 1991), MEF2B (Pollock and Treisman, 1991; Molkentin *et al.*, 1996), MEF2C (Martin *et al.*, 1993) and MEF2D (Breitbart *et al.*, 1993)]. The MEF2 proteins are extremely similar in their MADS domains and contain an additional region of homology in the carboxy-terminus of the MADS domain called the MEF2-specific domain. MEF2 proteins can homo- and heterodimerize to form different sets of MEF2 complexes (Martin *et al.*, 1994) and are potent transcriptional activators of promoters containing MEF2 elements, especially those in skeletal myocytes that also contain E-boxes. Moreover, direct injection of the RNA encoding the amphibian homologue of MEF2D, SL1, in *Xenopus* animal pole explants resulted in the induction of cardiac MLC2 transcripts (but not some others like α-MHC) indicating that MEF2 proteins are direct regulators of cardiac transcription (Chambers *et al.*, 1994). Like SRF, MEF2 proteins have been shown to acquire muscle-specific transcriptional activity by interacting with cofactors such as the myogenic bHLH

(Molkentin *et al.*, 1995; Kaushal *et al.*, 1994). In cardiac muscle, however, no interacting proteins have been found yet although one candidate might be Nkx2-5 since the SRF/Nkx2-5 interaction is mediated by the well-conserved MADS box. In fact, a putative MEF2/Nkx2-5 interaction might represent the true relevant *in vivo* MADS/NK2 interaction during cardiogenesis.

MEF2 mRNAs are widely expressed in the mouse but MEF2A, MEF2C and MEF2D transcripts seem to be particularly enriched in the muscle cell types. In addition, the MEF2 hnRNA undergoes extensive alternative splicing which gives rise to many cardiac-enriched forms (Martin *et al.*, 1994). In particular, MEF2C and MEF2B transcripts appear in the cardiogenic region as early as 7.5-8 dpc, at least one day before they are detected in the developing somites at day 8.5-9 dpc whereas the MEF2D and MEF2A transcripts appear in the heart forming region 12 hours later (Edmondson *et al.*, 1994). Precocious expression of XMEF2a in Xenopus resulted in the precocious expression of cardiac genes such as the XMHCa gene and lead to the formation of a larger heart, a phenotype also observed for Nkx2-5 and GATA-6 (Fu and Izumo, 1995). The importance of MEF2 proteins in myogenesis (including cardiogenesis) has been assessed in *Drosophila* where the pan-MEF2 ortholog (D-Mef2) was found to be essential for the terminal differentiation of all muscle lineages but not for their commitment (Bour *et al.*, 1995; Lilly *et al.*, 1995). Moreover, the recent gene knockout of MEF2C gene revealed that MEF2 proteins seem to have an important role in cardiogenesis. The MEF2C targeted disruption is embryonic lethal at 10.5 dpc and the lethality is caused by severe cardiac malformations (Lin *et al.*, 1997). The hearts of mutant embryos have a significant reduction in contractions and seem to have an absence of autonomous ventricular contraction. Moreover, rightward looping does not occur and there was no sign of right ventricle formation in the resulting hypoplastic ventricular chamber. Interestingly, the expression of cardiomyocyte markers such as ANF, MLC1a and dHAND were severely down-regulated in the mutant embryos, indicating that the defect might be of myocytic origin (Lin *et al.*, 1997). The role of MEF2 genes in the terminal differentiaton of muscle cells cannot be assessed yet due to an obvious genetic redundancy but, even though the generation of double and triple knockouts will be informative about the role of MEF2 protein in early

cardiogenesis, it is clear that MEF2 proteins are required for cardiogenesis.

1.3.2.4. HAND family.

The importance of the MyoD gene family for skeletal myogenesis led numerous investigators to search for equivalent tissue-specific bHLH master regulators. This search was especially done by investigators in the cardiac field since the heart is a striated muscle and shares a subset of contractile proteins with skeletal muscle. Extensive screens were done and led to the identification of numerous class A and B bHLH proteins of unknown function. bHLH proteins are classified in three broad classes. Class A proteins are defined by their ability to homo- and heter-dimerize with class A and B proteins whereas class B bHLH proteins bind E-box consensus sequences (CANNTG) only in heterodimer form with class A bHLH proteins (also called the ubiquitous bHLH proteins). Class C proteins only form homodimers. The analysis of cardiac promoters established the importance for E-box proteins of cardiac transcription. In cardiac muscle, several ubiquitous bHLH proteins are expressed, suggesting that some bHLH proteins distantly related to the MyoD family might play in important role in cardiac transcription. Moreover, Id proteins (inhibitors of class A and B bHLH proteins) are expressed at high level in some stages of cardiac development (Evans and O'Brien, 1993; Springhorn *et al.*, 1992), indicating that some class A or B bHLH proteins play some roles in cardiac transcription. Finally, the laboratory of David Bader reported the presence of a bHLH-related antigen in the developing heart. Collectively, these observations provide an indirect basis for the search of cardiac bHLH proteins .

Many reports have established that the bHLH-zipper protein USF was one of the major binding E-box binding protein in the heart (Navankasattusas *et al.*, 1994). However, its pattern of expression is not enriched in the myocardium and even though its activity might be modified in a cell-specific manner, it is still not a class B protein and is not inhibited by Id. However, the dHAND (also known as thing1, Hxt) and eHAND proteins (also known as thing2, Hed) are expressed in the developing myocardium as well as in other embryonic and extra-embryonic structures (Cserjesi *et al.*, 1995; Cross *et al.*, 1995; Srivastava *et al.*, 1997; Biben and Harvey, 1997). These two proteins are class B bHLH

proteins and were cloned in an interaction trap assay using a class A bHLH (Cserjesi *et al.*, 1995; Hollenberg *et al.*, 1995). Both genes are strongly expressed in extra-embryonic structures, neural-crest derived branchial arches and in the pericardium. Within the heart, the expression patterns of dHAND and eHAND are very dynamic, and dHAND mRNA is detected throughout the endocardium and the myocardium at the linear heart tube stage although with a notable gradient in the cranial domain of the heart (Biben and Harvey, 1997) and a notable enrichment on the right side of the developing myocardium (Srivastava *et al.*, 1997). In looping embryos, dHAND transcript levels decrease throughout the heart (especially in the caudal region) and by 9.0 dpc, its expression is restricted to the outflow tract (Biben and Harvey, 1997). On the other hand, eHAND expression is detected throughout the myocardium (but not in the endocardium) at the linear heart tube stage and, as looping proceeds, its expression become restricted to the left side of the heart, in the developing left ventricle and atrium. It is noteworthy to add that this expression pattern is not conserved in evolution since dHAND and eHAND expression is not asymmetric in chicken (Srivastava *et al.*, 1995). Thus, to understand their role in cardiac development, gene inactivation studies in chicken and mice have been carried out. In chicken, antisense knock-downs of HAND expression are not lethal although they result in an arrest of cardiac development at the looping stage only when specific antisense directed against both genes are added to the developing embryo (Srivastava *et al.*, 1995). This result suggests that HAND proteins are functionally redundant in the chick embryo. However, in mice the recent targeted disruption of the dHAND gene is somewhat discordant with the work in chicken since the dHAND gene disruption is embryonic lethal and is not fully compensated by eHAND (Srivastava *et al.*, 1997). The authors of these studies argue that these discrepancies might be due to species-specific differences. Mice with a disrupted dHAND gene die *in utero* at around 10.5 dpc due to severe cardiac defects and neural-crest defects. The aortic sac is grossly dilated in dHAND null mice and the hearts of the mutants fail to loop. At the molecular level, the expression of GATA-4 is severely reduced, although still present, in dHAND$^{-/-}$ mice (Srivastava *et al.*, 1997). The dHAND targeted disruption leads to a phenotype strikingly similar to the MEF2C knockout and MEF2 might be, along with dHAND, part of a common regulatory cascade leading to cardiac looping (Srivastava *et al.*,

1997). However, in the dHAND knockout, the defect is proposed to be partly due to the reduction of GATA-4 expression which is unaffected in the MEF2C knockout. Thus, it is still unclear what is the hierarchy of events leading to the cardiac looping defects obseved in dHAND and MEF2C null mutations. Altogether, this work indicate that dHAND is required for neural-crest-dependent cardiac morphogenesis and for cardiac looping, perhaps through the involvement of GATA-4.

1.3.2.5. TEF-1

The role of TEF-1 in cardiac transcription and development was independently unraveled by genetic and biochemical studies. Biochemical studies of the M-CAT binding factor identified TEF-1 as MCBF (Farrance *et al.*, 1992). TEF1 is a member of the rapidly expanding class of TEA domain transcription factors. The TEA domain is embedded in the DNA-binding domain of TEA proteins, and is composed of approximately 75 aa, structured in three putative α-helices. So far, four TEF genes have been cloned and the analysis of their expression revealed that none of them is restricted to muscle cells (Jacquemin *et al.*, 1996). However, alternative splicing of the TEF genes may give rise to cardiac-enriched isoforms (Stewart *et al.*, 1994) or, alternatively, the transcriptional activity of TEF proteins might be modified by their interaction with cardiac restricted factors to generate a muscle-specific activity as suggested by Larkin *et al.* (Larkin *et al.*, 1996). The importance of the TEF-1 gene in cardiogenesis was underlined by a retroviral gene trap experiment which disrupted the TEF-1 gene (Chen *et al.*, 1994). TEF-$1^{-/-}$ embryos die *in utero*, between 11 and 12 dpc, of cardiac malformations that involve ventricular wall thinning and severe reduction in trabeculation (Chen *et al.*, 1994). None of the presumptive transcriptional targets are affected by the TEF-1 mutation and it is still unclear if this defect is cell-autonomous. However if it is an intrinsic myocardial defect, the simplest explanation for this would be genetic redundancy among TEA family members or, as evidenced in the *in vivo* studies on the α-MHC promoter by the group of Jeffrey Robbins, by some sort of promoter plasticity that would compensate for the lack of M-CAT activity (Subramaniam *et al.*, 1991).

Collectively, cardiac transcriptional regulators play critical roles in almost all

aspects of cardiogenesis. How they collaborate together to generate the dynamic patterns of gene expression during cardiogenesis is only starting to emerge and will be an important field of study for years to come. Moreover, it is still unclear how these cardiogenic regulators are turned on by the cardiogenic signals emanating from the anterior endoderm and the outer ectoderm and the understanding of how cardiac inducers initiate cardiogenesis will be of paramount importance.

1.3.3. Post-transcriptional control of cardiac gene expression.

Most of the studies aimed at the understanding of cardiac-specific gene regulation have examined regulation of transcription. However, some evidence from biochemical and genetic studies indicate that post-transcriptional regulation plays a significant role in regulation of cardiogenesis and cardiac gene expression

1.3.3.1. Splicing control.

Studies on transcription factor and contractile protein regulation genes have established that many muscle-specific or developmentally regulated splice variants are expressed in the myocardium. These alternative splicing events are important since they can change the properties of the affected transcripts. For example, the muscle-specific splicing variant of MEF2D, MEF2D-1b possesses an alternative exon just 3' of the exon encoding the MADS domain (Martin *et al.*, 1994). This unique protein domain could be a novel protein-protein interaction domain which is required in muscle-specific transcription. However, very few studies have looked at the molecular regulation of splicing in the myocardium. The most characterized alternative splicing event in the myocardium is the alternative splicing of the cTNT transcript. It has been used as a model to study the mechanism of 3' site selection in muscle cells. The cTNT is composed of 18 exons, one of which (exon 5) is either present in embryonic heart and skeletal muscle or skipped in adult cardiac and skeletal muscle tissues. Interestingly, this developmental regulation results in a transition from >90% of mRNAs that include exon 5 in the embryo to >95% of mature transcripts that exclude this exon in the adult (Cooper and Ordahl, 1985). Mutational analyses of the skipped exon, carried out in transfection assays, established that some

residues within the exon are required for exon skipping *in vivo* and *in vitro* (Cooper and Ordahl, 1989). These studies identified a purine rich exon splicing element (ESE). Since mutation of the ESE does not block the splicing of the intron between exon 5 and 6, this element is required in 3' splice selection and is a true alternative splicing enhancer. Biochemical studies identified SR proteins as ESE-binding proteins. Interestingly, the affinity of the SR complex for the ESE is correlated with the strength of the enhancer *in vitro* and *in vivo* implying that the binding of SR proteins to the enhancer is the major determinant in this splicing event. However, in the case of the developmental regulation of this splicing event, it is not clear if the switch is caused by the loss of an important SR protein or by the expression of an inhibitory protein.

1.3.3.2. Translation control.

Very few studies have looked at translational regulation of gene expression in cardiac muscle although some observations suggest that it might actually play an important role. For example, MEF2 transcripts are expressed in a wide variety of tissues. However, both the binding activity and the presence of protein seem to be restricted to the muscle and brain cell types. Unfortunately, the mechanism by which the control of MEF2 mRNA translation occurs is unknown. However, studies on the translational regulation of the retinoic acid receptor-β2 (RARβ2) have implicated a cardiac and brain-specific translation regulatory mechanism. The 5' UTR of the RARβ2 transcript is long (461 nt) and possesses five, partially overlapping, uORFs (section 1.1.3.1) that might regulate its translation. Analyses in transgenic mice revealed that mutation of the uORFs altered significantly the expression of a lacZ reporter only in cardiac and brain tissues, unraveling a cardiac-restricted translational regulation pathway (Zimmer *et al.*, 1994). In cardiac tissue, the uORF4 negatively regulates the translation of the downstream ORF (Reynolds *et al.*, 1996). This model could now be used to define the biochemical activity involved in cardiac-restricted translational regulation and it would be interesting if a cardiac-restricted nonsense-mediated mRNA decay would be involved.

Collectively, even though there is little documented on the subject, these studies give interesting insights about post-transcriptional regulation of cardiac gene expression.

There is no doubt that the availability of cardiac-restricted post-transcriptional regulation events such as those involved in cTNT splicing and RARβ2 translation will enable a molecular analysis of these events and a better understanding of their respective roles in cardiac transcription and regulation.

1.4. HYPOTHESIS

This introduction focused on how the regulation of gene expression, with a particular emphasis on gene transcription, controls the fate of the precardiac and cardiac cells. Furthermore, I tried to underscore that the cardiac genetic program is initiated and regulated by the combinatorial action of multiple regulatory pathways. However, the nature of these pathways remains ill-defined and only handful have been unequivocally demonstrated to be involved in the control of cardiac transcription. Additionally, how these pathways interact to generate spatial and temporal diversity remains to be explored. It is thus of paramount importance to characterize novel cardiac regulatory pathways and to understand how they interact with each other to generate the diversity of gene expression patterns in the developing and mature myocardium. The main goal of my doctoral work was to identify and characterize novel cardiac regulatory pathways in order to understand how they interact to generate the expression pattern of a model gene, the gene encoding for the atrial natriuretic factor.

CHAPTER 2

The Atrial Natriuretic Factor gene is a downstream target for Nkx2.5 in the myocardium.

Daniel Durocher, Chen-Yi Chen, Ali Ardati, Robert J. Schwartz and Mona Nemer

Published in *Molecular & Cellular Biology*, Vol 16(9):4648-4655 (1996).

Previous studies from our laboratory established that the ANF promoter is organized in modules. The cardiac enhancer of the ANF promoter is divided in two regions (A and B) that contains elements necessary for high promoter activity in either the "adult" ventricle (A-domain) or for the embryonic myocardium and the post-natal atria (B-domain) [reviewed in (Durocher and Nemer, 1997)]. However, it was found that the proximal promoter (located between -135 and -150 bp) is necessary for high basal cardiac activity. This work is thus a characterization of this region that led to the discovery of a natural NK2 element (NKE) which binds the essential cardiac protein Nkx2-5.

This is a multi-authored authored work. Chen-Yi Chen is a post-doctoral fellow from Robert J. Schwartz's laboratory that cloned and provided the Nkx2-5 cDNA and, in addition, communicated unpublished information of paramount importance regarding the synthetic consensus sequence of the NKE. Ali Ardati generated Fig. 2.4A which is a footprint of the proximal ANF promoter region. All the remaining of the work was carried out by myself under the supervision of Mona Nemer.

2.1. ABSTRACT

The recently described NK2 family of homeodomain proteins encode key developmental regulators. In *Drosophila*, two members of this family, bagpipe and tinman, are respectively required for visceral and cardiac mesoderm formation. In vertebrates, *tinman* appears to be represented by a family of closely related NK2 genes, including Nkx-2.5, that is expressed at an early stage in precardiac cells. Consistent with a role for Nkx-2.5 in heart development, inactivation of the Nkx-2.5 gene in mice causes severe cardiac malformations and embryonic lethality. However, little is known about the molecular action of Nkx-2.5 and its targets in cardiac muscle. In this paper, we report the identification and characterization of a functional and highly conserved Nkx-2.5 responsive element termed the *NKE* in the proximal region of the cardiac ANF promoter. This NKE is composed of two, near-consensus, NK2 binding sites that are each able to bind purified Nkx-2.5. The NKE is sufficient to confer cardiac-specific activity to a minimal TATA containing promoter and is required for Nkx-2.5 activation of the ANF promoter in heterologous cells. Interestingly, in primary cardiocyte cultures, the NKE contributes to ANF promoter activity in a chamber and developmental specific manner suggesting that Nkx-2.5 and/or other related cardiac proteins may play a role in chamber specification. This work provides the identification of a direct target for NK2 homeoproteins in the heart and lays the foundation for further molecular analyses of the role of Nkx-2.5 and other NK2 proteins in cardiac development.

2.2. INTRODUCTION

The homeodomain is a 60 amino acid motif that folds into a very stable helix-turn-helix structure which is able to bind DNA with high affinity. Both genetic and biochemical data have shown that homeodomain proteins are sequence-specific transcriptional regulators most often involved in developmental processes like body segmentation, cell differentiation and organogenesis. The homeodomain itself is a very conserved unit found in all eukaryotes from yeast and protozoa to higher vertebrates. Moreover, specific members of the homeodomain superfamily are conserved throughout evolution in both primary sequence and function (17,18,48).

Recently, amino acid sequence comparison and low-stringency cloning unraveled a new family of homeoproteins termed NK2 whose members are found from planarians (Dth-1, Dth-2) (15,16) to leech (lox10) (38), *Drosophila* (NK2, NK3, NK4) (25), *C.* elegans (Ceh-22) (41), and vertebrates (Nkx-2.1 to -2.6) (20,29,45). The homeodomain of NK2 proteins is characterized by the presence of a very conserved tyrosine at position 54; in other homeoproteins like Antennapedia (Antp), this residue, a methionine, is known to directly contact DNA bases in the binding site (43). This amino acid difference between NK2 and Antp homeodomains may be responsible for the divergent target sites of NK2 proteins which have been found either through promoter analysis of target genes, as for the Nkx-2.1 (20) or by *in vitro* site selection as for Nkx-2.5 (8) and Nkx-2.1/TTF-1 (12). Outside the homeodomain, NK2 proteins possess a conserved segment of hydrophobic residues called the conserved peptide (CP) found just C-terminal of the homeodomain. This conserved peptide may be involved in protein-protein interaction or it may serve as a transcriptional inhibitory domain since its deletion in Nkx-2.5 and Nkx-2.1 enhances their transactivation properties (8,10).

Similar to other homeoproteins, NK2s control key steps in development. In *C.* elegans, *Ceh*-22 is involved in pharyngeal muscle development (41). In *Drosophila*, NK2 (or *vnd*) is thought to be involved in early neurogenesis by regulating the *achaete-scute* complex (22), NK3 (or *bagpipe*) has been shown to be required for visceral muscle formation (5) and NK4 (tinman) is required for pre-cardiac mesoderm formation (6,29). In

vertebrates, Nkx-2.1 is a key regulator of thyroid and lung gene expression (7,20) and is also essential for thyroid and lung organogenesis (26) and function (1). The presumptive mammalian homologue of tinman, Nkx-2.5, is expressed in the primordial spleen, the pharyngeal mesoderm, and the pre-cardiac mesoderm where its expression precedes that of other cardiac markers (27,29). Consistent with a role for Nkx-2.5 in the heart, targeted disruption of the Nkx-2.5 gene arrests heart development at the looping stage causing severe morphogenetic defect of the heart and embryonic lethality (31). This phenotype which is less severe than the *tinman* phenotype, may be due to genetic redundancy between Nkx-2.5, Nkx-2.6, and Nkx-2.3 which are all present in vertebrate myocardia (11,29,51). Elucidating the molecular mechanisms of action of these factors and identification of their target genes in the heart are critical to understanding their respective roles in myocardial development.

In this paper, we present data showing that the cardiac atrial natriuretic factor promoter (4) is a target for Nkx-2.5 and contains a Nkx-2.5 response element termed NKE. This NKE which harbors two near consensus binding sites for Nkx-2.5, interacts specifically with purified Nkx-2.5 and represents one of the first identified natural binding sites for Nkx-2.5 on a cardiac promoter.

2.3. MATERIALS AND METHODS

Cell cultures and transfections. Primary cardiocyte cultures were prepared from 1 or 4 day old Sprague-Dawley rats and kept in serum-free medium as described previously (3). L cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. Transfections were carried out using calcium phosphate precipitation 24 h after plating. 36 h post-transfection, cells were harvested and luciferase activity was assayed with an LKB luminometer and the data was recorded automatically. In all experiments, RSV-hGH was used as internal control and the amount of reporter was kept at 3 μg per dish; the total amount of DNA was kept constant (usually 8 μg). Unless otherwise stated, the results reported were obtained from at least four independent experiments with at least two different DNA preparations for each plasmid.

Plasmids. ANF-luciferase promoter constructs were cloned in the PXP-2 vector as described previously by Argentin et al (3). Mutations were generated by PCR and sequenced. The internal deletion of nucleotides -105 to -84 was constructed by inserting an oligonucleotide encompassing the -135 to -106 bp in the *Hin*dIII site of the -50 bp construct. Plasmids containing the NKE sites upstream of the ANF -50 bp minimal promoter were generated by polymerizing one or multiple copies of a *Bam*HI-*Bgl*II NKE synthetic oligonucleotide in the *Bam*HI site of the -50 ANF-luciferase plasmid. The MBP-Nkx-2.5 and pEMSV-Nkx-2.5 plasmids were described in Chen et al. 1995 (8).

Recombinant protein and antibody production. After transformation of BL21(DE3) *E.* coli strain with the MBP fusion vectors derived from pMalc-2 (New England Biolabs), individual colonies were picked and grown in 50 ml 2XYT up to an O.D. of 0,6. IPTG was then added at a final concentration of 0,5 mM. Bacterial cultures were then grown at 37°C for another hour. The cultures were centrifuged and the bacteria lysed by sonication. The purification procedure on amylose columns was followed according to the manufacturer's instructions. Anti-Nkx-2.5 polyclonal antibody was generated by immunization of rabbits with purified MBP-Nkx-2.5 according to standard procedures. Immune serum was affinity purified and characterized by Western blotting.

DNA-binding assays. Whole cell extracts were prepared from cardiac myocyte cultures as previously described (2). Protein concentration was assayed by Bradford assay. Binding reactions for electrophoretic mobility shift assays (EMSA) were performed on ice for 20 min in the presence of 500 ng of poly(dA-dT) and 500 ng of poly(dI-dC) in 60 mM KCl, 6 mM $MgCl_2$, 10 mM Tris-Cl (pH 7.9), 1 mM DTT, 12% glycerol, 20 000 cpm of double-stranded synthetic oligonucleotides and using typically 6-8 μg of whole cell extract or fractional nanogram amount of affinity-purified MBP-fusion protein. For antibody blocking assays, typically 200 ng of affinity purified α-Nkx-2.5 or an unrelated antibody (N.R.) was incubated in the presence of the protein extract 1 h prior to the addition of the probe. Footprinting experiments using whole-tissue nuclear extracts were performed as previously described (3).\

Table 2.1. Oligonucleotides used in this study

oligonucleotide	sequence[a]	ref
NKE	GCCGCCG**CAAGTGACAGAATG**GGA	this study
MA	GCCGCCGCA<u>T</u>A<u>T</u>GA**CAGAATG**GGA	this study
MB	GCCGCCG**CAAGTG**ACAG<u>CC</u>A<u></u>GGGA	this study
DM	GCCGCCGCA<u>T</u>A<u>T</u>GACAG<u>CC</u>A<u></u>GGGA	this study
2.1	GCCCAG**TCAAGTG**CA	(20)

[a]NKE motifs are bolded, mutations are underlined

2.4. RESULTS

The ANF proximal promoter harbors a NK2-like cardiac element. The atrial natriuretic factor (ANF) gene, which encodes the heart major secretory product, is an early marker of cardiomyocyte differentiation (53). ANF transcription is tightly regulated spatially during cardiac development. Previous studies in our laboratory have identified within the first 700 bp of the ANF promoter, two regions necessary for proper spatio-temporal expression of the gene (3). The one, located between -50 and -135 bp, is required for basal activity and for the function of the cardiac-specific enhancer located between -135 and -700 bp. A complex array of molecular pathways converges at the ANF promoter as exemplified by the identification of chamber- and stage-specific regulatory elements in the ANF enhancer (3,32) and the presence of an inducible, α1-adrenergic response element (the PERE) within the proximal promoter (2). However, except for GATA-4 that activates the ANF promoter (19) via the GATA binding site found in the proximal promoter (F. Charron and M. Nemer, in preparation), the identity of the trans-acting factors that modulate ANF transcription remains largely unknown.

Since the PERE appeared to contribute little basal activity in cardiomyocytes (2), we carried out deletion analysis of the proximal promoter in order to identify the DNA elements and their cognate nuclear proteins that account for transcriptional activity of the -135 to -50 bp region. As shown in Figure 2.1, sequence analysis reveals the presence of a number of consensus motifs including a GATA binding site juxtaposed to an AT-rich

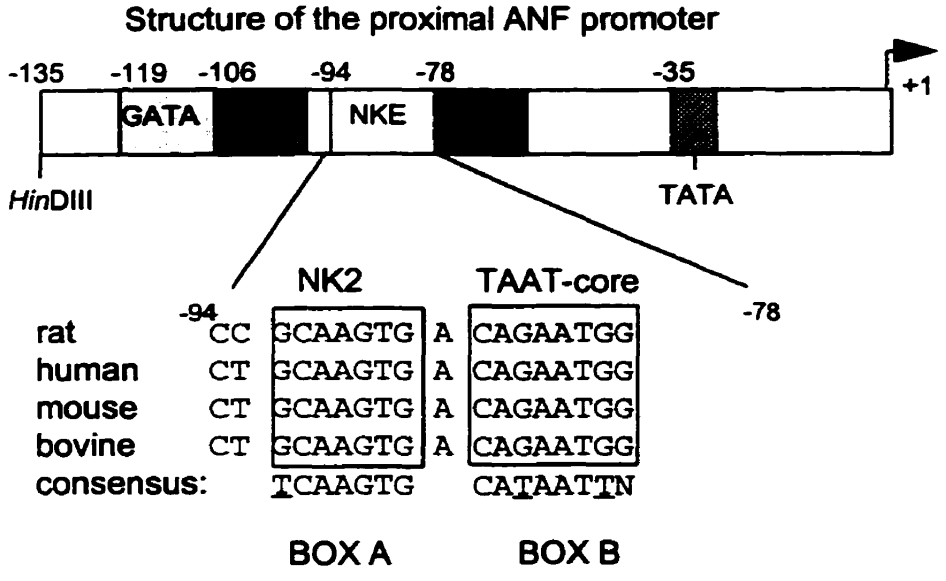
element and a region of high homology to newly identified NK2 binding sites (8,19). This organization is identical between ANF promoters from different species. Results of the deletion analysis reveal that the GATA and A/T-rich element as well as the putative NK2 binding site (termed NKE) contribute to the transcriptional activity of the proximal promoter in 1 d old ventricular cardiomyocytes (Fig. 2.1b). Indeed, deletion of sequences up to -106 bp removes both GATA and A/T-rich elements and leads to a 2.5-fold drop in promoter activity. Furthermore, deletion of sequences between -91 and -50 bp decreases promoter activity by over 12-fold. Sequences between -91 and -78 account for most of this effect since, consistent with previous work (2), internal deletion of sequences between -78 and -57 bp which harbor the PERE reduces promoter activity only modestly. Interestingly, the -91/-78 region contains two highly conserved motifs with homologies to consensus NK2 binding sites (Fig. 2.1). An oligonucleotide spanning this region is able to activate the minimal TATA containing ANF promoter by 10- to 15-fold in atrial or ventricular cardiomyocytes but not in non-cardiac cells (Fig. 2.2a).

The relevance of the NKE to ANF promoter activity was also assessed in the context of the maximally active -700 bp promoter both in atria and ventricles of 1 and 4 d old rats. We had previously shown that different transcription pathways were active in these stage-specific cultures (3). As shown in Figure 2.3a, the NKE contributes similarly to ANF promoter activity in 1 d old atrial and ventricular cardiomyocytes. Unexpectedly, this effect is lost in more differentiated ventricular myocytes and in fibroblast L cells where removal of the NKE leads to a consistent 2- to 10-fold activation, respectively (Fig. 2.3b and 2.3c). In contrast, promoter activity is most dependent on the NKE in differentiated atrial myocytes (Fig. 2.3b) prepared simultaneously from the same 4 d old rat hearts. These changes in promoter activity including the marked upregulation in L cells are not seen with a similar internal deletion that removes sequences upstream of the NKE (data not shown). Thus, it is unlikely that the gain of activity is simply the consequence of moving positive elements closer to the transcription initiation site; rather, it appears that the NKE itself harbors overlapping negative elements. This presence of such negative regulators within tissue-specific elements has already been observed (2,3). Thus, the NKE, in collaboration with other regulatory elements of the ANF promoter, contributes to cardiac transcription in

Figure 2.1.

Localization of a NK2-like binding site in the proximal ANF promoter. **(A)** Structural organization of the proximal ANF promoter. Putative regulatory elements of the ANF promoter are boxed and their location relative to the transcription start site is indicated. The PERE corresponds to the α_1-adrenergic resonsive sequence (2). The Nkx-2.5 consensus is based on the *in vitro* site selection studies of Chen and Schwartz (8). This structure is identical for rat, mouse, human and bovine ANF promoters. **(B)** The NKE is critical for ANF promoter activity. Transient transfection of rANF-luciferase vectors in primary cardiocyte cultures derived from dissected ventricular tissues of 1 d old rats. The results are expressed relative to the activity of the RSV-luciferase promoter and represent the mean +/-SD of at least four independent determinations. The shaded box represents the putative NKE (-91 to -78 bp).

A


Structure of the proximal ANF promoter
-135
-119
-106
-94
-78
-35
+1
GATA
NKE
HinDIII
TATA
NK2
TAAT-core
-94
-78
rat CC GCAAGTG A CAGAATGG
human CT GCAAGTG A CAGAATGG
mouse CT GCAAGTG A CAGAATGG
bovine CT GCAAGTG A CAGAATGG
consensus: TCAAGTG CATAATTN
BOX A
BOX B

B

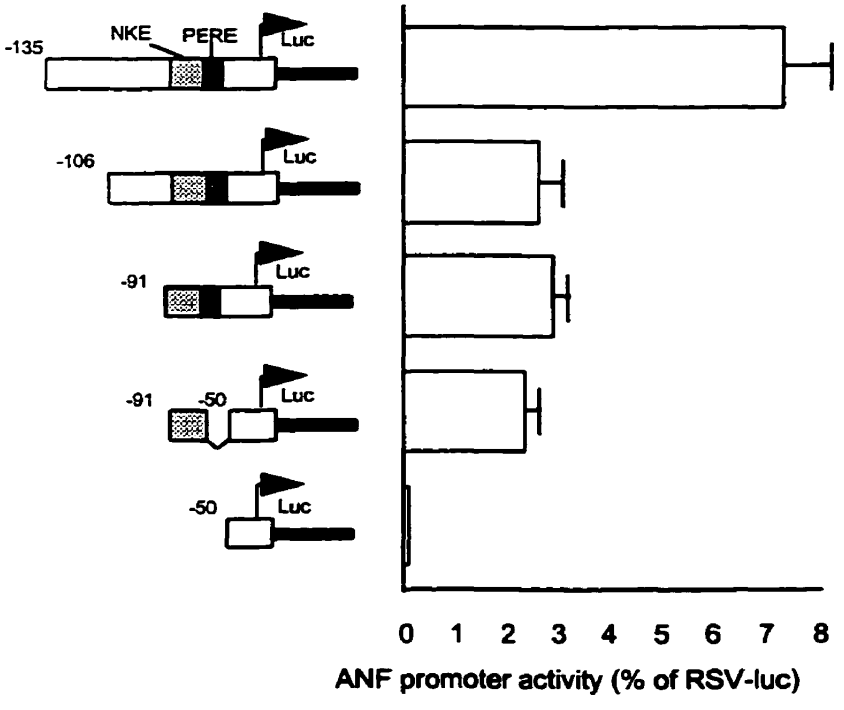
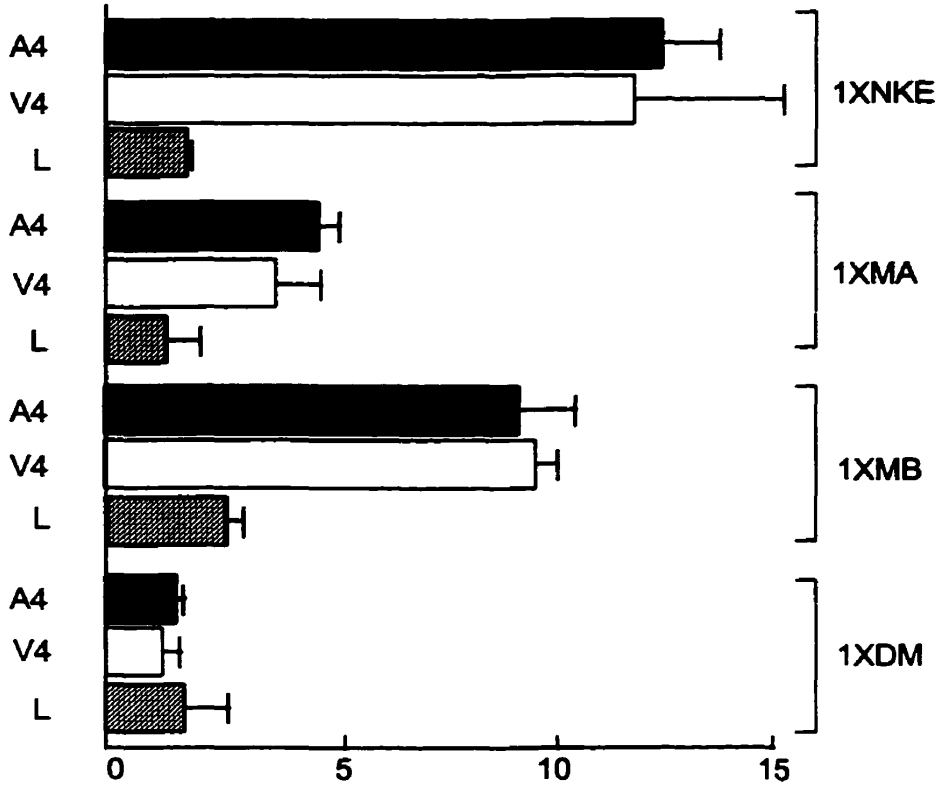

-135
NKE
PERE
Luc
-106
Luc
-91
Luc
-91
-50
Luc
-50
Luc
0 1 2 3 4 5 6 7 8
ANF promoter activity (% of RSV-luc)

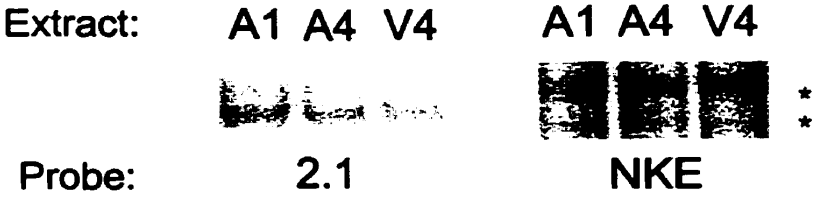
Figure 2.2.

The NKE requires both A and B boxes to be fully active. **(A)** An oligonucleotide containing one copy of the NKE is sufficient to confer cardiac specificity to a minimal ANF promoter in myocytes from 4 d old atria (A4) or ventricles (V4) but is inactive in fibroblast L cells. MA, MB and DM correspond to mutant NKEs as depicted in Table 2.1. Transfections were carried out as detailed in Material and Methods and the results (n=4) are expressed as fold activation over the minimal ANF-50 construct (± SD). **(B)** Nkx-2.5-like binding activity remains the same in different stage- and chamber-specific myocyte cultures. Whole cell extracts derived from 1 or 4 d old atrial myocytes (A1 and A4) and 4 d old ventricular myocytes (V4) were assayed for their ability to bind the NKE or the 2.1 probe. The panel represents a blow-up of the Nkx-2.5-related binding and corresponds to the complex abolished by the Nkx-2.5 antibody and by an unlabelled high affinity Nkx-2.5 binding site as shown in Figure 2.4c. The doublet seen likely corresponds to different Nkx-2.5 isoforms which are also obtained in L cells transfected with the Nkx-2.5 expression vector.

A


A4
V4
L
1XNKE
A4
V4
L
1XMA
A4
V4
L
1XMB
A4
V4
L
1XDM
0
5
10
15
Fold Activation (over -50 bp ANF)

B


Extract:
A1 A4 V4
A1 A4 V4
Probe:
2.1
NKE

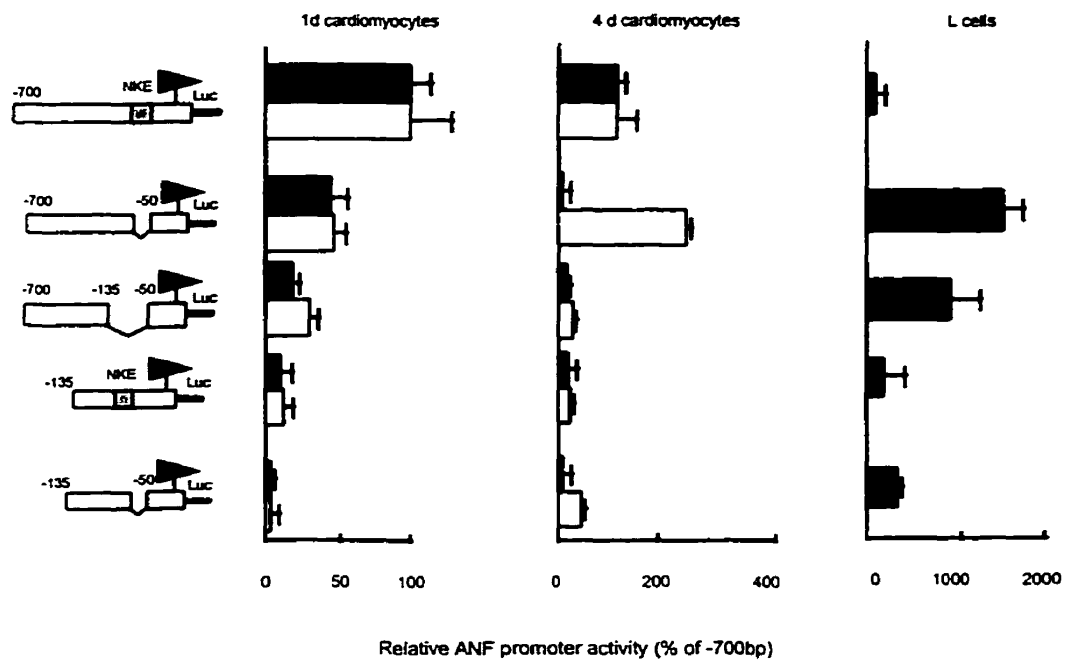


Figure 2.3.

The NKE contributes to ANF promoter activity in a developmental and chamber-specific manner. Transfections were carried out in primary cardiomyocytes derived from 1 or 4 d old rats as described in Material and Methods. Open bars represent ventricular myocytes and black-filled bars represent atrial myocytes. Note that the NKE appears to be most critical to ANF promoter activity in 4 d old atria (Fig. 2.3b) whereas it has little or even inhibitory effect in 4 d old ventricles (Fig. 3b) and in fibroblasts (Fig. 2.3c). The results are expressed relative to the activity of the -700 bp promoter taken as 100% in each cell type. For cardiomyocytes, n=8 and for L cells n=4.

a chamber- and stage-specific manner.

Nkx-2.5 binds and activates the NKE. Next, we analyzed the interaction of this element with cardiac nuclear extracts and with purified Nkx-2.5 protein. DNase I footprinting revealed the presence of two contiguous footprints between -56 to -76 and -77 to -94 that encompass the previously characterized α_1-adrenergic response element (PERE) and the NKE, respectively (Fig. 2.4a). The NKE footprint was restricted to cardiac extracts (Fig. 2.4a) indicating the presence of a tissue-specific DNA binding activity. Because the NKE contains homologies to NK2 binding sites, the footprint might correspond to binding of cardiac NK2 proteins.

Nkx-2.5 is presently the only well characterized NK2 protein in mammalian heart. *In vitro* site selection studies have revealed that Nkx-2.5 binds with high affinity to sequences corresponding to the consensus: TCAAGTG and with moderate affinity, *Antp*-like binding sites which contains the typical TAAT core (8). The NKE possesses two putative NK2 binding sites (Fig. 2.1), the first one CCAAGTG (box A) diverges at a single nucleotide from the high affinity consensus TCAAGTG, while box B consists of a modified TAAT core (CAGAATGG instead of CATAATTN). In EMSA, oligonucleotide probes containing both or either boxes (Table 2.1) form specific complexes with bacterially expressed *or in vitro* translated Nkx-2.5 (Fig. 2.4b and data not shown) which are competed by the unlabelled high affinity synthetic Nkx-2.1 binding site (Fig. 2.4b). Interestingly, direct binding and displacement studies indicate that the two boxes within the NKE can independently interact with Nkx-2.5 and only mutation of both sites completely abolishes this interaction (Fig. 2.4b and data not shown).

EMSAs were also carried out with cardiac extracts. As shown in Fig. 2.4c, incubation of atrial myocyte extracts with the NKE produced two major retarded complexes, one of which was completely abolished in presence of a Nkx-2.5-specific antibody. Moreover, this complex was effectively abrogated in presence of excess unlabelled oligonucleotides that retained the ability to bind purified Nkx-2.5 (Fig. 2.4b) but was unaffected by the NKE mutant (DM) that no longer interacted with Nkx-2.5. Together, these results strongly suggest that Nkx-2.5 or a highly related protein present in cardiomyocytes interacts with the NKE. Furthermore and consistent with the high level

activity of the NKE in atrial and ventricular myocytes, the Nkx-2.5-related complex was present at similar level in both atrial and ventricular myocyte extracts.

In order to ascertain the functional relevance of Nkx-2.5 to ANF transcription, ANF reporter constructs were cotransfected in heterologous cells with Nkx-2.5 expression vector. As shown in Fig. 2.5a, Nkx-2.5 is a potent activator of the full length but not the minimal ANF promoter. This dose-dependent activation is already seen with 0.25 μg of expression vector and reaches 10- to 15-fold when 2.5 to 5 μg of Nkx-2.5 vector are used. The closely related Nkx-2.1 protein was also able to transactivate the ANF promoter; other homeodomain proteins expressed or not in the heart were also tested but none could activate the ANF promoter including Gax, Mhox, Otx1, Oct1 and Pit1 (data not shown). The inductive effect of Nkx-2.5 was observed with all ANF promoters that retained the NKE but not with constructs lacking the NKE (Fig. 2.5b and data not shown). Interestingly and consistent with the binding data (Fig. 2.4), the presence of either site A or B was sufficient for NKE transactivation by Nkx-2.5; however, this transactivation was abolished with the mutant NKE in which both sites are mutated and which is no longer able to bind Nkx-2.5 (Fig. 2.5c). Together, the data presented indicate that the ANF promoter is a target for Nkx-2.5 and possibly other NK2 related proteins in the heart.

2.5. DISCUSSION

The NK2 class of homeodomain proteins plays a central role in cardiac development. However the molecular circuitry underlying the action of NK2 proteins in the myocardium remains unknown. The present work provides data showing that Nkx-2.5 trans-activates the cardiac ANF promoter via a novel cardiac-specific *cis*-element, the NKE which interacts specifically with recombinant Nkx-2.5 and Nkx-2.5 related antigens in cardiac extracts. Together the data suggest that the ANF gene may be a direct downstream target for NK2 proteins in the myocardium.

The NKE is a new cardiac *cis*-element. At the level of the ANF promoter, the site where Nkx-2.5 binds harbors the characteristics of a new cardiac specific *cis*-element. So far only a limited number of such elements have been identified within cardiac-specific

Figure 2.4.

Interaction of cardiac and recombinant proteins with the NKE. (A) DNase I footprinting experiments were carried out using a 5' end-labeled 200 bp rANF (-135 to +50) fragment and 10 μg of tissue nuclear extracts prepared from the indicated tissues of 1 d old rats according to Argentin et al. (3). The position of the protected regions were assigned based on the simultaneously run Maxam-Gilbert sequencing reaction (G track). The -56/-77 footprint (PERE) corresponds to the α_1-adrenergic response element previously characterized (2) while the -77/-94 footprint covers the NKE. (B) Bacterially produced Nkx-2.5 can bind to the ANF NKE. Nkx-2.5 fusion protein production and binding conditions are described in Material and Methods. Nkx-2.5-DNA interaction was assayed over the NKE probe. Unlabelled competitor oligonucleotides were used at 250-fold molar excess. The sequence of the oligonucleotides used is depicted in Table 2.1. (C) Endogenous Nkx-2.5 or Nkx-2.5-related antigens present in atrial extracts bind the NKE. Whole cell atrial myocyte extracts were assayed for their ability to form complexes with the NKE or a mutant oligonucleotide which retains box A (MB). Both purified MBP-Nkx-2.5 and atrial extracts were pre-incubated with an affinity purified anti-Nkx-2.5 antibody or an unrelated purified antibody (N.R.). Pre-incubation with the anti-Nkx-2.5 abolished the formation of the MBP-Nkx-2.5-DNA complex and the upper complex formed in atrial extracts. This upper complex was also abolished by unlabelled competitor oligonucleotides that bind Nkx-2.5 (MA, MB, 2.1) but not by a mutant NKE site (DM) that no longer binds Nkx-2.5. In all cases, unlabelled oligonucleotides were used at 150-fold molar excess.

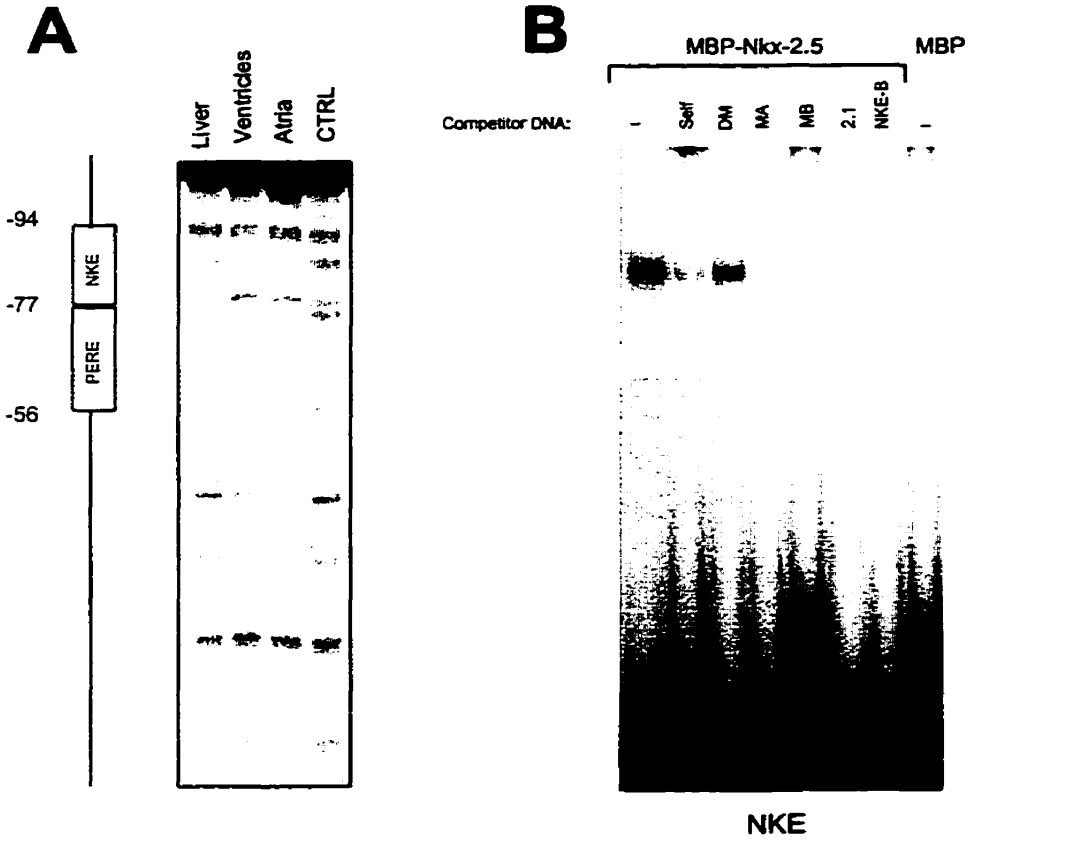
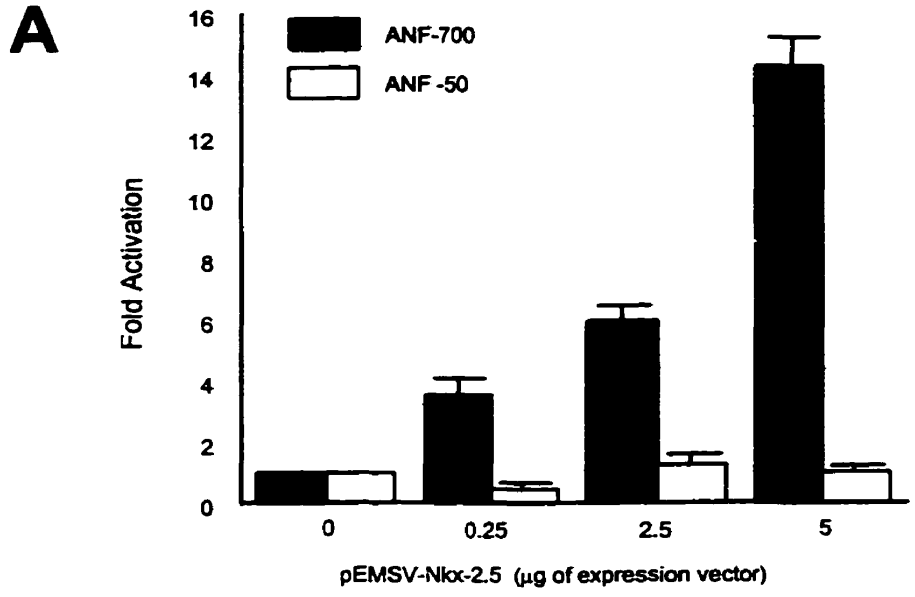
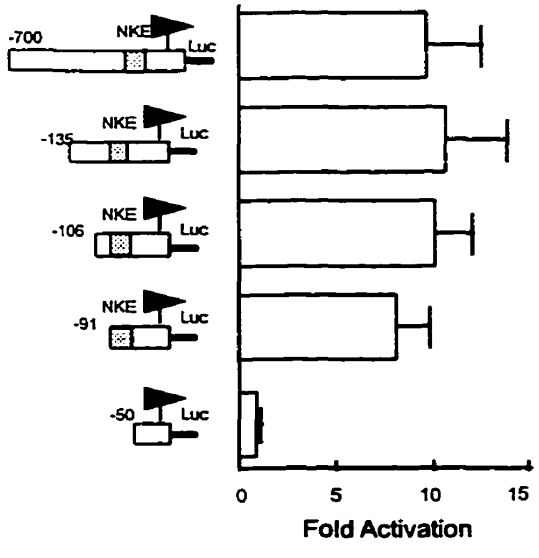
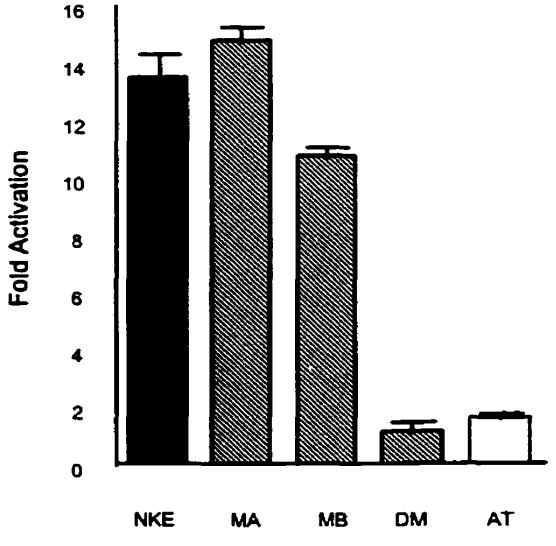

A
Liver
Ventricles
Atria
CTRL
-94
-77
-56
NKE
PERE
B
MBP-Nkx-2.5
MBP
Competitor DNA:
–
Self
DM
MA
MB
2.1
NKE-B
–
NKE
C
MBP-Nkx-2.5
Atrial Extracts
Atrial Extracts
Antibody:
N.R.
α-Nkx-2.5
Competitor DNA:
MBP-Nkx-2.5
Nkx-2.5
NKE
NKE
MB

Figure 2.5.

(A) Nkx-2.5 activates the rANF -700 promoter in a dose-dependent manner. Transient cotransfection experiments were done in L cells using the calcium phosphate method as described in Material and Methods. The results represent the mean +/-SD of at least four independent experiments. (B) The Nkx-2.5-responsive region maps to the NKE. Cotransfection experiments were done in L cells using 2 μg of pEMSV-Nkx-2.5 as described in a). The results represent the mean (+/-SD) of at least four independent experiments. (C) Nkx-2.5 can transactivate the NKE. Cotransfection experiments were done in L cells using 5 μg of the expression vector and 3 μg of various reporter constructs corresponding to different oligonucleotides (Table 2.1) fused in one copy upstream of the -50 bp ANF promoter. MA and MB correspond to mutations of the A or B box of the NKE whereas DM is a double mutant of both. Not all "TAAT cores" containing sequences are responsive to Nkx-2.5 since an AT-rich region encompassing nucleotides -600 to -565 of the ANF promoter (AT) is unaffected by Nkx-2.5. Solid bars represent the ANF minimal promoter. promoters; they include the


A
ANF-700
ANF -50
Fold Activation
16
14
12
10
8
6
4
2
0
0
0.25
2.5
5
pEMSV-Nkx-2.5 (µg of expression vector)


B
-700
NKE
Luc
-135
-106
-91
-50
0
5
10
15
Fold Activation


C
Fold Activation
16
14
12
10
8
6
4
2
0
NKE
MA
MB
DM
AT

WGATAR (19,21,35), M-CAT (36,47,50), CArG (3,33,47,52), and Mef2 (23,36,39,46) motifs that are known to participate in cardiac-specific transcription. Interestingly, most cardiac promoters contain several of these elements suggesting that cardiac transcription is controlled by a complex regulatory network. For example, the ANF promoter contains, in addition to the NKE, WGATAR, CArG and AT-rich/Mef2 elements that may function in different combinatorial pathways to direct proper spatial and stage-specific transcription of the ANF gene (discussed below). A search of different cardiac promoters revealed the presence of NKE-like elements within the upstream regulatory regions of the cardiac MLC2 (40) and β myosin heavy chain (MHC) genes (28) (Table 2.2). Incidentally, expression of these two genes is altered in mice homozygous for a null Nkx-2.5 allele which display defective chamber specification (31). However, whether MLC2 or βMHC are direct targets for Nkx-2.5 will need to be tested. Nevertheless, it is intriguing that the ANF NKE and the putative NKEs on the MLC2 and βMHC promoters all contain an E box motif CAAGTG that can interact with some bHLH factors (Durocher and Nemer, unpublished data). In skeletal muscle, myogenic bHLH factors are key regulators of tissue-specific transcription and cell differentiation (42). Although no myocardial specific bHLH factors have been identified yet, mutational analysis of some cardiac promoters have revealed the importance of regulatory elements containing E box motifs for their activity in primary cardiomyocyte cultures (34,37). Given the data with the ANF NKE, it is possible that these E box motifs may in fact lie within functional NKEs. It is also possible that overlapping NKEs and E boxes might reflect competitive regulation of cardiac transcription by NK2 and bHLH factors during development. The identification of a transcriptional target for NK2 factors in the myocardium will now allow direct testing of this hypothesis.

Finally, it is noteworthy that, on the ANF promoter, the NKE is located within the first 100 bp of upstream sequences similar to the two functional Nkx-2.1 binding sites in the thyroglobin (20) and lung surfactant B (7) promoters. In addition, in all three cases, the NKE is composed of a tandem of two sites that can independently bind Nkx-2.1 or -2.5 protein. In the case of the ANF NKE, mutation of either site decreases NKE activity in cardiomyocytes with mutation of the A site resulting in the most significant loss of activity (30% of wild-type); however, only mutation of both sites completely abolishes the NKE

activity. In this respect, it should be noted that the ANF NKE is an imperfect match of the high affinity site selected *in vitro* (8) and has a lower apparent affinity to Nkx-2.5 than these sites. Thus, the presence of both sites may be required for productive interaction with cofactors or with components of the transcriptional machinery; alternatively, it may simply insure an adequate local concentration of Nkx-2.5 or, may enhance the binding affinity to the protein. In this respect, it should be noted that the ANF NKE is an imperfect match of the high affinity sites selected *in vitro* (8) and has a lower apparent affinity to Nkx-2.5 than these sites. This in turn might provide added *in vivo* specificity in discriminating between structurally related proteins that are coexpressed in the myocardium but that may serve distinct developmental functions as previously suggested from *in vivo* studies of the conserved low affinity binding sites for the *fushi tarazu* homeo protein (17).

Stage and chamber specificity of NKE activity. One of the unexpected results of this study was the finding that NKE activity within the ANF promoter parallels the stage and chamber specificity of the endogenous ANF gene and the transfected ANF promoter, i.e. NKE contribution to promoter activity was highest in differentiated atria and lowest in differentiated ventricles. So far, few identified cardiac *cis*-elements possess such properties. Many cardiac genes are expressed in a developmental- and chamber-specific manner (reviewed by Lyons et al) (30) and some cardiac promoters including ANF display spatial specificity in transgenic mice (13,14,24,44,49). However, neither the regulatory elements nor the transcription factors that are responsible for chamber-specific expression have been identified. The mechanism underlying chamber-specific activity of the NKE is not readily understood since the pattern of expression of Nkx-2.5 within the heart does not show such specification; nevertheless, inactivation of the Nkx-2.5 gene unraveled a role for Nkx-2.5 in heart regionalization (31). Given that the transcriptional activity of the NKE *per se* does not differ between atrial and ventricular myocytes (Fig. 2.2b), it is likely that the chamber-specific activity of the NKE reflects interactions between Nkx-2.5 and other factors bound to neighboring promoter elements. In this respect, Chen et al. have shown cooperative interactions between Nkx-2.5 and serum response factor (SRF), a MADS box factor, over the cardiac α actin promoter (9). This interaction occurs at the SRE and does not seem to be dependent on Nkx-2.5 binding. Perhaps SRF or another MADS box factor,

Mef2, bound at upstream elements could interact with Nkx-2.5 and account for developmental and/or spatial specificity. Alternatively, the interaction of Nkx-2.5 with chamber-specific cofactors (activators or repressors) might result in the observed spatial specificity of the NKE. It is also noteworthy that the NKE is located near a functional GATA binding site and an AT-rich region which can interact with members of the MADS box family. Further studies will need to examine the interesting possibilities of cooperative interaction between Nkx-2.5 and other cardiac factors such as GATA-4 or Mef2 and their integration at the level of the ANF and other cardiac promoters.

2.6. ACKNOWLEDGMENTS

We are grateful to Michel Chamberland for oligonucleotide synthesis, to Lise Laroche for secretarial assistance and to members of the Nemer's laboratory for discussions and critical reading of this manuscript. This work was supported by grants from the Canadian Medical Research Council (MRC) and the National Institutes of Health (RO150422). DD is a recipient of the Canada 1967 fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC), and MN is MRC Scientist.

Table 2.2. NKE-like motifs in cardiac promoters

PROMOTER	NKE motif	POSITION
		-
ANF (4)	GCAAGTG, CAGAATG	92, -84
β-MHC (28)	CCAAGTG, TCAAGTG,	-1737, -1257
	CAGAATG, CAGAATG	-1691, -1545
MLC-2 (40)	A: GCAAGTC, GCAAGTG	-190, -1194

2.7. REFERENCES

1. **Acebron, A., P. Aza-Blanc, D. L. Rossi, L. Lamas, and P. Santisteban.** 1995. Congenital human thyroglobulin defect due to low expression of the thyroid-specific transcription factor TTF-1. J. Clin. Invest. **96**:781-785.
2. **Ardati, A. and M. Nemer.** 1993. A nuclear pathway for a_1-adrenergic receptor signaling in cardiac cells. EMBO J. **12**:5131-5139.
3. **Argentin, S., A. Ardati, S. Tremblay, I. Lihrmann, L. Robitaille, J. Drouin, and M. Nemer.** 1994. Developmental stage-specific regulation of atrial natriuretic factor gene transcription in cardiac cells. Mol. Cell. Biol. **14**:777-790.
4. **Argentin, S., M. Nemer, J. Drouin, G. K. Scott, B. P. Kennedy, and P. L. Davies.** 1985. The gene for rat atrial natriuretic factor. J. Biol. Chem. **260**:4568-4571.
5. **Azpiazu, N. and M. Frasch.** 1993. tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. Genes Dev. **7**:1325-1340.
6. **Bodmer, R.** 1993. The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. Dev. **118**:719-729.
7. **Bohinski, R. J., R. Di Lauro, and J. A. Whitsett.** 1994. The lung-specific surfactant protein B gene promoter is a target for thyroid transcription factor 1 and hepatocyte nuclear factor 3, indicating common factors for organ-specific gene expression along the foregut axis. Mol. Cell. Biol. **14**:5671-5681.
8. **Chen, C. Y. and R. J. Schwartz.** 1995. Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. J. Biol. Chem. **270**:15628-15633.
9. **Chen, C. Y. C., T. McQuinn, M. J. Frankovsky, and R. J. Schwartz.** 1996. Cardiac a-actin gene, a downstream target for Tinman homologue, Nkx-2.5, and serum resonse factor. Submitted
10. **Defelice, M., G. Damante, M. Zannini, H. Francislang, and R. Dilauro.** 1995. Redundant domains contribute to the transcriptional activity of the thyroid

transcription factor 1. J. Biol. Chem. **270**:26649-26656.

11. **Evans, S. M., W. Yan, M. P. Murillo, J. Ponce, and N. Papalopulu**. 1995. Tinman, a drosophila homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates - xnkx-2.3, a second vertebrate homologue of tinman. Dev. **121**:3889-3899.
12. **Fabbro, D., G. Tell, L. Pellizzari, A. Leonardi, C. Pucillo, R. Lonigro, and G. Damante**. 1995. Definition of the DNA-binding specificity of TTF-1 homeodomain by chromatographic selection of binding sequences. Biochem. Biophys. Res. Commun. **213**:781-788.
13. **Field, L. J.** 1988. Atrial natriuretic factor-SV40 T antigen transgenes produce tumors and cardiac arrythmias in mice. Science **239**:1029-1033.
14. **Franz, W. M., D. Breves, K. Klingel, G. Brem, P. H. Hofschneider, and R. Kandolf**. 1993. Heart-specific targeting of firefly luciferase by the myosin light chain-2 promoter and developmental regulation in transgenic mice. Circ. Res. **73**:629-638.
15. **Garcia-Fernandez, J., J. Baguna, and E. Salo**. 1991. Planarian homeobox genes: cloning, sequence analysis, and expression. Proc. Natl. Acad. Sci. USA **88**:7338-7342.
16. **Garcia-Fernandez, J., J. Baguna, and E. Salo**. 1993. Genomic organization and expression of the planarian homeobox genes Dth-1 and Dth-2. Dev. **118**:241-253.
17. **Gehring, W. J., M. Affolter, and T. Burglin**. 1994. Homeodomain proteins. [Review]. Annu. Rev. Biochem. **63**:487-526.
18. **Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, and K. Wüthrich**. 1994. Homeodomain-DNA recognition. Cell **78**:211-223.
19. **Grépin, C., L. Dagnino, L. Robitaille, L. Haberstroh, T. Antakly, and M. Nemer**. 1994. A hormone-encoding gene identifies a pathway for cardiac but not skeletal muscle gene transcription. Mol. Cell. Biol. **14**:3115-3129.
20. **Guazzi, S., M. Price, M. De Felice, G. Damante, M. G. Mattei, and R. Di Lauro**. 1990. Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a

novel DNA binding specificity. EMBO J. **9**:3631-3639.

21. **Ip, H. S., D. B. Wilson, M. Heikinheimo, Z. Tang, C. N. Ting, M. C. Simon, J. M. Leiden, and M. S. Parmacek.** 1994. The GATA-4 transcription factor transactivates the cardiac muscle-specific troponin C promoter-enhancer in nonmuscle cells. Mol. Cell. Biol. **14**:7517-7526.
22. **Jimenez, F., L. E. Martin-Morris, L. Velasco, H. Chu, J. Sierra, D. R. Rosen, and K. White.** 1995. vnd, a gene required for early neurogenesis of Drosophila, encodes a homeodomain protein. EMBO J. **14**:3487-3495.
23. **Johnson, J. E., B. J. Wold, and S. D. Haushchka.** 1989. Muscle creatine kinase sequence elements regulating skeletal and cardiac muscle expression in transgenic mice. Mol. Cell. Biol. **9**:3393-3399.
24. **Kelly, R., S. Alonso, S. Tajbakhsh, G. Cossu, and M. Buckingham.** 1995. Myosin light chain 3F regulatory sequences confer regionalized cardiac and skeletal muscle expression in transgenic mice. J. Cell Biol. **129**:383-396.
25. **Kim, Y. and M. Nirenberg.** 1989. Drosophila NK-homeobox genes. Proc. Natl. Acad. Sci. USA **86**:7716-7720.
26. **Kimura, S., Y. Hara, T. Pineau, P. Fernandez-Salguero, C. H. Fox, J. M. Ward, and F. J. Gonzalez.** 1996. The *T/ebp* null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. Genes Dev. **10**:60-69.
27. **Komuro, I. and S. Izumo.** 1993. Csx: a murine homeobox-containing gene specifically expressed in the developing heart. Proc. Natl. Acad. Sci. USA **90**:8145-8149.
28. **Liew, C. C., M. J. Sole, K. Yamauchi-Takihara, B. Kellam, D. H. Anderson, L. P. Lin, and J. C. Liew.** 1990. Complete sequence and organization of the human cardiac beta-myosin heavy chain gene. Nucleic Acids Res. **18**:3647-3651.
29. **Lints, T. J., L. M. Parsons, L. Hartley, I. Lyons, and R. P. Harvey.** 1993. Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. Dev. **119**:419-431.
30. **Lyons, G. E.** 1994. In situ analysis of the cardiac muscle gene program during embryogenesis. Trends Cardiovasc. Med. **4**:70-77.

31. **Lyons, I., L. M. Parsons, L. Hartley, R. Li, J. E. Andrews, L. Robb, and R. P. Harvey**. 1995. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev. **9**:1654-1666.

32. **McBride, K., L. Robitaille, S. Tremblay, S. Argentin, and M. Nemer**. 1993. Fos/jun repression of cardiac-specific transcription in quiescent and growth stimulated myocytes is targeted at a tissue-specific cis-element. Mol. Cell. Biol. **13**:600-612.

33. **Miwa, T. and L. Kedes**. 1987. Duplicated CArG box domains have positive and mutually dependent regulatory roles in expression of the human alpha-cardiac actin genes. Mol. Cell. Biol. **7**:2803-2813.

34. **Molkentin, J. D., R. S. Brogan, S. M. Jobe, and B. E. Markham**. 1993. Expression of the alpha-myosin heavy chain gene in the heart is regulated in part by an E-box-dependent mechanism. J. Biol. Chem. **268**:2602-2609.

35. **Molkentin, J. D., D. V. Kalvakolanu, and B. E. Markham**. 1994. Transcription factor GATA-4 regulates cardiac muscle-specific expression of the a-myosin heavy-chain gene. Mol. Cell. Biol. **14**:4947-4957.

36. **Molkentin, J. D. and B. E. Markham**. 1994. An M-CAT binding factor and an RSRF-related A-rich binding factor positively regulate expression of the alpha-cardiac myosin heavy-chain gene in vivo. Mol. Cell. Biol. **14**:5056-5065.

37. **Moss, J. B., T. C. McQuinn, and R. J. Schwartz**. 1994. The avian cardiac alpha-actin promoter is regulated through a pair of complex elements composed of E boxes and serum response elements that bind both positive- and negative-acting factors. J. Biol. Chem. **269**:12731-12740.

38. **Nardelli-Haefliger, D. and M. Shankland**. 1993. Lox10, a member of the NK-2 homeobox gene class, is expressed in a segmental pattern in the endoderm and in the cephalic nervous system of the leech Helobdella. Dev. **118**:877-892.

39. **Navankasattusas, S., H. Zhu, A. V. Garcia, S. M. Evans, and K. R. Chien**. 1992. A ubiquitous factor (HF-1a) and a distinct muscle factor (HF-1b/MEF-2) form an E-box-independent pathway for cardiac muscle gene expression. Mol. Cell. Biol. **12**:1469-1479.

40. **Nudel, U., J. M. Calvo, M. Shani, and Z. Levy**. 1984. The nucleotide sequence of a

rat myosin light chain 2 gene. Nucleic Acids Research **12**:7175-7186.

41. **Okkema, P. G. and A. Fire**. 1994. The Caenorhabditis elegans NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. Dev. **120**:2175-2186.

42. **Olson, E. N. and W. H. Klein**. 1994. bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes Dev. **8**:1-8.

43. **Otting, G., Y. Q. Qian, M. Billeter, M. Muller, M. Affolter, W. J. Gehring, and K. Wuthrich**. 1990. Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. **9**:3085-3092.

44. **Parsons, W. J., J. A. Richardson, K. H. Graves, R. S. Williams, and R. W. Moreadith**. 1993. Gradients of transgene expression directed by the human myoglobin promoter in the developing mouse heart. Proc. Natl. Acad. Sci. USA **90**:1726-1730.

45. **Price, M., D. Lazzaro, T. Pohl, M. G. Mattei, U. Ruther, J. C. Olivo, D. Duboule, and R. Di Lauro**. 1992. Regional expression of the homeobox gene Nkx-2.2 in the developing mammalian forebrain. Neuron **8**:241-255.

46. **Qasba, P., E. Lin, M. D. Zhou, and M. A. Siddiqui**. 1992. A single transcription factor binds to two divergent sequence elements with a common function in cardiac myosin light chain-2 promoter. Mol. Cell. Biol. **12**:1107-1116.

47. **Rindt, H., J. Gulick, S. Knotts, J. Neumann, and J. Robbins**. 1993. In vivo analysis of the murine beta-myosin heavy chain gene promoter. J. Biol. Chem. **268**:5332-5338.

48. **Scott, M. P., J. W. Tamkun, and G. W. I. Hartzell**. 1989. The structure and function of the homeodomain. Biochem. Biophys. Acta **989**:25-48.

49. **Seidman, C. E., E. V. Schmidt, and J. G. Seidman**. 1991. cis-Dominance or rat atrial natriuretic factor gene regulatory sequences in transgenic mice. Can. J. Physiol. Pharmacol. **69**:1486-1492.

50. **Stewart, A. F. R., S. B. Larkin, I. K. G. Farrance, J. H. Mar, D. E. Hall, and C. P. Ordahl**. 1994. Muscle-enriched TEF-1 isoforms bind M-CAT elements from

muscle-specific promoters and differentially activate transcription. J. Biol. Chem. **269**:3147-3150.

51. **Tonissen, K. F., T. A. Drysdale, T. J. Lints, R. P. Harvey, and P. A. Krieg**. 1994. XNkx-2.5, a Xenopus gene related to Nkx-2.5 and tinman: evidence for a conserved role in cardiac development. Dev. Biol. **162**:325-328.

52. **Vincent, C. K., A. Gualberto, C. V. Patel, and K. Walsh**. 1993. Different regulatory sequences control creatine kinase-M gene expression in directly injected skeletal and cardiac muscle. Mol. Cell. Biol. **13**:1264-1272.

53. **Zeller, R., K. D. Bloch, B. S. Williams, R. J. Arceci, and C. E. Seidman**. 1987. Localized expression of the atrial natriuretic factor gene during cardiac embryogenesis. Genes Dev. **1**:693-698.

CHAPTER 3

The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors

Daniel Durocher, Frédéric Charron, René Warren, Robert J. Schwartz and Mona Nemer

Published in *EMBO Journal*, vol 16, pp 5687-5696 (1997)

In this study, we further examined how the NKE was contributing to basal cardiac ANF promoter activity. We emphasized this work on a possible functional interaction between GATA-4 and Nkx2-5, the NKE binding protein that was suggested by the close proximity of the NKE and the GATA element and by the evolutionary conservation of the phasing between these two elements. In this chapter, we indeed report a functional and physical interaction between GATA-4 and Nkx2-5.

This is a multi-authored work in which Frédéric Charron generated Figure 3.4B and the bulk of the GATA-4 deletions. René Warren was a summer student under my supervision who helped me setting up the conditions for the pull-down assay. Our collaborator, Robert Schwartz from the Baylor College of Medicine in Houston provided some important Nkx2-5 mutants. Finally, I generated all the other results and constructions as well as writing the paper under the supervision of Mona Nemer.

3.1. ABSTRACT

The tissue-restricted GATA–4 transcription factor and Nkx2-5 homeodomain protein are two of the earliest markers of precardiac cells. Although both proteins are essential for heart formation, neither one is sufficient to initiate cardiogenesis. However, overexpression of GATA–4 or Nkx2-5 enhances cardiac development in committed precursors suggesting that each of these proteins likely interacts with a cardiac cofactor. We have tested whether GATA–4 and Nkx2-5 are in fact cofactors for each other using transcription and binding assays. For this, we took advantage of the cardiac atrial natriuretic factor (ANF) promoter which is presently the only known target for Nkx2-5. Coexpression of GATA–4 and Nkx2-5 resulted in synergistic activation fo the ANF promoter in heterologous cells. The synergy, which requires the DNA binding site for both factors, involves physical interaction between Nkx2-5 and GATA-4 as evidenced *in vitro* and *in vivo*. This interaction maps to the carboxy-terminal zinc finger of GATA–4 and a C-terminus extension; similarly, a C-terminally extended homeodomain of Nkx2-5 is required for GATA–4 binding. The structure/function study suggests that binding of GATA–4 to the C-terminus autorepressive domain of Nkx2-5 may induce a conformational change that unmasks Nkx2-5 activation domains. Remarkably, the other myocardial GATA factor, GATA-6, is unable to substitute for GATA–4 for binding or functional interaction with Nkx2-5. Thus, the molecular interaction, which appears to have been evolutionary conserved between specific members of the GATA zinc finger family and Nkx2-5, may impart functional specificity to GATA factors and provide cooperative cross-talk between two pathways that are critical for the early events of cardiogenesis. Moreover, given the coexpression of GATA proteins and members of the NK2 class of homeodomain in other tissues, the GATA/Nkx partnership may represent a paradigm for transcription factor interaction during organogenesis.

3.2. INTRODUCTION

The GATA family of transcription factors are key developmental regulators that have been conserved throughout evolution (Fu and Marzluf, 1990; Spieth *et al.*, 1991; Platt *et al.*, 1996; Coffman *et al.*, 1996; Stanbrough *et al.*, 1995; Winick *et al.*, 1993). Various family members have been shown to alter transcription of target genes via binding to the consensus WGATAR sequence through a DNA-binding domain consisting of two adjacent zinc-fingers of the C2/C2 family. Sequence-specific DNA-binding requires the C-terminal zinc-finger and the N-terminal finger may stabilize the DNA-protein complex via electrostatic interactions with the phosphate backbone (Whyatt *et al.*, 1993). This DNA-binding domain is the most highly conserved region among the various GATA proteins. In vertebrates, six members have been identified so far and they can be divided into two subgroups based on sequence homology and tissue distribution. The first subgroup which includes GATA-1, -2, and -3 is largely restricted to the hematopoietic system where all three GATA factors have been shown to play essential, non-redundant functions (Fujiwara *et al.*, 1996; Pandolfi *et al.*, 1995; Ting *et al.*, 1996; Tsai *et al.*, 1994). Remarkably, arrested proerythroblasts lacking GATA-1 express several GATA-1 target genes although they are unable to achieve terminal erythroid differentiation (Weiss *et al.*, 1994) raising the possibility that GATA-2 - which is coexpressed with GATA-1 in proerythroblasts - may partially substitute for GATA-1. Consistent with this, GATA factors appear to be functionally interchangeable in some (Visvader *et al.*, 1995; Blobel *et al.*, 1995) but not all (Weiss *et al.*, 1994; Briegel *et al.*, 1993) *in vitro* assays. Taken together with the *in vivo* data, these results suggest that functional specificity of GATA proteins likely involves interactions with other cell restricted cofactors. Consistent with this hypothesis, GATA-1 was found to interact with the erythroid specific LIM protein RBTN2 and to be present in complexes containing RBTN2 and the hematopoietic basic helix-loop-helix protein SCL/TAL1 (Osada *et al.*, 1995). GATA-1 was also shown to cooperate with the ubiquitous SP1 protein and with two other erythroid factors, the basic leucine zipper NFE-2 (Walters and Martin, 1992; Gong and Dean, 1993) and the zinc finger EKLF (Gregory *et al.*, 1996; Merika and Orkin, 1995) for transcriptional activation of erythroid promoters/enhancers.

At least, in the case of SP1 and EKLF, the interaction was also observed with GATA-2 and involved direct contact through the DNA-binding domains (Merika and Orkin, 1995). Thus, the identity of the proteins that serve as cofactor(s) to impart functional specificity of GATA proteins in the hematopoietic system remains essentially unknown.

The other subclass of vertebrates GATA factors includes GATA-4, -5, and -6 whose expression is restricted to the heart and gut (Grépin *et al.*, 1994; Arceci *et al.*, 1993; Jiang and Evans, 1996; Laverriere *et al.*, 1994; Kelley *et al.*, 1993). All three genes are transcribed at very early stages of Xenopus, avian and mouse cardiac development (Heikinheimo *et al.*, 1994; Jiang and Evans, 1996; Laverriere *et al.*, 1994; Kelley *et al.*, 1993; Morrisey *et al.*, 1996). Within the heart, transcripts for GATA-4, -5, and -6 are found in distinct cell types with GATA-5 mRNA predominantly found in endocardial cells while GATA-4 and -6 are present in the myocardium (Kelley *et al.*, 1993; Grépin *et al.*, 1994; Morrisey *et al.*, 1996). The spatial and temporal expression of GATA-4 together with various functional studies are consistent with an important role of this GATA family member in cardiogenesis. Thus, GATA-4 was found to be a potent transactivator of several cardiac specific promoters (Grépin *et al.*, 1994; Ip *et al.*, 1994; Molkentin *et al.*, 1994); inhibition of GATA-4 expression in an *in vitro* model of cardiogenesis altered survival of precardiac cells and inhibited terminal cardiomyocyte differentiation (Grépin *et al.*, 1995; Grépin *et al.*, 1997). Moreover, targeted inactivation of the GATA-4 gene blocks formation of the primitive heart tube indicating a crucial role for GATA-4 in heart development (Kuo *et al.*, 1997; Molkentin *et al.*, 1997). However, ectopic expression of GATA-4 is not sufficient to initiate cardiac differentiation or to activate the cardiac genetic program although it markedly potentiates cardiogenesis (Jiang and Evans, 1996; Grépin *et al.*, 1997) suggesting cooperative interaction between GATA-4 and other cardiac transcription factors.

Genetic studies in *Drosophila melanogaster* have identified the gene *Tinman* as a key regulator of heart differentiation. In *tin* embryos, flies lack the dorsal vessel, the fly structure homologous to the heart, as a result of defects in late mesoderm specification (Bodmer *et al.*, 1990; Bodmer, 1993; Azpiazu and Frasch, 1993). Presumptive homologues of *Tinman* have been cloned in vertebrates (Nkx2-3, Nkx2-5 and Nkx2-7) and are expressed in the myocardium (reviewed in (Harvey, 1996)(Lints *et al.*, 1993; Tonissen *et al.*, 1994;

Evans *et al.*, 1995; Schultheiss *et al.*, 1995; Lee *et al.*, 1996; Chen and Fishman, 1996; Buchberger *et al.*, 1996). Targeted disruption of the Nkx2-5 gene in mice leads to embryonic death due to cardiac morphogenetic defects (Lyons *et al.*, 1995). However, gain-of-function studies in zebrafish *Danio rerio* and *Xenopus laevis* indicate that ectopic expression of Nkx2-5 results in enhanced myocyte recruitment but is not sufficient to initiate cardiac gene expression or differentiation (Chen and Fishman, 1996; Cleaver *et al.*, 1996) suggesting that Nkx2-5 acts in concert with other transcription factors to specify the cardiac phenotype.

Since the cardiac-specific atrial natriuretic factor (ANF) promoter is a transcriptional target for both GATA-4 and Nkx2-5 (Grépin *et al.*, 1994; Durocher *et al.*, 1996), we used it to investigate functional cooperation between Nkx2-5 and GATA-4 in the heart. We present data showing that Nkx2-5 and GATA-4 specifically cooperate in activating ANF and other cardiac promoters, and physically interact both *in vitro* and *in vivo*. This molecular interaction provides the interesting possibility that instead of being part of the same epistatic group, the two pathways collaborate in the early events of cardiogenesis.

3.3. RESULTS

GATA-4 and Nkx2-5 synergistically activate cardiac transcription. Recent studies from our laboratory have demonstrated that two cardiac-specific transcriptional pathways converge on the atrial natriuretic factor (ANF) promoter (Durocher *et al.*, 1996; Grépin *et al.*, 1994), ANF being the major secretory product of embryonic and postnatal cardiomyocytes. The region of the ANF promoter which is essential for high basal cardiac activity (Argentin *et al.*, 1994) harbors a GATA element located at -120 bp in the rat promoter which binds with high affinity all the members of the cardiac GATA subfamily (Charron et al., in preparation), and the NKE which binds Nkx2-5 and is required for ANF promoter and enhancer function (Durocher *et al.*, 1996). As seen in Fig. 3.1A, the nucleotide composition of these elements as well as their phasing are conserved across species suggesting an evolutionary pressure to maintain important regulatory pathways.

This led us to investigate whether GATA proteins and Nkx2-5 could functionally interact at the level of the ANF promoter. We tested this hypothesis by cotransfecting GATA-4 and Nkx2-5 expression vectors in non-cardiac cells (HeLa cells) at limiting DNA concentrations (Durocher *et al.*, 1996; Grépin *et al.*, 1994) in order to avoid squelching. Under these conditions, GATA-4 and Nkx2-5 were able to activate synergistically the ANF promoter but not control promoters lacking GATA and NKE sites such as TK109 (Fig. 3.1B). This cooperative response was not caused by transactivation of the CMV promoter which drives Nkx2-5 and GATA-4 expression since the co-expression of both vectors does not alter the level of either Nkx2-5 or GATA-4 protein (Fig. 3.6B).

The relevance of this synergy to cardiac transcription was further assessed by cotransfecting Nkx2-5 and GATA-4 with other cardiac promoters including ANF, β-MHC, α-MHC and the BNP reporters. Under the conditions used in Figure 3.2A, a subset of promoters that contain both NKE and GATA elements could be synergistically activated by the combination of Nkx2-5 and GATA-4 (Fig. 3.2A). Thus, BNP promoter which is a GATA-4 target (Argentin *et al.*, 1994) responds synergistically to Nkx2-5 and GATA-4. Interestingly, sequences with high homology to the NKE are present around -385 and -437 bp and are conserved across species; promoter fragments lacking these putative NKEs are no longer responsive to Nkx2-5/GATA synergy (data not shown). On the other hand, neither additive nor synergistic effects were observed on the α-MHC and the β-MHC promoters in response to Nkx2-5 and GATA-4 at all different DNA concentrations tested (Fig. 3.2A and data not shown). These data suggest that only a subgroup of cardiac genes are targeted by both transcription factors and that both NKE and GATA sites are required for synergy. This hypothesis was further tested using ANF promoter mutants deleted of either the GATA or the NKE elements. As shown in Figure 3.2B, there appears to be an absolute requirement for both elements to achieve synergy. The same results were obtained using BNP promoters containing only GATA sites or heterologous promoters with multimerized GATA elements (data not shown). This suggests that, in natural promoters, both proteins have to be recruited at the promoter or require a conformational change induced upon DNA-binding. Indeed, a GATA-4 mutant that no longer binds DNA because one of the zinc-coordinating cysteins in the carboxy-terminal zinc finger was mutated, no

longer supports Nkx2-5 synergy (Fig. 3.2C).

Since multiple GATA and homeobox proteins are expressed in the heart, we investigated the specificity of the synergy. In cotransfection assays using the proximal -135 bp ANF promoter as reporter, we found that Nkx2-5 was able to cooperatively activate transcription of the ANF reporter only with GATA-4 and GATA-5 (Fig 3.3A). No synergy was observed with either GATA-1 or GATA-6. Since GATA-6 is as potent as GATA-4 in transactivating the ANF promoter (our unpublished data), the results suggest that transcriptional cooperativity between Nkx2-5 and GATA proteins requires specific molecular/structural determinants on the GATA-4, -5 proteins. The same approach was used to identify homeoproteins that could cooperate with GATA-4 including other NK2 proteins (TTF-1/Nkx2-1) (Lints *et al.*, 1993; Guazzi *et al.*, 1990), Hox proteins (HoxB3), Pou proteins (Oct1) (Sturm *et al.*, 1988) or *bicoid*-related homeoboxes (Ptx1) (Lamonerie *et al.*, 1996). Transcriptional cooperativity was observed with the members of the *Antp* superfamily Nkx2-5, TTF-1, and HoxB3 but not with Oct1 and Ptx1 (Fig 3.3B). These results suggest that the *Antp*-type homeodomain plays an important role in the specificity of the synergy.

We then used deletion mutants of both Nkx2-5 and GATA-4 to map the domains involved in synergy over the ANF promoter. All mutant proteins were tested for expression level and nuclear localization (data not shown). The analyses revealed that, in addition to the DNA-binding domain (Fig. 3.2C), two GATA-4 domains are required for the synergy, one located within the N-terminal 119 amino acids of the protein, and the second in the C-terminal 121 amino acids (Fig. 3.4). Both domains contain GATA-4 activation domains although the presence of either domain is sufficient for transactivation of GATA-dependent promoters in heterologous cells (Fig. 3.4B). The observation that both domains are required for cooperative interaction with Nkx2-5, suggest that synergistic interaction between these two domains may be required for Nkx2-5-induced transcriptional activation of the ANF promoter or that each domain fulfills a distinct function. Mutational analysis of Nkx2-5 showed that, while the homeodomain is critical for Nkx-GATA synergy, domains outside the homeobox, particularly sequences C-terminus of the homeodomain, are also important (Fig. 3.5A). Thus, neither the homeodomain (122-203), nor in fusion with the N-

FIGURE 3.1.

Nkx2-5 and GATA-4 can cooperate transcriptionally. (**A**) Structural organization of the proximal ANF promoter. Regulatory elements of the ANF promoter are boxed, and their location relative to the transcription start are indicated. The PERE corresponds to the phenylephrine response element (Ardati and Nemer, 1993), the NKE to the NK2 response element (Durocher *et al.*, 1996). The Nkx2-5 binding site and the GATA elements of the ANF promoter are conserved among species (rat, human, bovine and mouse promoters) and their spacing is conserved (20 bp, 2 turns of the DNA double helix). (**B**) GATA-4 and Nkx2-5 synergistically activate the ANF promoter. HeLa cells were transiently cotransfected as described in the Methods section using CMV-driven expression vectors for GATA-4 and Nkx2-5 or the backbone vector as control (pCGN) in conjunction with either ANF-luciferase or Tk109-luciferase reporters. The results, expressed as fold induction of reporter constructs, are from one representative experiment (out of at least 4) and represent the mean of a duplicate.

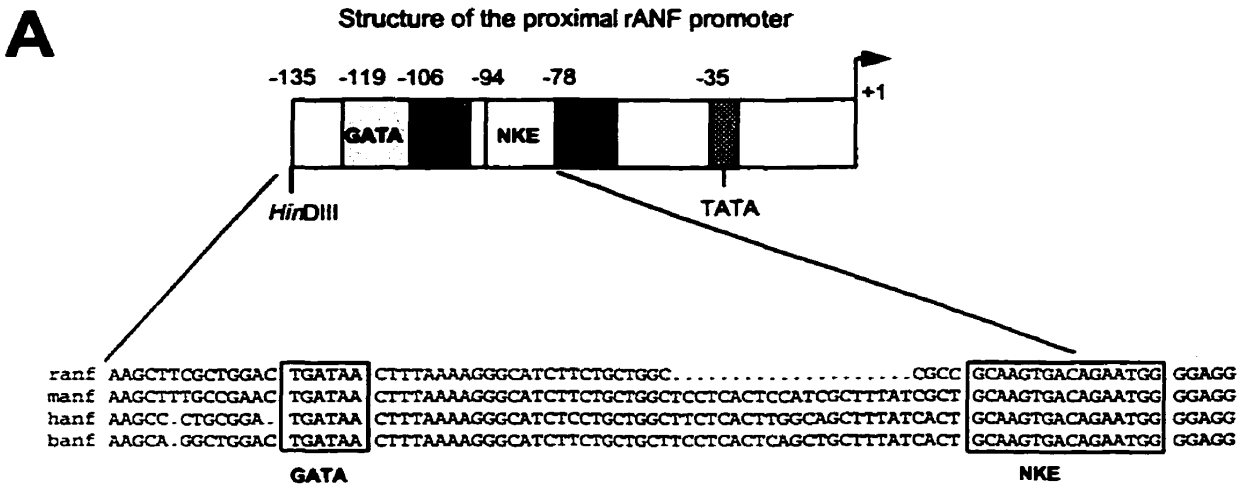
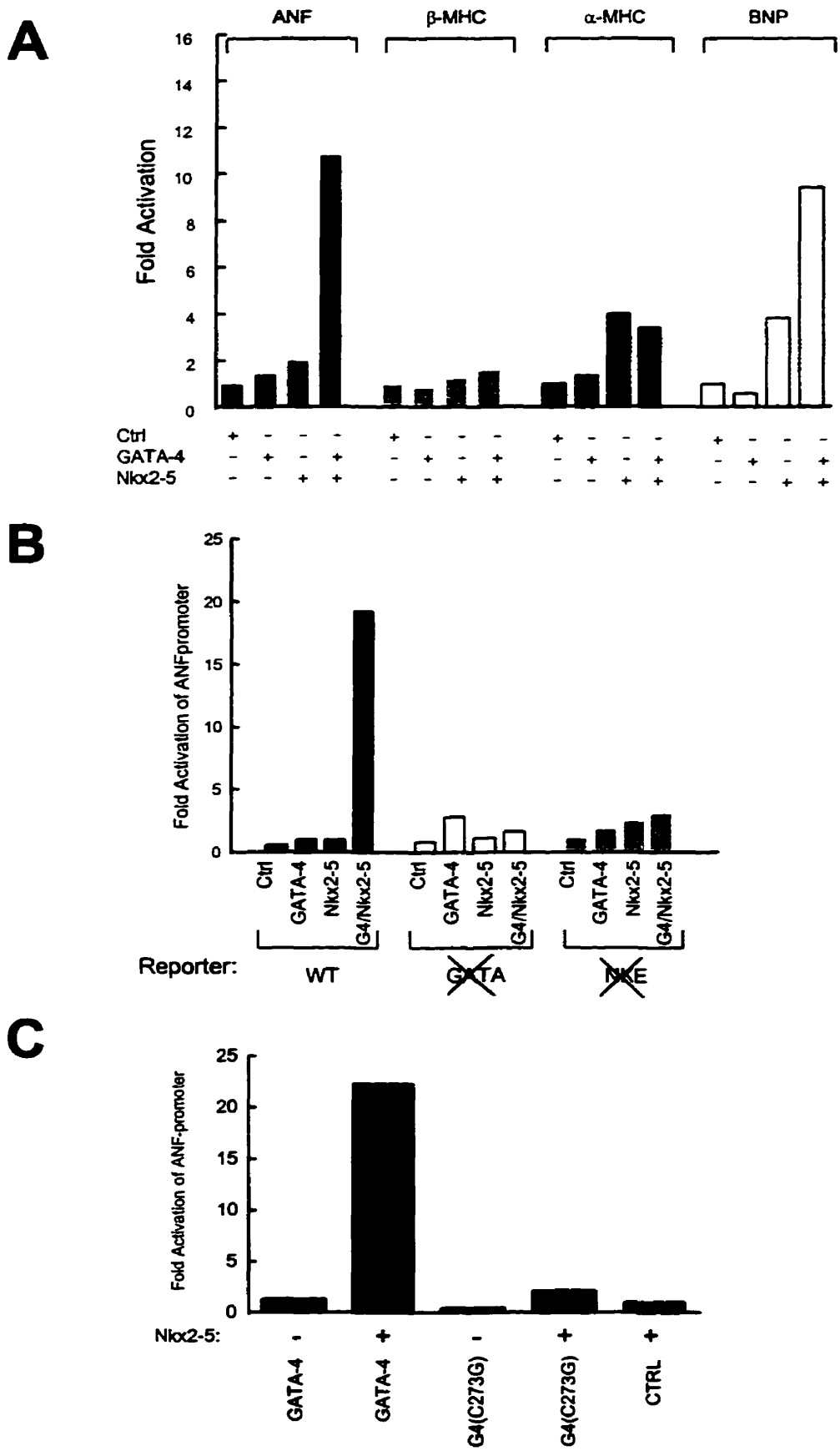

A
Structure of the proximal rANF promoter
-135
-119
-106
-94
-78
-35
+1
GATA
NKE
HinDIII
TATA
ranf AAGCTTCGCTGGAC TGATAA CTTTAAAAGGGCATCTTCTGCTGGC.....CGCC GCAAGTGACAGAATGG GGAGG
manf AAGCTTTGCCGAAC TGATAA CTTTAAAAGGGCATCTTCTGCTGGCTCCTCACTCCATCGCTTTATCGCT GCAAGTGACAGAATGG GGAGG
hanf AAGCC.CTGCGGA. TGATAA CTTTAAAAGGGCATCTCCTGCTGGCTTCTCACTTGGCAGCTTTATCACT GCAAGTGACAGAATGG GGAGG
banf AAGCA.GGCTGGAC TGATAA CTTTAAAAGGGCATCTTCTGCTGCTTCCTCACTCAGCTGCTTTATCACT GCAAGTGACAGAATGG GGAGG
GATA
NKE
B
ANF-luc
Fold Induction
30
25
20
15
10
5
0
Ctrl
GATA-4
Nkx2-5
G4/Nkx2-5
Tk109-luc
Fold Induction
30
25
20
15
10
5
0
Ctrl
GATA-4
Nkx2-5
G4/Nkx2-5

FIGURE 3.2.

A subset of cardiac promoters are synergistically activated by the Nkx2-5/GATA-4 combination. (**A**) HeLa cells were transiently cotransfected as described in Fig. 3.1 using various cardiac promoters linked to the luciferase reporter. ANF represents the rat ANF -135 construct; β-MHC, the rat -667 bp promoter; α-MHC, the rat -613 bp promoter whereas BNP represents the rat -2 kb promoter. (**B**) The synergy between GATA-4 and Nkx2-5 requires both binding sites in the context of the ANF promoter. Transient cotransfection in Hela cells were carried out as described in previous figures and the promoter described represents either the -135 bp promoter (WT), the Δ-106/-135 bp promoter which removes the GATA element and the Δ -57/-106 bp promoter which removes the NKE site. (**C**) GATA-4 binding to DNA is required for synergy. A point mutant of GATA-4 (C273G) which does not bind DNA *in vitro* and which cannot activate GATA-dependent promoters was used in a cotransfection assay with or without Nkx2-5.


A
ANF
β-MHC
α-MHC
BNP
Fold Activation
16
14
12
10
8
6
4
2
0
Ctrl
GATA-4
Nkx2-5
B
Fold Activation of ANFpromoter
25
20
15
10
5
0
Ctrl
GATA-4
Nkx2-5
G4/Nkx2-5
Reporter:
WT
GATA
NKE
C
Fold Activation of ANF-promoter
25
20
15
10
5
0
Nkx2-5:
GATA-4
GATA-4
G4(C273G)
G4(C273G)
CTRL

terminal regions of Nkx2-5 (1-203) is able to stimulate GATA-4 activity. Deletion of the entire C-terminal region totally impairs the ability of Nkx2-5 to stimulate GATA-4 transcription while partial deletions of the C-terminus (1-246 and Δ204-246) reduce consistently the extent of synergy observed whitout completely abolishing it. This result suggests that these two regions of the C-terminus are only partially redundant or that the functional interaction between Nkx2-5 and GATA-4 requires an «extended» homeodomain in the C-terminus. The C-terminus is not known as a transcriptional activation domain, in fact, it appears to be an autorepression domain since its deletion leads to superactivation (Fig. 3.5B); thus, the requirement for the C-terminus suggests that GATA-4 physically interacts with Nkx2-5 to cause a conformational change and derepress (or unmask) Nkx2-5 activation domains.

GATA-4 and Nkx2-5 physically interact *in vitro* and *in vivo*. We first assessed possible physical interaction between Nkx2-5 and GATA-4 using pull-down assays with immobilized MBP-Nkx2-5 and *in vitro* translated, ^{35}S-labeled GATA-4 (Fig. 3.6A). MBP-Nkx2-5 was able to retain GATA-4 specifically while a MBP-LacZ control fusion could not retain GATA-4 on the column and the immobilized Nkx2-5 could not interact with labeled control luciferase (Fig. 3.6A). The interaction between GATA-4 and Nkx2-5 was also observed in the presence of ethidium bromide, suggesting that this interaction occurs without DNA bridging (data not shown). In order to ascertain the *in vivo* relevance of this interaction, co-immunoprecipitations were performed on nuclear extracts from 293 cells transfected with expression vectors for wild-type GATA-4 or HA-tagged Nkx2-5 alone or in combination. Nuclear extracts from these transfected cells were then incubated with the monoclonal antibody 12CA5 which recognizes the HA epitope. Immunocomplexes were separated on SDS-PAGE, subjected to Western blotting and visualized with the anti-GATA-4 antibody. As seen in Figure 3.6B, GATA-4 protein was precipitated by the 12CA5 mAb solely when both proteins were expressed, implying either a direct or indirect contact with Nkx2-5.

Next, deletion mutants of GATA-4 were generated in order to map the region(s) of GATA-4 protein involved in physical interaction with Nkx2-5. Figure 3.7A displays the results of the binding studies and the left panel of Figure 3.8A shows the structure of the

deletion mutants and summarizes their behavior in pull-down and transfection assays. The Nkx2-5 interaction domain seems to map to the second zinc-finger and a C-terminal basic region that is not part of any known activation domain of GATA-4. This localization is consistent with the observation that the physical interaction requires zinc ions since pull-down assays in the presence of chelating agents (EDTA and phenanthroline) abolish the interaction (data not shown). Unfortunately the requirement of the Nkx2-5 binding domain for the synergy could not be assessed since it is part of a basic region essential for the nuclear targeting of GATA-4 (Charron et al., unpublished results). Interestingly, neither GATA-1 nor GATA-6, which do not transcriptionally synergize with Nkx2-5, could be retained on the MBP-Nkx2-5 column suggesting that physical interaction is required for functional cooperativity.

The same approach was also used to map the GATA-4 interaction domain on Nkx2-5. A series of Nkx2-5 deletion mutants were bacterially produced in fusion with MBP, quantified and assayed for their ability to interact with ^{35}S-labeled GATA-4. The results of these binding assays revealed that both the homeodomain and its C-terminal region are required for physical interaction (Fig. 3.7B and 3.8A). The homeodomain by itself (122-203) or extended to contain the N-terminal domain (1-203) were insufficient for physical interaction. However, when the homeodomain was fused to parts of the C-terminal (1-246 and Δ204-246) the fusion proteins regained the ability to bind GATA-4 suggesting that the C-terminal extension provided an essential docking site for the GATA protein or was required for the proper folding of the homeodomain. It is noteworthy that these results are in complete agreement with the transfection data and indicate that the determinants of Nkx2-5 and GATA-4 interaction reside mostly in the homeodomain and a C-terminal extension. Collectively, the results also suggest that functional synergy between Nkx2-5 and GATA-4 requires physical interactions of the two proteins.

3.4. DISCUSSION

Transcription factors GATA–4 and Nkx2-5 are two of the earliest markers of precardiac cells and, as evidenced by gene inactivation studies (Molkentin *et al.*, 1997; Kuo

FIGURE 3.3.

The synergy is specific for a subset of cardiac GATA proteins and *Antp*-type homeoproteins. **(A)** Cotransfection assays in HeLa cells using various GATA expression vectors were done in the presence (+) or absence (-) of the Nkx2-5 expression vector. Ctrl represents the backbone vector for most of the GATA constructs (pCGN). **(B)** Cotransfection assays in HeLa cells were performed in the presence of various homeodomain protein expression vectors in absence (-) or in presence (+) of GATA-4 (+, 0.1 μg; ++, 0.25 μg).

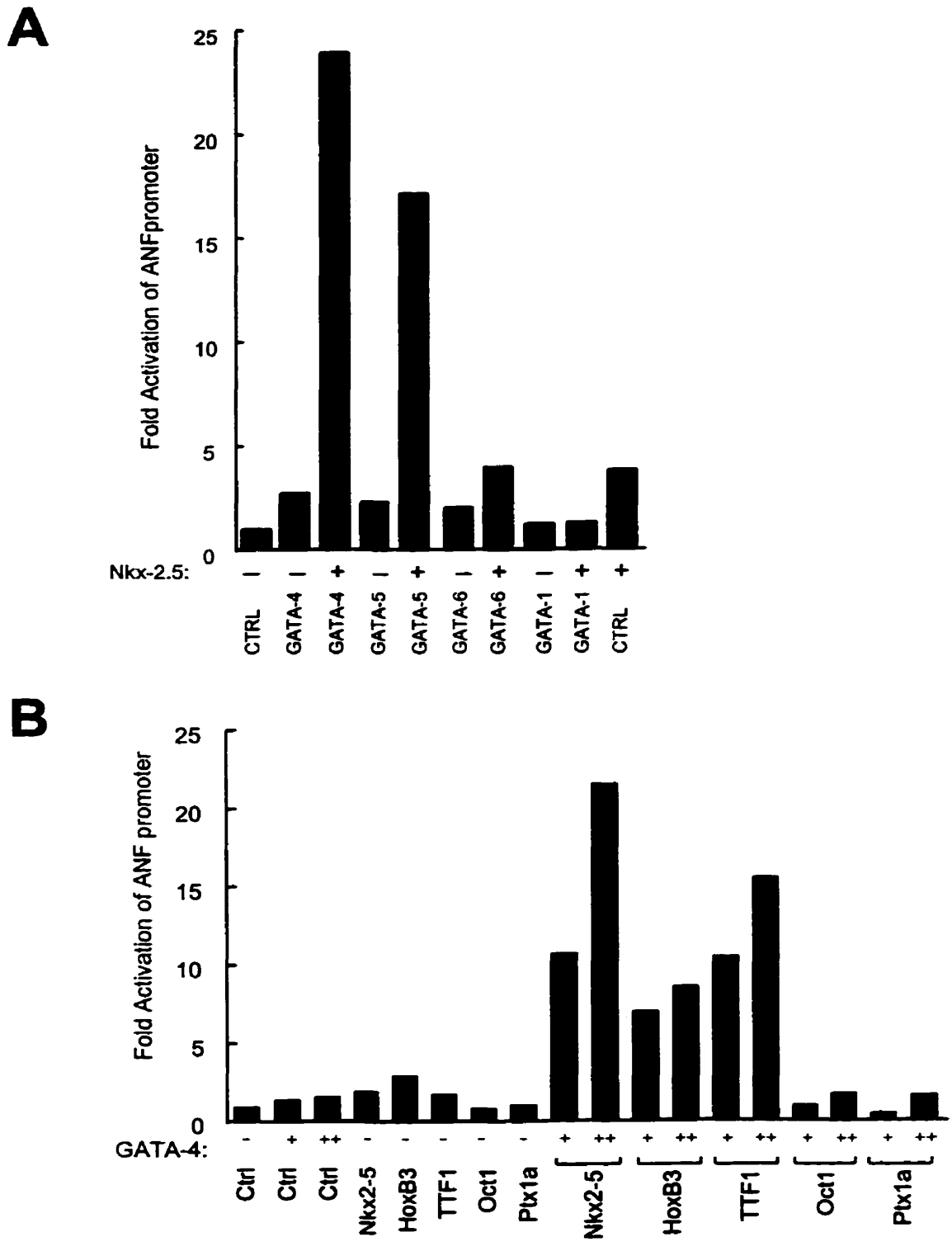
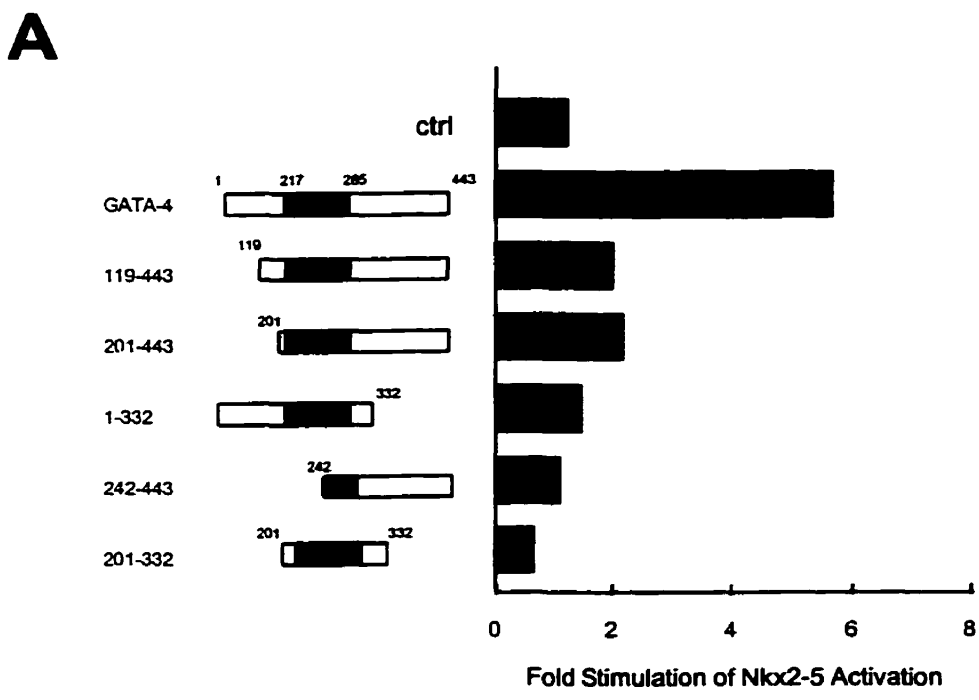

A
Fold Activation of ANFpromoter
25
20
15
10
5
0
Nkx-2.5: – – + – + – + – + +
CTRL
GATA-4
GATA-4
GATA-5
GATA-5
GATA-6
GATA-6
GATA-1
GATA-1
CTRL
B
Fold Activation of ANF promoter
25
20
15
10
5
0
GATA-4: - + ++ - - - - - + ++ + ++ + ++ + ++ + ++
Ctrl
Ctrl
Ctrl
Nkx2-5
HoxB3
TTF1
Oct1
Ptx1a
Nkx2-5
HoxB3
TTF1
Oct1
Ptx1a

FIGURE 3.4.

The synergy requires both activation domains of GATA-4. A) GATA-4 vectors (50 ng/35mm dish) expressing truncated GATA-4 proteins able to translocate to the nucleus were used in cotransfection assays with or without the Nkx2-5 expression vector. *Ctrl* represents the backbone vector. The results are expressed as fold stimulation of Nkx2-5 activation (equivalent to Fold synergy where the value of 1 represents no synergy, i.e. the ratio between the activity of the reporter in the presence of the GATA deletion mutant plus Nkx2-5 over the activity of the reporter only in the presence of the GATA deletion mutant). B) GATA-4 activation domains are located both at the C- and N-termini. GATA-4 vectors were transfected in HeLa cells at the dose of 0.2 μg/dish with the ANF-135 bp-luciferase reporter.


A
ctrl
GATA-4
1
217
285
443
119-443
119
201-443
201
1-332
332
242-443
242
201-332
201
332
0
2
4
6
8
Fold Stimulation of Nkx2-5 Activation

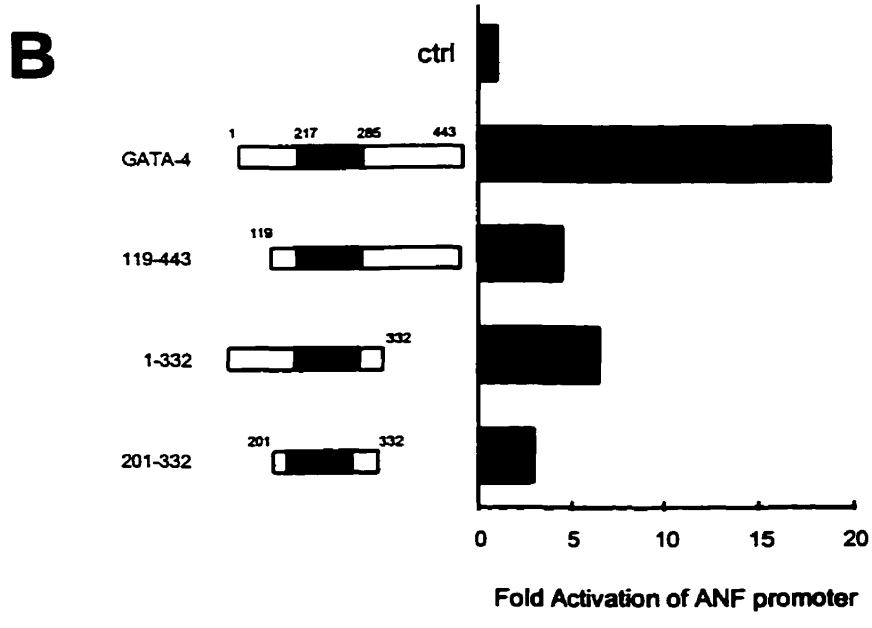
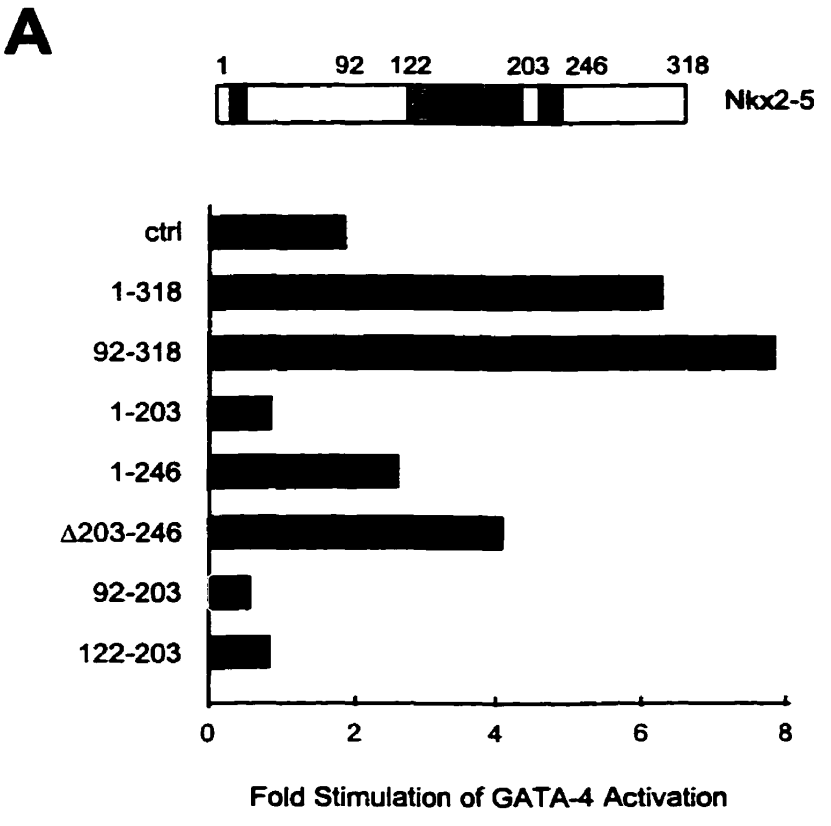
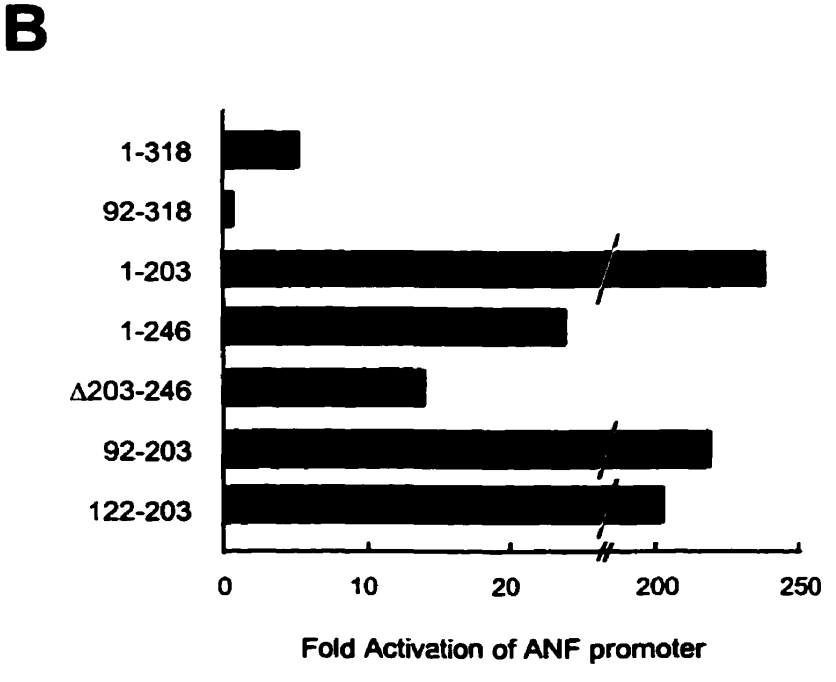

B
ctrl
GATA-4
1
217
285
443
119-443
119
1-332
332
201-332
201
332
0
5
10
15
20
Fold Activation of ANF promoter

FIGURE 3.5.

The synergy requires the C-terminus of Nkx2-5. **(A)** CMV-driven vectors, expressing various deletion of Nkx2-5, were used in cotransfection assays as described in Fig 3.4A, whereNkx2-5 concentration was kept at (0,5 μg/35 mm dish). *Ctrl* represents the backbone vector without insert. The data is expressed as fold stimulation of GATA-4 activation which is calculated by the ratio of the reporter activation when GATA-4 and Nkx2-5 expression vectors are present over the reporter activation when GATA-4 alone is present. **(B)** The C-terminus domain of Nkx2-5 is an auto-inhibitory domain that masks an activation domain located N-terminal of the homeodomain. Cotransfections in HeLa cells were carried out with an optimal dose of pCGN-Nkx2-5 constructs (2 μg/dish) on the ANF-135 bp promoter. The results depict the mean of six independent experiments.


A
1
92
122
203
246
318
Nkx2-5
ctrl
1-318
92-318
1-203
1-246
Δ203-246
92-203
122-203
0
2
4
6
8
Fold Stimulation of GATA-4 Activation


B
1-318
92-318
1-203
1-246
Δ203-246
92-203
122-203
0
10
20
200
250
Fold Activation of ANF promoter

et al., 1997; Lyons *et al.*, 1995), both play critical roles in cardiogenesis. The data presented here show that GATA–4 and Nkx2-5 interact physically and synergistically to activate cardiac transcription suggesting functional convergence of two critical cardiac pathways.

Modulation of Nkx2-5 activity by GATA-4. Members of the GATA family of transcription factors (GATA-1, -2, and -3) have been shown to interact with other classes of nuclear proteins containing Lim domain (Osada *et al.*, 1995), zinc finger (Merika and Orkin, 1995; Gregory *et al.*, 1996), and basic leucine zipper (Walters and Martin, 1992; Gong and Dean, 1993; Kawana *et al.*, 1995) motifs. The present work demonstrates that GATA factors are also able to interact with homeodomain containing proteins of the NK2 and *Antp* type. This GATA/Nkx interaction is so far the first example of zinc finger/homeodomain interaction in vertebrates. The only other known zinc finger/homeodomain cooperation is in *Drosophila*, where it was recently shown that the orphan nuclear receptor αFtz-F1 is a cofactor for the homeodomain protein Ftz (Guichet *et al.*, 1997; Yu *et al.*, 1997); in this case, the physical association between αFtz-F1 and Ftz is thought to enhance the binding of Ftz to its lower affinity target sequences (Yu *et al.*, 1997; Guichet *et al.*, 1997) much in the same way that Extradenticle and Pbx modulate the DNA binding activity of Hox proteins (Phelan *et al.*, 1995; Lu and Kamps, 1996; Peltenburg and Murre, 1997). The interaction of GATA–4 with Nkx2-5 does not appear to result in cooperative DNA binding since neither protein appears to alter the affinity or sequence specificity of the other; moreover, the presence of both GATA and NKE sites does not enhance either GATA–4 or Nkx2-5 binding to their sites as evidenced by gel shift assays using nuclear extracts containing both proteins or each one separate (data not shown). Instead, the data suggest that GATA–4 interaction with Nkx2-5 serves to unmask the activation domains of Nkx2-5 as illustrated in Figure 3.8B; this would be reminiscent of the Extradenticle-induced conformational change, that switches Hox proteins from repressors to activators (Chan *et al.*, 1996; Peltenburg and Murre, 1997).

The region of GATA–4 that contacts Nkx2-5 spans the second zinc finger and a ≈ 40 amino acid C terminal extension (Fig. 3.8A, left panel). This represents a highly conserved segment among the cardiac GATA–4, -5, and -6 proteins with an overall 85-95%

homology; notable differences between GATA-4 and -6 (but not GATA-4 and -5) that may account for the differential interaction with Nkx2-5 are found in the hinge region (AA 243-270) preceding the second zinc finger and three non-conservative changes that affect phosphorylatable residues (H244S, N250S, S262P). The differential interaction of GATA proteins with Nkx2-5 reveals for the first time differences between GATA proteins in an *in vitro* assay.

Is Nkx2-5 a specificity cofactor for GATA–4? Two GATA proteins, GATA–4 and -6, are present in the myocardium and both are potent activators of cardiac transcription. However, inactivation of the GATA–4 gene arrests cardiac development at a very early stage despite marked upregulation of GATA-6 arguing for specificity of GATA–4 and -6 function (Kuo *et al.*, 1997; Molkentin *et al.*, 1997). The upregulation of GATA-6 might, at least partially, account for ANF expression in presumed cardiogenic cells of the GATA-4$^{-/-}$ embryos (Molkentin *et al.*, 1997) much like upregulation of GATA-2 in GATA-1$^{-/-}$ pre-erythroblasts might explain globin gene expression in the absence of GATA-1. It should be however pointed out that ANF transcription is controlled by multiple pathways in complex spatio-temporal manners (Argentin *et al.*, 1994; Durocher *et al.*, 1996); thus, presence of ANF transcripts in GATA-4$^{-/-}$ cells may reflect activation or maintenance of more complex compensatory pathways; moreover, GATA-5 which can also cooperate with Nkx2-5 in the heart does not seem to be restricted to the endocardial cells before the primitive heart tube stage (Morrisey *et al.*, 1997). The exact reason for which GATA-4 is obviously essential for primitive heart developement must await further biochemical and genetic studies. Nevertheless, the available evidence clearly indicates that GATA-6 is unable to fully substitute for GATA-4 with respect to cardiogenesis. Similarly, despite their seemingly interchangeable role in some *in vitro* assays (Visvader *et al.*, 1995; Blobel *et al.*, 1995), the hematopoietic members of the GATA family are clearly non redundant (Fujiwara *et al.*, 1996; Pandolfi *et al.*, 1995; Ting *et al.*, 1996; Tsai *et al.*, 1994). Unfortunately, the molecular basis for GATA factor specificity has yet to be unraveled. The data presented suggest that interaction of GATA proteins with other tissue-restricted transcription factors might be the underlying mechanism for functional specificity of the GATA family members. Thus, Nkx2-5 may be the specificity cofactor for GATA–4 while

other homeodomain proteins of the NK2 or *Antennapedia* class may fulfill a similar function for GATA-6 in the myocardium.

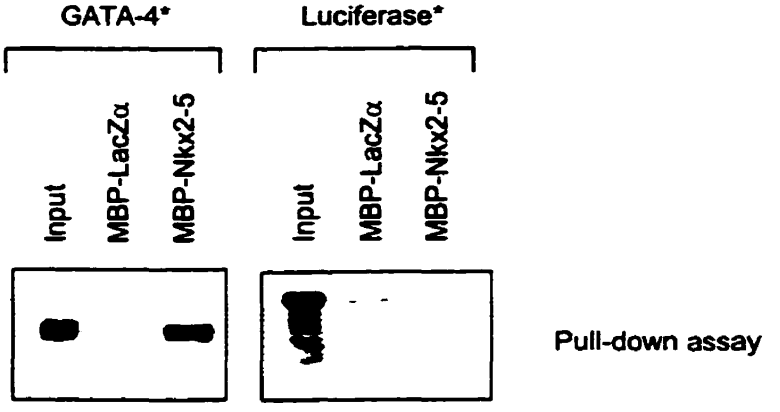
The presence of a cofactor for GATA proteins is likely the case for the hematopoietic system. Indeed, in a recent publication, Weiss et al. demonstrated that the presence of the GATA-1 zinc fingers was essential for erythroid differentiation. Interestingly, the homologous region of GATA-3 (which is not co-expressed with GATA-1) but not the entire GATA-2 (which is upregulated in GATA-1-/- pre-erythroblasts) could functionally substitute for GATA-1 zinc fingers suggesting that interaction of zinc fingers with an - as yet unidentified - nuclear factor may be an important determinant for definitive erythropoiesis (Weiss *et al.*, 1997). Since GATA proteins and other members of the NK2 family are also coexpressed in other tissues such as spleen (GATA-5 and Nkx2-5), gut (GATA-5, -6 and Xbap) (Lints *et al.*, 1993; Morrisey *et al.*, 1996; Newman *et al.*, 1997; Morrisey *et al.*, 1997); it is thus tempting to speculate whether GATA/Nkx partnership may represent a paradigm for transcription factor interaction during cell fate determination.

It is noteworthy that, at least in cardiac muscle, such paradigm appears to have been evolutionary conserved. Indeed, in *Drosophila*, the cardiac promoter of the transcription factor D-mef which is a target for Tinman, contains juxtaposed GATA and NKE sites (Table 3.1); while the NKEs are necessary, they are not sufficient for cardiac expression raising the possibility of an interaction with other factors (Gajewski *et al.*, 1997). Moreover, in *C. elegans*, two members of the GATA family have been described [elt-1 and -2, (Spieth *et al.*, 1991; Hawkins and McGhee, 1995)] whose expression is found in gut and perhaps pharyngeal muscles, and GATA elements are necessary for tissue-specific transcription in those tissues (Egan *et al.*, 1995; Okkema and Fire, 1994). Moreover; at least one member of the NK2 family, CEH-22 , is also expressed in *C. elegans* pharyngeal muscle and has been implicated in activation of the muscle-specific myosin heavy chain (Myo-2) enhancer (Okkema and Fire, 1994). Interestingly, the Myo-2 enhancer requires the closely linked GATA and NKE sites (Table 3.1) for muscle expression (Okkema and Fire, 1994). Thus, at least in muscle cells the GATA and Nkx interactions appear to have been evolutionary conserved.

FIGURE 3.6.

GATA-4 and Nkx2-5 physically interact *in vitro* and *in vivo*. **(A)** GATA-4 and Nkx2-5 interact *in vitro*. Pull-down protein-protein binding assays were performed using immobilized, bacterially produced MBP fusions (MBP-Nkx2-5 and MBP-LacZ as control) and either ^{35}S-labelled GATA-4 or luciferase protein. After incubation, the protein complexes were spun, extensively washed and separated on a 10% SDS-PAGE. Labeled proteins were visualized and quantified by autoradiography on phosphor plates. **(B)** GATA-4 and Nkx2-5 interact *in vivo*. Nuclear extracts from the simian kidney cell line 293 transfected with either the backbone vectors (mock), GATA-4 expression vector alone (GATA-4), HA-tagged Nkx2-5 (Nkx2-5), or a combination of GATA-4 and HA-Nkx2-5 (GATA-4/Nkx2-5) were used for immunoprecipitation. 60 μg of nuclear extract were incubated with the mAb 12CA5 and precipitated with protein-G agarose. The top panel represents the immunocomplex separated by electrophoresis and blotted with an anti-GATA-4 polyclonal antibody. The bottom two panels represents western blots on the transfected nuclear extracts (20 μg) using either the anti-GATA-4 Ab (middle panel) or the anti-HA (12CA5) mAb. The white ghost bands are produced by the immuglobulin heavy chains that comigrate with GATA-4 on SDS-PAGE.

A


GATA-4*
Luciferase*
Input
MBP-LacZα
MBP-Nkx2-5
Input
MBP-LacZα
MBP-Nkx2-5
Pull-down assay

B

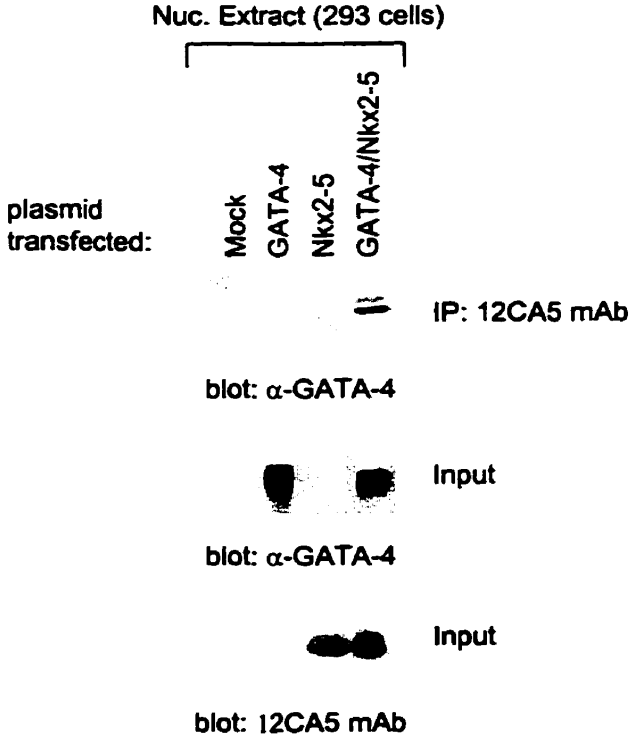
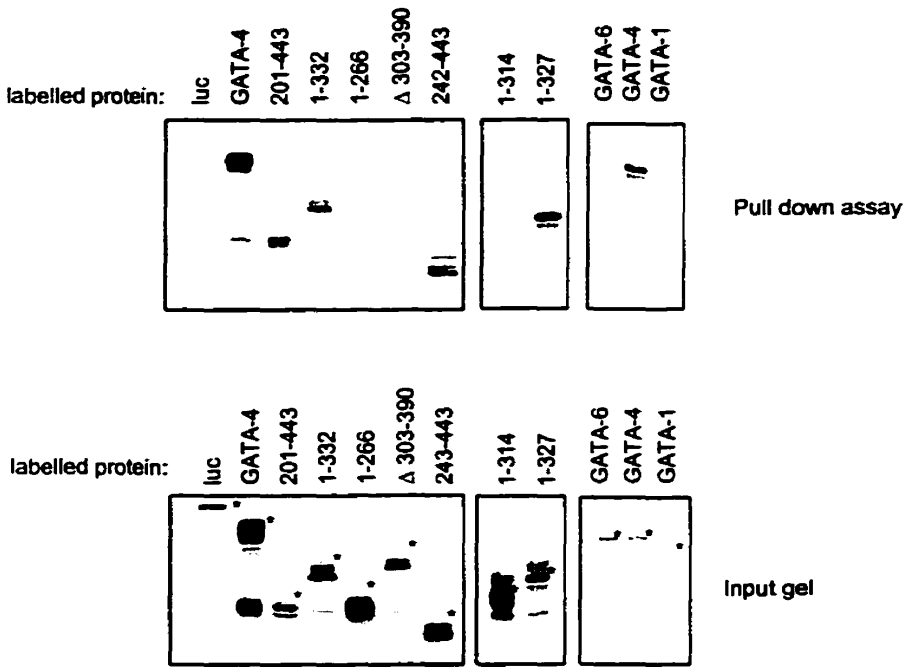

Nuc. Extract (293 cells)
plasmid transfected:
Mock
GATA-4
Nkx2-5
GATA-4/Nkx2-5
IP: 12CA5 mAb
blot: α-GATA-4
Input
blot: α-GATA-4
Input
blot: 12CA5 mAb

FIGURE 3.7.

The physical interaction maps near the C-terminal zinc-finger of GATA-4 and to the C-terminally extended homeodomain of Nkx2-5. **(A)** Luciferase (luc) or deletion mutants of GATA-4 were translated and labelled with ^{35}S-methionine to be subsequently used in pull-down assays with full-length MBP-Nkx2-5 as described in Fig. 3.6. **(B)** A series of immobilized deletion mutants of Nkx2-5, in fusion with the maltose binding protein (MBP), were produced in bacteria, quantified on gel, and used in pull down assays with *in vitro* translated GATA-4. *ΔHD* represents Nkx2-5 Δ122-203. Protein complexes were separated by electrophoresis and GATA-4 protein was visualized by autoradiography on phosphor plates.

A


labelled protein:
luc
GATA-4
201-443
1-332
1-266
Δ 303-390
242-443
1-314
1-327
GATA-6
GATA-4
GATA-1
Pull down assay
labelled protein:
luc
GATA-4
201-443
1-332
1-266
Δ 303-390
243-443
1-314
1-327
GATA-6
GATA-4
GATA-1
Input gel

B

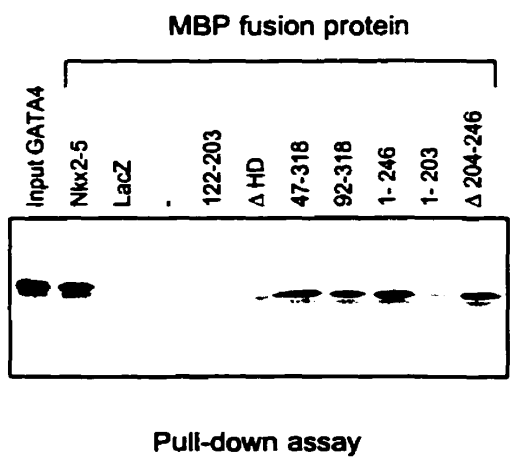
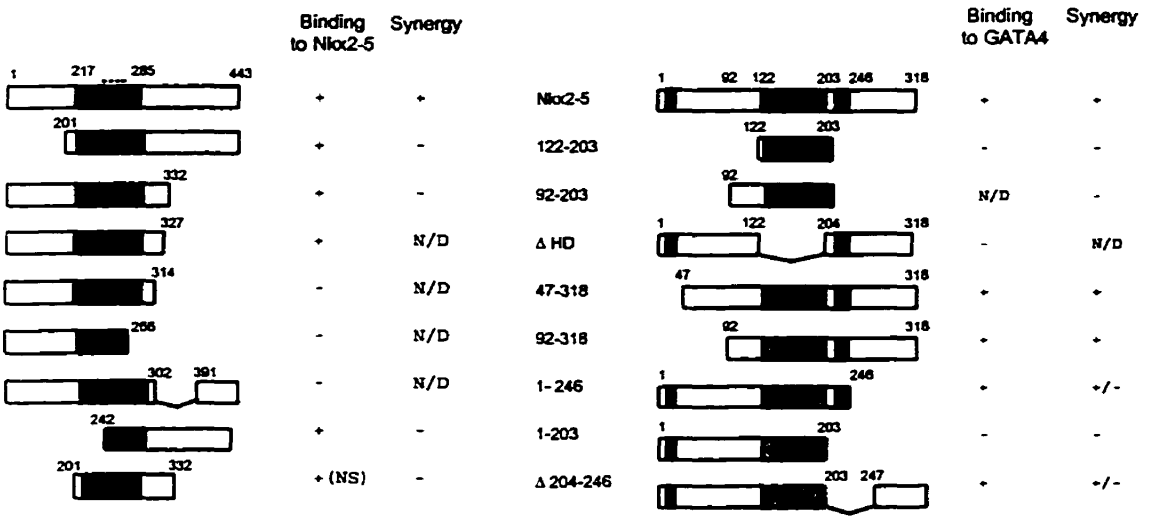

MBP fusion protein
Input GATA4
Nkx2-5
LacZ
-
122-203
Δ HD
47-318
92-318
1- 246
1- 203
Δ 204-246
Pull-down assay

Figure 3.8.

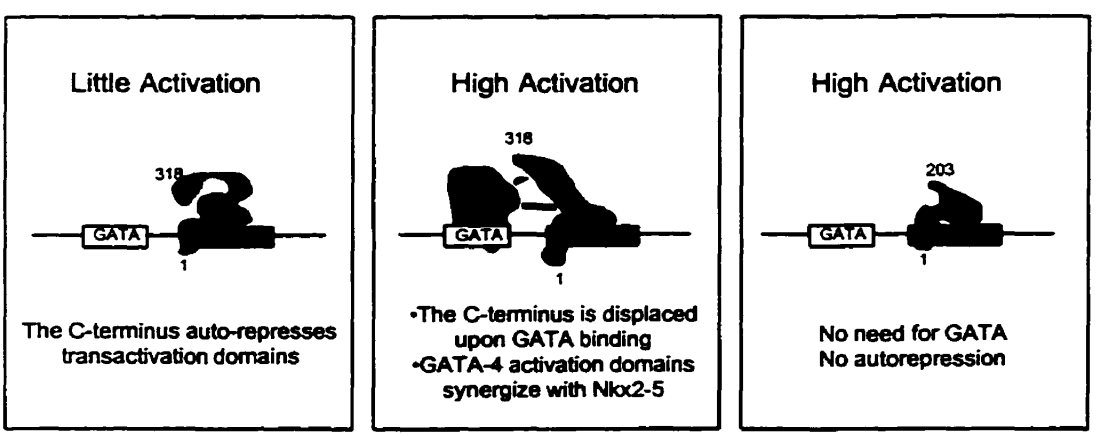
(A) Schematic summary of the Nkx2-5/GATA-4 interaction. The left panel represent the activity of GATA-4 deletions both in transfection experiments and in pull-down assays. N/D represents constructs that were not tested. Constructs deleting the 266-332 region cannot be used in cotransfections since they do not translocate into the nucleus. The right panel summarizes the activities of Nkx2-5 deletion mutants in cotransfections with GATA-4 or in pull-down assays. The asterisks (*) on the GATA-4 molecule represents the amino acids that are conserved between GATA-4 and -5 but not in GATA-6, they are: H244S, N250D, L261V, S262P where the second amino acid represent the residue present on GATA-6 at the equivalent position. (B) Model of Nkx/GATA synergy. Synergistic interactions between Nkx2-5 and GATA-4 require the binding of GATA-4 and Nkx2-5 to their cognate binding sites. GATA-4 displaces the C-terminal auto-inhibitory region of Nkx2-5 and liberates the Nkx2-5 activation domain. The GATA-4 activation domains can then synergize with Nkx2-5 activation domains.

A



B

Model for GATA/Nkx synergy



3.5. MATERIALS AND METHODS

Cell cultures and transfections. HeLa and 293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum. Transfections were carried out using calcium phosphate precipitation 24 h after plating. 36 h post-transfection, cells were harvested and luciferase activity was assayed with an LKB luminometer and the data was recorded automatically. In all experiments, RSV-hGH was used as internal control and the amount of reporter was kept at 3 μg per dish; the total amount of DNA was kept constant (usually 8 μg). Unless otherwise stated, the results reported were obtained from at least four independent experiments with at least two different DNA preparations for each plasmid. Primary cardiocyte cultures were prepared from 1 or 4 day old Sprague-Dawley rats and kept in serum-free medium as described previously (Argentin *et al.*, 1994).

Plasmids. ANF-luciferase promoter constructs were cloned in the PXP-2 vector as described previously (Argentin *et al.*, 1994; Durocher *et al.*, 1996). The construction of the various pCG-GATA-4 vectors was based on the original rat GATA-4 cDNA described in Grépin *et al.*, (1994). The position of the mutation/deletion is indicated on the figures. All constructs were sequenced and functionally tested for nuclear translocation and DNA-binding activity following transfection in L cells as previously described (Grépin *et al.*, 1994). pRSET-GATA-4 derivatives for *in vitro* translation were constructed by insertion of the XbaI-BamHI fragment of the corresponding pCG-GATA-4 construct into the NheI-BamHI sites or NheI-BglII sites of pRSETA (Invitrogen corp.). MBP-Nkx2-5 (Δ203-246) was obtained by the insertion of an oligonucleotide corresponding to aa 198-203 in the PflMI-SacII sites of Nkx2-5. The SphI-SacII fragment corresponding to the deletion was then transferred in MBP-Nkx2-5. The other MBP-Nkx2-5 deletions were described in Chen et al. (Chen and Schwartz, 1995).

Recombinant protein production. After transformation of BL21(DE3) *E.* coli strain with the MBP fusion vectors derived from pMalc-2 (New England Biolabs), individual colonies were picked and grown in 50 ml 2XYT up to an O.D. of 0,6. Induction of the recombinant proteins and their purification were done as previously described

(Durocher *et al.*, 1996). *In vitro* translation of GATA-4 and Nkx2-5 derivatives were performed with rabbit reticulocyte lysates using the TNT coupled in vitro transcription/translation system (Promega Corp., Madison, WI)

Protein-protein binding assays. *In vitro* binding studies were performed with MBP-Nkx2-5 derivatives purified from bacteria and coupled to amylose-sepharose beads (New England Biolabs). GATA-4 derivatives were labeled with ^{35}S methionine during *in vitro* translation and typically 2-8 µL of labeled GATA proteins were incubated in the presence of 300 ng of immobilized Nkx2-5 fusion proteins in 400 µL of 1X binding buffer (150 mM NaCl, 50 mM Tris-Cl pH 7.5, 0.3% Nonidet P-40, 10 mM $ZnCl_2$, 1 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride, 0,25% BSA) for 2 hours at 4°C with agitation and then centrifuged for 2 minutes at 13 000 rpm at room temperature. Beads were washed 3 times by vortexing in 500 mL of binding buffer at room temperature and protein complexes were released after boiling in Laemmli buffer and the proteins were resolved by SDS-PAGE. Labeled proteins were visualized and quantified by autoradiography on phosphor storage plates (PhosphorImager, Molecular Dynamics).

Immunoprecipitations and immunoblots. Immunoprecipitations on nuclear extracts of transfected 293 cells were done using 60 µg of nuclear extract. Extracts were pre-cleared with 20 µL of normal rabbit serum and 15 µL of agarose beads-coupled protein G (Sigma Chemicals) for 2 hours at 4°C. Binding reactions were carried out with 40 µL of 12CA5 antibody in 500 µL of 1X binding buffer without BSA as described in the protein-protein binding assays paragraph for 2 hours at 4°C, with agitation without protein G beads and for an additional 2 hours with 15 µL of protein G beads. Bound immunocomplexes were washed four times in 1X binding buffer and were resuspended in 20 mL of 1X Laemmli buffer, boiled and subjected to SDS-PAGE electrophoresis. Proteins were transfered on Hybond-PVDF membrane and were subjected to immunoblotting. GATA-4 antibody (Santa-Cruz biotechnolgy) was used at a dilution of 1/1000 and was revealed with biotinylated anti-goat antibody (dilution 1/12000) and avidin-biotinylated horseradish peroxidase (HRP) complex (Vectastain). The 12CA5 (anti-Ha) antibody was used at a dilution of 1/500 and was a generous gift of Benoit Grondin and Muriel Aubry (Grondin *et al.*, 1996). The secondary antibody was anti-mouse-HRP and the antigens were visualized

with chemiluminescence (Kodak).

Table 3.1. Putative NK2/GATA interaction

Gene	species	Promoter sequence[a]	GATA	Nkx	Ref
ANF	vertebrates	**TGATAACTT**(N_{20})**CGCCGCAAGTG**	GATA-4	Nkx2-5	this study
Myo-2	*C. elegans*	**TAAAGTGGTTGTGTGGATAA**	elt-2 (?)	Ceh-22	(Okkema and Fire, 1994)
D-Mef	*Drosophila*	**GGATAAGGGGCTCAAGTGG CACTTGAGACCGGGGCTCGC-TATCG**	pannier (?)	Tinman	(Gajewski *et al.*, 1997)

[a]Residues in bold represent either GATA or NKE motifs

3.6. ACKNOWLEDGMENTS

We are grateful to Michel Chamberland for oligonucleotide synthesis, to Lise Laroche for secretarial assistance and to members of the Nemer laboratory for discussions and critical reading of this manuscript. This work was supported by grants from the Canadian Medical Research Council (MRC) and the National Institutes of Health (RO150422). DD is a recipient of the Canada 1967 fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC), FC holds a studentship from the Heart and Stroke Foundation of Canada (HSFC) and MN is a MRC Scientist.

3.7. REFERENCES

Arceci,R.J., King,A.A., Simon,M.C., Orkin,S.H. and Wilson,D.B. (1993) Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol. Cell. Biol.* **13**, 2235-2246.

Ardati,A. and Nemer,M. (1993) A nuclear pathway for α_1-adrenergic receptor signaling in cardiac cells. *EMBO J.* **12**, 5131-5139.

Argentin,S., Ardati,A., Tremblay,S., Lihrmann,I., Robitaille,L., Drouin,J. and Nemer,M. (1994) Developmental stage-specific regulation of atrial natriuretic factor gene transcription in cardiac cells. *Mol. Cell. Biol.* **14**, 777-790.

Azpiazu,N. and Frasch,M. (1993) tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. *Genes Dev.* **7**, 1325-1340.

Blobel,G.A., Simon,M.C. and Orkin,S.H. (1995) Rescue of GATA-1-deficient embryonic stem cells by heterologous GATA-binding proteins. *Mol. Cell. Biol.* **15**, 626-633.

Bodmer,R., Jan,L.Y. and Jan,Y.N. (1990) A new homeobox-containing gene, msh-2, is transiently expressed early during mesoderm formation of Drosophila. *Development*, **110**, 661-669.

Bodmer,R. (1993) The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. *Development*, **118**, 719-729.

Briegel,K., Limi,K.C., Plank,C., Beug,H., Engel,J.D. and Zenke,M. (1993) Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. *Genes Dev.* **7**, 1097-1109.

Buchberger,A., Pabst,O., Brand,T., Seidl,K. and Arnold,H.H. (1996) Chick Nkx-2.3 represents a novel family member of vertebrate homologues to the drosophila homeobox gene tinman - differential expression of CNkx-2.3 and CNkx-2.5 during heart and gut development. *Mech. Dev.* **56**, 151-163.

Chan,S.K., Popperl,H., Krumlauf,R. and Mann,R.S. (1996) An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J.* **15**, 2476-2487.

Chen,C.Y. and Schwartz,R.J. (1995) Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. *J. Biol. Chem.* **270**, 15628-15633.

Chen,J.N. and Fishman,M.C. (1996) Zebrafish *tinman* homolog demarcates the heart field and initiates myocardial differentiation. *Development*, **122**, 3809-3816.

Cleaver,O.B., Patterson,K.D. and Krieg,P.A. (1996) Overexpression of the tinman-related genes XNKX-2.5 and XNKX-2.3 in xenopus embryos results in myocardial hyperplasia. *Development*, **122**, 3549-3556.

Coffman,J.A., Rai,R., Cunningham,T., Svetlov,V. and Cooper,T.G. (1996) Gat1p, a GATA family protein whose production is sensitive to nitrogen catabolite repression, participates in transcriptional activation of nitrogen-catabolic genes in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **16**, 847-858.

Durocher,D., Chen,C.Y., Ardati,A., Schwartz,R.J. and Nemer,M. (1996) The ANF promoter is a downstream target for Nkx-2.5 in the myocardium. *Mol. Biol. Cell*, **16**, 4648-4655.

Egan,C.R., Chung,M.A., Allen,F.L., Heschl,M.F., Van Buskirk,C.L. and McGhee,J.D. (1995) A gut-to-pharynx/tail switch in embryonic expression of the Caenorhabditis elegans ges-1 gene centers on two GATA sequences. *Dev. Biol.* **170**, 397-419.

Evans,S.M., Yan,W., Murillo,M.P., Ponce,J. and Papalopulu,N. (1995) Tinman, a drosophila homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates - xnkx-2.3, a second vertebrate homologue of tinman. *Development*, **121**, 3889-3899.

Fu,Y.H. and Marzluf,G.A. (1990) nit-2, the major nitrogen regulatory gene of Neurospora crassa, encodes a protein with a putative zinc finger DNA-binding domain. *Mol. Cell. Biol.* **10**, 1056-1065.

Fujiwara,Y., Browne,C.P., Cunniff,K., Goff,S.C. and Orkin,S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA*, **93**, 12355-12358.

Gajewski,K., Kim,Y., Lee,Y.M., Olson,E.N. and Schulz,R.A. (1997) D-mef2 is a target for Tinman activation during Drosophila heart development. *EMBO J.* **16**, 515-522.

Gong,Q. and Dean,A. (1993) Enhancer-dependent transcription of the epsilon-globin promoter requires promoter-bound GATA-1 and enhancer-bound AP-1/NF-E2. *Mol. Cell. Biol.* **13**, 911-917.

Gregory,R.C., Taxman,D.J., Seshasayee,D., Kensinger,M.H., Bieker,J.J. and Wojchowski,D.M. (1996) Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. *Blood*, **87**, 1793-1801.

Grépin,C., Dagnino,L., Robitaille,L., Haberstroh,L., Antakly,T. and Nemer,M. (1994) A hormone-encoding gene identifies a pathway for cardiac but not skeletal muscle gene transcription. *Mol. Cell. Biol.* **14**, 3115-3129.

Grépin,C., Robitaille,L., Antakly,T. and Nemer,M. (1995) Inhibition of transcription factor GATA-4 expression blocks *in vitro* cardiac muscle differentiation. *Mol. Cell. Biol.* **15**, 4095-4102.

Grépin,C., Nemer,G. and Nemer,M. (1997) Enhanced cardiogenesis in embryonic stem cells overexpressing the GATA-4 transcription factor. *Development*, **124**, 2387-2395.

Grondin,B., Bazinet,M. and Aubry,M. (1996) The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. *J. Biol. Chem.* **271**, 15458-15467.

Guazzi,S., Price,M., De Felice,M., Damante,G., Mattei,M.G. and Di Lauro,R. (1990) Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J.* **9**, 3631-3639.

Guichet,A., Copeland,J.W., Erdelyi,M., Hlousek,D., Zavorszky,P., Ho,J., Brown,S., Percival-Smith,A., Krause,H.M. and Ephrussi,A. (1997) The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. *Nature*, **385**, 548-552.

Harvey,R.P. (1996) NK-2 homeobox genes and heart development. *Dev. Biol.* **178**, 203-216.

Hawkins,M.G. and McGhee,J.D. (1995) elt-2, a second GATA factor from the nematode Caenorhabditis elegans. *J. Biol. Chem.* **270**, 14666-14671.

Heikinheimo,M., Scandrett,J.M. and Wilson,D.B. (1994) Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. *Dev.*

Biol. **164**, 361-373.

Ip,H.S., Wilson,D.B., Heikinheimo,M., Tang,Z., Ting,C.N., Simon,M.C., Leiden,J.M. and Parmacek,M.S. (1994) The GATA-4 transcription factor transactivates the cardiac muscle-specific troponin C promoter-enhancer in nonmuscle cells. *Mol. Cell. Biol.* **14**, 7517-7526.

Jiang,Y.M. and Evans,T. (1996) The Xenopus GATA-4/5/6 genes are associated with cardiac specification and can regulate cardiac-specific transcription during embryogenesis. *Dev. Biol.* **174**, 258-270.

Kawana,M., Lee,M.E., Quertermous,E.E. and Quertermous,T. (1995) Cooperative interaction of gata-2 and ap1 regulates transcription of the endothelin-1 gene. *Mol. Cell. Biol.* **15**, 4225-31: 3.

Kelley,C., Blumberg,H., Zon,L.I. and Evans,T. (1993) GATA-4 is a novel transcription factor expressed in endocardium of the developing heart. *Development*, **118**, 817-827.

Kuo,C.T., Morrisey,E.E., Anandappa,R., Sigrist,K., Lu,M.M., Parmacek,M.S., Soudais,C. and Leiden,J.M. (1997) GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048-1060.

Lamonerie,T., Tremblay,J.J., Lanctôt,C., Therrien,M., Gauthier,Y. and Drouin,J. (1996) PTX1, a *bicoid*-related homeobox transcription factor involved in transcription of pro-opiomelanocortin (POMC) gene. *Genes Dev.* **10**, 1284-1295.

Laverriere,A.C., MacNeill,C., Mueller,C., Poelmann,R.E., Burch,J.B. and Evans,T. (1994) GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J. Biol. Chem.* **269**, 23177-23184.

Lee,K.H., Xu,Q. and Breitbart,R.E. (1996) A new *tinman*-related gene, *nkx2.7*, anticipates the expression of *nkx2.5* and *nkx2.3* in Zebrafish heart and pharyngeal endocerm. *Dev. Biol.* **180**, 722-731.

Lints,T.J., Parsons,L.M., Hartley,L., Lyons,I. and Harvey,R.P. (1993) Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development*, **119**, 419-431.

Lu,Q. and Kamps,M.P. (1996) Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a

model of a Pbx1-Hox-DNA complex. *Mol. Cell. Biol.* **16**, 1632-1640.

Lyons,I., Parsons,L.M., Hartley,L., Li,R., Andrews,J.E., Robb,L. and Harvey,R.P. (1995) Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev.* **9**, 1654-1666.

Merika,M. and Orkin,S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor gata-1 with the kruppel family proteins sp1 and eklf. *Mol. Cell. Biol.* **15**, 2437-47: 5.

Molkentin,J.D., Kalvakolanu,D.V. and Markham,B.E. (1994) Transcription factor GATA-4 regulates cardiac muscle-specific expression of the a-myosin heavy-chain gene. *Mol. Cell. Biol.* **14**, 4947-4957.

Molkentin,J.D., Lin,Q., Duncan,S.A. and Olson,E.N. (1997) Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061-1072.

Morrisey,E.E., Ip,H.S., Lu,M.M. and Parmacek,M.S. (1996) GATA-6 - a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* **177**, 309-322.

Morrisey,E.E., Ip,H.S., Tang,Z.H., Lu,M.M. and Parmacek,M.S. (1997) GATA-5 - a transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev. Biol.* **183**, 21-36.

Newman,C.S., Grow,M.W., Cleaver,O., Chia,F. and Krieg,P. (1997) Xbap, a vertebrate gene related to bagpipe, is expressed in developing craniofacial structures and in anterior gut muscle. *Dev. Biol.* **181**, 223-233.

Okkema,P.G. and Fire,A. (1994) The Caenorhabditis elegans NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. *Development*, **120**, 2175-2186.

Osada,H., Grutz,G., Axelson,H., Forster,A. and Rabbitts,T.H. (1995) Association of erythroid transcription factors: complexes involving the lim protein rbtn2 and the zinc-finger protein gata1. *Proc. Natl. Acad. Sci. USA*, **92**, 9585-9: 2.

Pandolfi,P.P., Roth,M.E., Karis,A., Leonard,M.W., Dzierzak,E., Grosveld,F.G., Engel,J.D. and Lindenbaum,M.H. (1995) Targeted disruption of the GATA3 gene causes severe

abnormalities in the nervous system and in fetal liver haematopoiesis [see comments]. *Nature Genetics*, **11**, 40-44.

Peltenburg,L.T.C. and Murre,C. (1997) Specific residues in the PBX homeodomain differentially modulate the DNA-binding activity of Hox and engrailed proteins. *Development*, **124**, 1089-1098.

Phelan,M.L., Rambaldi,I. and Featherstone,M.S. (1995) Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell. Biol.* **15**, 3989-3997.

Platt,A., Langdon,T., Arst,H.N.J., Kirk,D., Tollervey,D., Sanchez,J.M. and Caddick,M.X. (1996) Nitrogen metabolite signalling involves the C-terminus and the GATA domain of the Aspergillus transcription factor AREA and the 3' untranslated region of its mRNA. *EMBO J.* **15**, 2791-2801.

Schultheiss,T.M., Xydas,S. and Lassar,A.B. (1995) Induction of avian cardiac myogenesis by anterior endoderm. *Development*, **121**, 4203-4214.

Spieth,J., Shim,Y.H., Lea,K., Conrad,R. and Blumenthal,T. (1991) elt-1, an embryonically expressed Caenorhabditis elegans gene homologous to the GATA transcription factor family. *Mol. Cell. Biol.* **11**, 4651-4659.

Stanbrough,M., Rowen,D.W. and Magasanik,B. (1995) Role of the GATA factors Gln3p and Nil1p of Saccharomyces cerevisiae in the expression of nitrogen-regulated genes. *Proc. Natl. Acad. Sci. USA*, **92**, 9450-9454.

Sturm,R.A., Das,G. and Herr,W. (1988) The ubiquitous octamer-binding protein Oct-1 contains a POU domain wit a homeo box subdomain. *Genes Dev.* **2**, 1582-1599.

Ting,C.N., Olson,M.C., Barton,K.P. and Leiden,J.M. (1996) Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature*, **384**, 474-478.

Tonissen,K.F., Drysdale,T.A., Lints,T.J., Harvey,R.P. and Krieg,P.A. (1994) XNkx-2.5, a Xenopus gene related to Nkx-2.5 and tinman: evidence for a conserved role in cardiac development. *Dev. Biol.* **162**, 325-328.

Tsai,F.Y., Keller,G., Kuo,F.C., Weiss,M., Chen,J., Rosenblatt,M., Alt,F.W. and Orkin,S.H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, **371**, 221-226.

Visvader,J.E., Crossley,M., Hill,J., Orkin,S.H. and Adams,J.M. (1995) The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Mol. Cell. Biol.* **15**, 634-641.

Walters,M. and Martin,D.I. (1992) Functional erythroid promoters created by interaction of the transcription factor GATA-1 with CACCC and AP-1/NFE-2 elements. *Proc. Natl. Acad. Sci. USA*, **89**, 10444-10448.

Weiss,M.J., Keller,G. and Orkin,S.H. (1994) Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **8**, 1184-1197.

Weiss,M.J., Yu,C.N. and Orkin,S.H. (1997) Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol.* **17**, 1642-1651.

Whyatt,D.J., deBoer,E. and Grosveld,F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *EMBO J.* **12**, 4993-5005.

Winick,J., Abel,T., Leonard,M.W., Michelson,A.M., Chardon-Loriaux,I., Holmgren,R.A., Maniatis,T. and Engel,J.D. (1993) A GATA family transcription factor is expressed along the embryonic dorsoventral axis in Drosophila melanogaster. *Development*, **119**, 1055-1065.

Yu,Y., Li,W., Su,K., Yussa,M., Han,W., Perrimon,N. and Pick,L. (1997) The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. *Nature*, **385**, 552-555.

CHAPTER 4

A vertebrate member of the UPF1 family is a sequence-specific regulator of transcription.

Daniel Durocher, Jean-Marc Juteau and Mona Nemer

Manuscript in preparation

The surprising finding that the NKE behaves as an atrial-specific element suggested that many transcriptional pathways collaborate in order to generate chamber-specific promoter activity. Thus, to further our knowledge on the molecular pathways leading to chamber-specificity, I undertook to study the CARE element which was shown to be necessary for embryonic and postnatal atrial ANF promoter activity. We undertook to clone CARE-binding proteins by a southwestern-based screen. This work describes the characterization of CATF1 which is a ssDNA-binding helicase that contains transcriptional activities. However, CATF1 does not seem to be the endogenous CAX protein and therefore, CATF1 represents a novel transcriptional pathway converging on the ANF promoter.

This is a multi-authored work. Jean-Marc Juteau initiated the molecular cloning of CATF and obtained the first partial clone. I cloned the remaining of the cDNA and performed all the characterization under the supervision of Mona Nemer.

ABSTRACT

Helicases play important roles in virtually all cellular processes involving nucleic acids such as DNA repair, replication, recombination, RNA transcription, splicing, and translation where they perform the vectorial unwinding of duplex nucleic acids. Recently, a number of putative helicases have been shown or proposed to be involved in the regulation of transcription such as the members of the Swi2/rad54 family. However, except for p70 Ku and the SV40 large T antigen very few of them so far harbor both *bona fide* helicase activity and features of classical transcriptional regulators such as sequence-specific DNA-binding and autonomous transcriptional activation. Here we report the isolation of such a protein which we called CATF1 and which is the ortholog of the previously described Sμbp-2 and RIP1a proteins. CATF1 was cloned by its ability to bind a cardiac-specific element of the ANF promoter, and harbors a functional helicase domain of the helicase superfamily I. CATF1 defines a new helicase family that comprises UPF1-, SEN1-, and DNA2-related helicases. Close CATF1 relatives are found in *S. cerevisiae* and in the archeal genomes of *Methanococcus jannaschii* and *Archaeoglobus fulgidus*, indicating that CATF1 defines an evolutionary conserved novel helicase subfamily. CATF1 was found to possess sequence-specific DNA-binding activity which is negatively modulated by ATP-binding. Additionally, only a helicase- truncated CATF1 could transactivate promoters containing its binding site, suggesting that CATF1 is a sequence-specific transactivator and that ATP binding could modulate CATF1 activity in vivo. Finally, in mammals, CATF1 transcripts are enriched in the embryonic and perinatal myocardium raising the possibility that CATF1 may play stage and tissue-specific roles

INTRODUCTION

The chemical nature of nucleic acids enables the formation of duplexes of either RNA-RNA, DNA-RNA or DNA-DNA molecules through extensive hydrogen bonding between complementary base pairs. However, almost all cellular processes that read, process or repair the genetic code require nucleic acid unwinding enzymes, the helicases. Helicases function at the single-stranded/double-stranded nucleic acid junction to catalyze the transient, vectorial and progressive unwinding of the nucleic acid molecule. This catalysis requires energy in the form of 5'-triphosphate hydrolysis. Thus, the helicase is also an NTPase that uses this energy to translocate along the DNA and to break the hydrogen bonds between bases [reviewed in (34,35)].

Helicases have been classified in a number of large protein families. Most helicases are part of three superfamilies (SF) termed SF1 to SF3 (23). Sequence analyses of *bona fide* helicases revealed that they harbor an NTP-binding domain, the so-called A and B motifs of the Walker box (24,27). This domain is shared by all the helicases characterized so far and reflects the energy requirement as ATP, to perform the vectorial unwinding of nucleic acids (35). Further computer-assisted analyses revealed that SF1 and SF2, but not SF3 members share five other conserved motifs. These motifs, together with the NTP-binding site compose the helicase signature that is found in the majority of helicases characterized so-far. These superfamilies have developed internal hierarchies with defined families and subfamilies within them (23). However, sequence conservation among family members does not guarantee functional conservation but rather is likely to represent conservation of biochemical properties. For example, some members of the swi2/rad54 family which do harbor the seven helicase subdomains have not been shown yet to be *bona fide* helicases, rather, they have been recently proposed to be ATP-driven "disrupters" of protein-DNA interactions (44). Thus, the presence of the seven-domains-helicase motif does not guarantee helicase activity and the helicase domain may have been used during evolution to generate a panel of related ATP-dependent biochemical activities involving nucleic acids. It has been estimated that in fact, helicase domain-encoding genes

would account for up to 1% of all protein coding genes (23). This reflects the involvement of helicases in all aspects of nucleic acid metabolism, genome maintenance and gene expression.

In recent years, genetic and biochemical studies have unraveled the direct involvement of helicases and nucleic acid-stimulated NTPases in the regulation of transcription in prokaryotes, viruses and eukaryotes. For instance, the SV40 large T antigen is a helicase from the SF3 class which is required for both transcriptional activation of the SV40 enhancer and viral replication (18). Interestingly, whereas the initiation of replication requires a functional helicase, transcriptional activation of the SV40 enhancer does not, implying that the large T antigen possesses ATP-dependent and independent activities (57). On the other hand, putative helicases from the rapidly growing SWI2 family (SF2 class) have been shown to be involved in transcriptional activation by both elegant biochemical and genetic studies; moreover, the transcriptional activity of swi2/snf2p has been shown to be ATP-dependent and seems to require the integrity of the seven helicase subdomains (32). However, both the SV40 T antigen and swi2p have to be recruited to promoters via DNA-bound proteins in order to perform their transcriptional functions. In contrast, two putative helicases from SF2, and another a one from SF3, harbor specific DNA-binding activities enabling them to interact directly with promoter or other chromosomal sites. The mammalian CHD1 protein has been cloned through its ability to bind an A/T rich sequence of the Vκ19A promoter (12). CHD1 possesses a helicase domain of the swi2/rad54 family and binds with a relaxed specificity to A/T rich sequences in the minor-groove of DNA (52). CHD1 also harbors a chromodomain outside the helicase domain suggesting that it might be involved in regulating chromatin structure in a sequence-specific manner, perhaps at scaffold/matrix attachment regions (S/MAR) which are enriched in A/T (6). Another SF2 helicase, the HIP116/Zbu protein was cloned by virtue of its interaction with either the Sph elements of the HIV LTR or a critical E-box of the MLC3 promoter (22,49). Since the two target sites are unrelated with respect to their primary sequence, HIP116/Zbu may recognize a structural DNA motif. Interestingly, HIP116/Zbu expression is restricted during development to certain tissues such as the heart suggesting that it might play a role in some developmental processes. Although both CHD1 and HIP116/Zbu bind DNA, no

transcriptional activity could be documented, implying that they either require ancillary proteins to accomplish their transcriptional function or that they act at other levels of gene regulation such as chromatin remodeling, replication or even splicing. Finally, an SF3 member, the Ku autoantigen p70 [also known as DNA Helicase II (54)] has been involved in the regulation of RNA polymerase I and II transcription (20,21,30,40); this protein, which possesses sequence-specific DNA-binding activity, has been characterized as the regulatory subunit of the DNA-dependent protein kinase (DNA-PK). Since it recruits the catalytic subunit the DNA-PK to DNA, this event affects transcription because the recruitment of DNA-PK results in the subsequent phosphorylation of DNA-bound, nearby transcriptional regulators (21). So-far, no member of the SF1 class has been linked to transcriptional regulation.

Here we present the molecular cloning and the biochemical characterization of the SF1 helicase protein CATF1/Sμbp2. CATF1 was cloned by virtue of its interaction with a tissue-restricted element present in the promoter of the cardiac-specific atrial natriuretic peptide (ANF) gene. Sequence analyses revealed that the CATF1 protein is evolutionarily conserved and putative homologues are present in yeast and archea. The data presented suggest that CATF1 is a prototype for a new family of helicases, tentatively named the UPF1 family, that share both sequence homologies and biochemical properties with the yeast Upf1 protein including *bona fide* helicase activity and ATP-dependent DNA-binding. In addition, CATF1 harbors unique activities such as sequence-specific DNA-binding and transcriptional activation properties.

MATERIALS AND METHODS.

Library screening and DNA sequence determination. A directional cardiocyte λgt22 cDNA library (superscript kit from Bethesda Research Laboratories) was prepared as described previously (25). The library was screened by an in situ filter binding technique using the polymerized rCARE oligonucleotide (5'-gatccGCAGAGGGAGCTGGGTGT-GGGCCAGCCGTCA-3') as described in (36) in the following binding buffer: 5 mM $MgCl_2$, 60 mM KCl, 10 mM Tris-Cl pH 7.4 and 1 mM DTT. The *SalI-NotI* insert of the

purified clone was subcloned in pBluescript KS (Stratagene inc. La Jolla CA). 5' sequences were cloned by RACE-PCR using cardiomyocyte cDNA and subcloned in the *Sma*I site of PTz19R. Three independent RACE clones were sequenced on double-stranded templates using the dideoxynucleotide method of Sanger. Primer walking generating contiguous sequences on both strands was used to fully determine the sequence.

Plasmids. The ANF luciferase plasmids used in this study were described previously (2). The primary CATF1 cDNA form the positive phage λ1 corresponding to aa 470 to 989 was inserted in the SalI-NotI sites of pBluescript KS. Full-length CATF1 cDNA was generated by subcloning the *Eco*RI-*Acc*I fragment of the RACE-PCR product in pKS-CATF1. This plasmid was digested with KpnI and XbaI to subclone the CATF1 cDNA without the 3' UTR into pcDNA3 (Invitrogen). CATF1 (470-989) was then inserted in the prokaryotic expression vector pmalc using the *Xba*I-*Bam*HI of pKS-CATF1. The same fragment was used to subclone in the XbaI-BamHI sites of pRSV-GAL4 to prepare GAL4-CATF1 fusion vector. pRSET-CATF1 was generated by introducing in frame the complete cDNA of CATF1 form pCDNA3 KpnI-BamHI in pRSETB. Site-directed mutagenesis of Lys281 to Arg on the pRSET-CATF1 vector was carried out by the Quickchange mutagenesis method (Stratagene) using Pfu polymerase and the following oligonucleotides: 5'-TCCTGGCACTGGGAGAACCACAACTG-TGGT-3' and 5'-ACCACAGTTGTGGTTCTCCCAGTGCCAGGA-3' following the manufacturer's recommendations.

Cell culture and transfections. Primary cardiocyte cultures were prepared from 1- to 4-day-old Sprague-Dawley rats and maintained in serum free medium as described previously (2). NIH 3T3, L, and simian kidney 293 cells were grown in Dulbecco's modified Eagle's medium supplemented with 5% fetal (NIH 3T3) and 10% fetal calf serum (L cells) respectively. DNA was transfected by the calcium phosphate precipitation technique. The amount of reporter plasmid was 3 μg per dish and the total amount of DNA (maximum 12 μg) was kept constant in all dishes by addition of plasmid pCG, pCGN or pcDNA3. Cells were harvested 36 hours after the transfection and luciferase activity was measured with an LKB luminometer and automatically recorded.

RNA extraction and reverse-transcription PCR. Total cellular RNA was isolated by the guanidium-thiocyanate-phenol-chloroform method (9). Northern (RNA) blots were performed using the glyoxal method as described previously (25). After transfer, the Nytran membranes were prehybridized and then hybridized with random primed ^{32}P-labeled CATF1 or β-actin cDNA probes. Blots were washed for 15 min at 65 °C in 0.1X SSC-0.1% sodium dodecyl sulfate (SDS) (1X SSC is 0.15 M NaCl plus 0.015 M sodium citrate). For PCR amplification of CATF1 transcripts, oligonucleotides 5'-CTCGTCACTTTGCACAT-3' (amino acids 572 to 578) and 5'-GGCTGGACTTGAGAGCC-3' (amino acids 668 to 774) were used on cDNAs made from various rat neonatal or adult tissues and from primary cardicyte cultures as described by Dagnino et al. (10). PCR products were separated by agarose gel electrophoresis, and Southern blots were probed with a labelled internal oligonucleotide (5'-ACATGCCGCCTAGCACG-3') corresponding to the amino acids 604 to 610 of the CATF1 sequence. The same cDNA samples were subjected to tubulin amplification as a control.

Recombinant protein production. XL1-blue (Stratagene) bacteria harboring the MBP-CATF1 fusion vector were grown overnight in 2XYT medium in the presence of 2% glucose and 100 μg/mL of ampicillin. 2% glycerol was then added to a 1/10 dilution of the culture which was grown until the OD_{600} was 0.5-0.6 units. MBP-CATF1 production was induced for 16 hours at 4°C by the addition of 1 mM final of isopropyl β-D-thiogalactoside (IPTG). Cells were then centrifuged and resuspendend in cold column buffer [200 mM Tris-Cl (pH 8.0), 5 mM EDTA, 2 mM phenylmethylsulfonyl fluoride (PMSF), 2 mM dithiothreitol and 1 mM NaN_3]. Cells were lysed by ultrasonication, discontinuously for 1 min. The lysates were spun and the soluble fraction was passed on an agarose-amylose affinity column as described in the manufacturer's protocol (New England Biolabs). His-CATF1 fusions were expressed in BL21(DE3) cells (Novagen). Expression was induced by adding 0.25 mM of IPTG to 1 liter of exponentially growing cells in 2XYT medium in the presence of 100 μg/mL of ampicillin. After 3 hours of induction at 37°C, cells were harvested by centrifugation and resuspended in 13 mL of

MCAC0 buffer [20 mM Tris-Cl (pH 7.9), 500 mM NaCl, 1 mM PMSF] and frozen overnight. The next morning, after thawing, cells were lysed by sonication and centrifuged to remove insoluble material. Three mL of a Ni^{2+} agarose slurry (50% vol/vol) (Invitrogen) was packed on a low pressure column (Bio-RAD) and washed with 10 bed volumes of MCAC0. Lysates were applied on the column at an elution rate of approximately 20 mL/hour and the column was washed with 10 bed volumes of MCAC0 plus 30 mM imidazole. Fusion proteins were eluted in 10 mL of MCAC0 plus 100 mM imidazole. Fractions were collected and proteins were detected by a Bradford assay (Bio-Rad) and the purity was estimated on an SDS-PAGE. Fractions were pooled and dialysed in storage buffer [20 mM Tris-Cl (pH 7.9), 150 mM NaCl, 10% glycerol, 1 mM DTT and 1 mM PMSF]. Protein fractions in storage buffer were snap-frozen in liquid nitrogen and stored at -80°C.

Whole cell extract preparation and in vitro DNA-binding assays. Whole cell extracts were prepared from approximately $5X10^6$ cells by hypotonic lysis of the cells. Cells were first washed with 4 mL of ice-cold PBS. Cells were then collected in 2 mL of ice-cold PBS/1 mM EDTA and spun in a microfuge. The cell pellet was then resuspended in WCEB [25 mM Hepes (pH 7.7), 0.3 mM NaCl, 15 mM $MgCl_2$, 0.2 mM EDTA and 0.1% Triton X-100] and cells were lysed by gentle agitation on a Nutator for 30 min. The cell lysate was then spun and the supernatent was kept for quantitation by the Bradford method. Extracts were aliquoted in small volumes and snap-frozen in dry ice and stored at -80°C for further use. Gel shift assays were performed by incubating affinity purified MBP or MBP-CATF1 (around 50 ng) or 10 μg of whole-cell protein extracts in a 20 μL reaction mixture containing 20 000 cpm of the gel-purified ssDNA or dsDNA rat CARE probe [prepared as previously described (25)] in a binding buffer containing 10 mM Tris-Cl (pH 7.9), 5 mM $MgCl_2$, 60 mM KCl, 1 μg of poly dI-dC. Reactions mixtures were incubated on ice for 30 minutes and the complexes were separated on a 5% polyacrylamide gel in 0.25X Tris-Borate-EDTA buffer at 250 V for 3 h at room temperature. Off-rate and on-rate analyses are described in the legend of Fig. 4.5. Gels

were transfered onto a Whatman paper, dried and radiolabeled complexes were visualized by autoradiography on phosphor storage plates (PhosphorImager, Molecular Dynamics).

Antibody production and antibody techniques. Polyclonal antibody CAA2 directed against rat CATF1 protein was generated by immunizing rabbits with the MBP-$CATF1_{470-989}$ fusion protein following the protocol described in (26). For western detection of CATF1, proteins were subjected to SDS-PAGE electophoresis, transfered on Hybond-PVDF membranes and incubated with the CAA2 antibody (dilution 1/2000). Immunoreactivity was revealed by an HRP-conjugated anti-rabbit goat antibody at a dilution of 1/50 000. Monoclonal 12CA5 antibody (anti-HA) was used at a dilution of 1/500 and was detected with an anti-mouse-HRP secondary antibody. Antigen-antibody complexes were visualized by chemilumiscence following the manufacturer's recommendation (Renaissance, NEN).

ATPase and Helicase assays. The hydrolysis of ATP was measured by the release of inorganic phosphate, $^{32}P_i$ from [γ-^{32}P]ATP by PEI-cellulose thin layer chromatography. Typically reactions were assembled in a final volume of 20 µL with or without nucleic acids as cofactors in a reaction mixture containing 25 mM Tris-Cl (pH 7.0), 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 1 µCi of [γ-^{32}P]ATP (>5000 ci/mmol, Amersham) and 300 ng of His-CATF1 protein. Reaction mixtures were incubated at 37°C for 30 min and stopped with I µL of EDTA at 500 mM. A sample of the reaction mixture (2 µL) was then spotted onto PEI-cellulose plates (J.T. Baker) and chromatography was carried out in 0.8M LiCl, 0.8 M acetic acid. Plates were dried and radioactivity was visualized and quantified by autoradiography on phosphor storage plates (PhosphorImager, Molecular Dynamics). Helicase activity was examined using a strand displacement assay. Helicase templates were generated by annealing 1 pmol of a 47-mer oligonucleotide corresponding to the polycloning site of M13mp19, on 2 µg of M13mp18. Labeling of the substrate was carried out by extension of the primer using the Klenow fragment of DNA polymerase I and dTTP, dGTP and [α-^{32}P]dATP. The 52 nucleotide extended primer was then purified by gel filtration on a G-100 sephadex column (Pharmacia Biotech). Helicase assays were performed in 20 µL reactions containing 20 mM Tris-Cl (pH 8.0), 50 mM KCl, 3 mM

$MgCl_2$, 1 mM DTT, 1 mM ATP (unless otherwise stated) with typically 5 fmol of the DNA helicase substrate and 50 ng of His-CATF1 fusion proteins and incubated at 37°C for 30 min. Reactions were stopped by the addition of 5 μL of stop/loading buffer (50% glycerol, 0.5% SDS, 10 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol). Samples (usually 5 μL) of the reaction mixture were electrophoresed on a 6% PAGE (30:1 in 1 X TBE) at 15 mA constant current. Gels were dried and autoradiographed. Bands were visualized and quantitated on a PhosphorImager (Molecular Dynamics)

RESULTS

Over the past years, intense work has been directed toward the identification of cardiac-specific transcriptional pathways. The analysis of cardiac promoters has so far identified only a limited number of truly cardiac-specific *cis*-elements [reviewed in (17)]. Two of these, the cardiac GATA and NKE elements have been discovered through analyses of the cardiac promoters of the natriuretic peptide hormone genes, ANF and BNP (16,25). More recent studies on the cardiac ANF promoter unraveled the role of a conserved element located within the enhancer and termed CARE for cardiac regulatory element (Ardati *et al.* manuscript in preparation). The CARE element is necessary for the high cardiac-specific activity of the ANF element in embryonic cardiomyocytes and in postnatal atrial myocytes.

Molecular cloning of a novel cardiac-restricted DNA-binding protein. In order to identify cardiac protein(s) interacting with the rCARE element of the ANF promoter, an *in situ* filter binding technique was used to screen approximately 1 X 10^7 plaques of a cardiocyte λgt22 cDNA library using the polymerized rCARE element (51). Only one positive clone, λ1, went through the steps of enrichment and was purified to homogeneity. Subcloning and subsequent sequencing of the λ1 insert revealed that it encodes an ORF lacking the initiator ATG codon. RACE-PCR was carried out to clone the 5' end of the cDNA and three identical clones were sequenced and found to encode the initiator ATG. The assembly and the conceptual translation of the RACE and λgt22 clones identified a

putative full-length ORF that we called CATF1 (for Cardiac ATPase Transcription Factor).

In order to examine the sites of expression for CATF1, Northern blot analyses on several neonatal tissues were carried out. A 4.4 kb hybridizing band corresponding to the CATF1 mRNA was found only in heart (Fig. 4.2A). This cardiac-restricted expression was further confirmed by semi-quantitative RT-PCR analyses on several neonatal and adult rat tissues. In the neonatal rat, CATF1 transcripts were detected consistently in tissues from the atrial and ventricular chambers of the heart and from the pancreas (Fig. 4.2B). Transcripts for CATF1 were not detected, at this stage, in all other tissues examined such as the lung, liver, stomach, intestine, kidney and skeletal muscle. However, in adult tissues CATF1 mRNA was found to be more widely expressed, particularly in various muscle tissues such as cardiac, skeletal and smooth muscles (Fig 4.2C and data not shown). This expression pattern suggests that CATF1 may be involved in developmental processes.

CATF1, is a vertebrate member of the class I helicase superfamily. Computer-assisted analysis of CATF1 coding sequence revealed the presence of a helicase motif that spans for most (70%) of the protein and a P-Q rich region at the C-terminus of the protein reminiscent of activation domains present in many transcription factors. The seven helicase subdomains are present on CATF1 including the consensus G/AxxGxGKS/T for the ATP-binding sequence (motif I) and a hydrophobic stretch followed by the residues DE/D the motif II, responsible for Mg^{++}/NTP binding which places CATF1 in (SF1) of the helicase superfamily I (23).

Databank analysis of the primary structure of CATF1 revealed that this factor is the rat ortholog of the hamster RIP1a (31) the human GF-1/Sμbp-2 (5,19) and the mouse Sμbp-2 (37). The helicase motif did not score well with other types of known helicase proteins including members of the helicase superfamily II (comprising members of the DEAD box family and the SWI2/rad54 family of DNA-dependent ATPases) as well as members of the helicase superfamily III (such as SV40 large T antigen) . The best scores

Figure 4.1.

Primary structure of CATF1. **(A)** CATF1/Sμbp-2 possesses an helicase motif. Predicted translation of rat CATF1 form the assembly of the λcatf fragment and the RACE-PCR fragment. The seven helicase subdomains as well as the PQ region are underlined. **(B)** CATF1/Sμbp-2 defines a new family of helicases. Here is The alignments between members of the UPF1 superfamily are depicted. Alignments were obtained using the BLAST2 program and manually compiled. Bold residues represent consensus residues of the helicase I superfamily whereas underlined residues are representative of the UPF1 family. Accession numbers for the aligned proteins: L10075 (mCATF); L15625 (hamCATF1); L14754 (humCATF1); P34243 (ykl017c); L77117 (MJU67467 or MJ_HEL); M76659 (scUPF1); P38859 (scDna2); Q00416 (scSEN1); and X52574 (mMOV10). **(C)** Schematic description of the helicase families. All UPF1 family members harbor a probability of non-relatedness inferior to $1e^{-30}$ as calculated by the BLAST2 program (filter off). Subfamily members were clustered with each other and have a probability of non-relatedness inferior to $1e^{-60}$. The Sμbp-2, CATF1 and RIP1a were considered orthologs.

A

```
   1  MASSTVESFV AQQLQLLELE RDAEVEERRS WQEHSSLKEL QSRGVCLLKL

  51  QVSGQRTGLY GQRLVTFEPR KFGPAAVLPS NSFTSGDIVG LYDTNESSQL

 101  ATGVLTRITQ KSVIVAFDES HDFQLNLDRE NTYRLLKLAN DTYKRLKKAL

 151  LTLKKYHSGP ASSLIDVLLG GSTPSPATEI PPVTFYNTTL DPSQKEAVSF

 201  ALAHKEVAII HGPPGTGKTT TVVEIILQAV KQGLKVLCSA PSNSSAVDNL
                    I                          IA

 251  VERLALCKKQ ILRLGHPARL LESVQQHSLD AVLARSDNAQ IVADIRRDID

 301  QVFGKNKKNQ DEREKGNFRN EIKLLRKELK EREEAAIVQS LSAADVVLAT

 351  NTGASTDGPL KLLPEDYFDV VVVDEPAQAL EASCWIPLLK APKCILAGDH
                              II

 401  KQLPPTTVSH KAALAGLSRS LMERLAEKHG AAVVRMLVVQ YRMHQAITRW
        III                                    IV

 451  ASEAMYHGQL TAHPSVAGHL LKDLPGVADT EETSVPLLLI DTAGCGLLEL

 501  EEEDSQSKGN PGEVRLVTLH IQALVGCWGP GCDIAVIAPY NLQVDLLRQS

 551  LSNKHPELEI KSVDGFQGRE KEAVILTFVR SNRKGEVGFL AEDRRINVAV
                          V

 601  TRARRHVAVI CDSHTVNNHA FLKTLVDYFT EHGEVRTAFE YLDDIVPENY
        VI

 651  THEGSRSHSC APKPKCPTTS VRKPASAQES RQEARAATGH SRRKPSEKPL

 701  GSQVQPQHSS KANGSDRTGG TDRTEHFRAM IEEFVASKEA QLEFPTSLSS

 751  HDRLRVHQLA EEFGLKHDST GEGKARHITV SRRSPAGSGS ATPQPPSPPS

 801  PAQAEPEPQV EQPVGQPHGP TQLDLKALHL ERLQRQQGSP AQTQLGGGSR
                      PQ-rich

 851  PQKAPQKKKK KEPKGPAMAL PSEEDFDALV SAVVKADNTC SFTKCSAAPP

 901  TLGQFCMHCS RRYCLSHHLP EIHGCGEKAR AHARQRISRE GVLYAGSGTK

 951  DRALDPAKRA QLQRKLDKKL GELSSQRTSK KKEKERGT
```

B

```
                    I                  Ia                 II                 III
  rCATF1   208- IIHGPPGTGKT -16- VLCSAPSNSAVD -122- VVVDECAQAL-14-ILAGDHKQLPPTTV-
  mCATF1   209- ........... -16- ...C....I... -122- ..........-14-.......R........-
hamCATF1   209- ........... -16- I..C....V... -122- ..........-14-.......R........-
humCATF1   210- ........... -16- ...C....I... -122- ..I.......-14-...............-
 Ykl017C   226- ........... -18- I.I.G...IS.. -125- .II..VS..M-19-V....N......IK-
  MJ_HEL   203- ........... -17- ..ATAD..I.A. -154- I.I..GS..M-15-.M.........VL-
  scUPF1   426- L.Q........ -17- ..VC....I... -104- .LI..ST..S-15-..V...Q..G.VIL-
  scDna2  1071- L.L.M...... -16- ..LTSYTH.... -73-- .IL..AS.IS-14-IMV...Y....LVK-
  scSEN1..1234- L.Q........ -41- I.IC....A... -169- .IL..AC.CT-15-.MV..PN.....VL-
  mMOV10   521- ..F........ -18- I.AC.....GA. -80-- IFI..AGHCM-24-V....PR..G.VLR-

                   IV                 V                  VI
  rCATF1    28- LVVQYRM -116- IKSVDGFQGREKEAVIL -16- DRRINVAVTRARRHVAVICD -376
  mCATF1    28- .T..... -116- ..................... -16- ..................... -381
hamCATF1    28- .T..... -116- ..................... -16- ..................... -377
humCATF1    28- .T..... -116- ..................... -16- ..................... -380
 Ykl017c    31- .N..... -125- .ST.........DVI.. -16- E..L...M..P..QLV.VGN -35
  MJ_HEL    22- .EI.... -116- VNT........N..IVI -13- L......I...K.KLILIGN -31
  scUPF1    26- .E..... -118- VA...A......DYI.. -16- P..L..GL...KYGLVILGN -156
  scDna2    27- .TL.... -142- .LTA.Q....D.KCI.I -17- L..V...M...KSKLIIIGS -131
  scSEN1    26- .D..... -119- FNTI.....Q...IILI -17- F..M...L...KTSIWVLGH -401
  mMOV10    38- .LRN..S -127- VG..EE...Q.RSVILI -23- PK.F.......KALLIVVGN -21
```

C

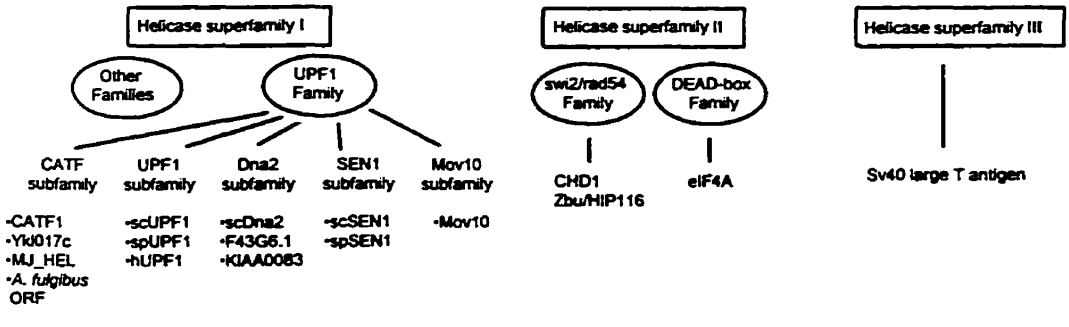


Figure 4.2.

CATF1 is dynamically expressed during postnatal development. **(A)** Northern blot of neonatal tissues. Each lane contains 15 μg of total RNA from 1- to 3-day-old neonatal rat tissues. The same blots were rehybridized sequentially with CATF1 and rat β-actin as described in Material and Methods. **(B)** PCR amplification of CATF1 transcripts in various neonatal tissues. PCR-amplified transcripts were separated by agarose gel electrophoresis, blotted on a nylon membrane, and hybridized with an internal oligonucleotide. The conditions and the sequence of the primers are found in the Material and Methods section. Please note that the PCR was carried under semi-quantitative conditions. **(C)** PCR amplification of CATF1 transcripts in various adult tissues. PCR reactions were carried out using the same primers and conditions as in (B).

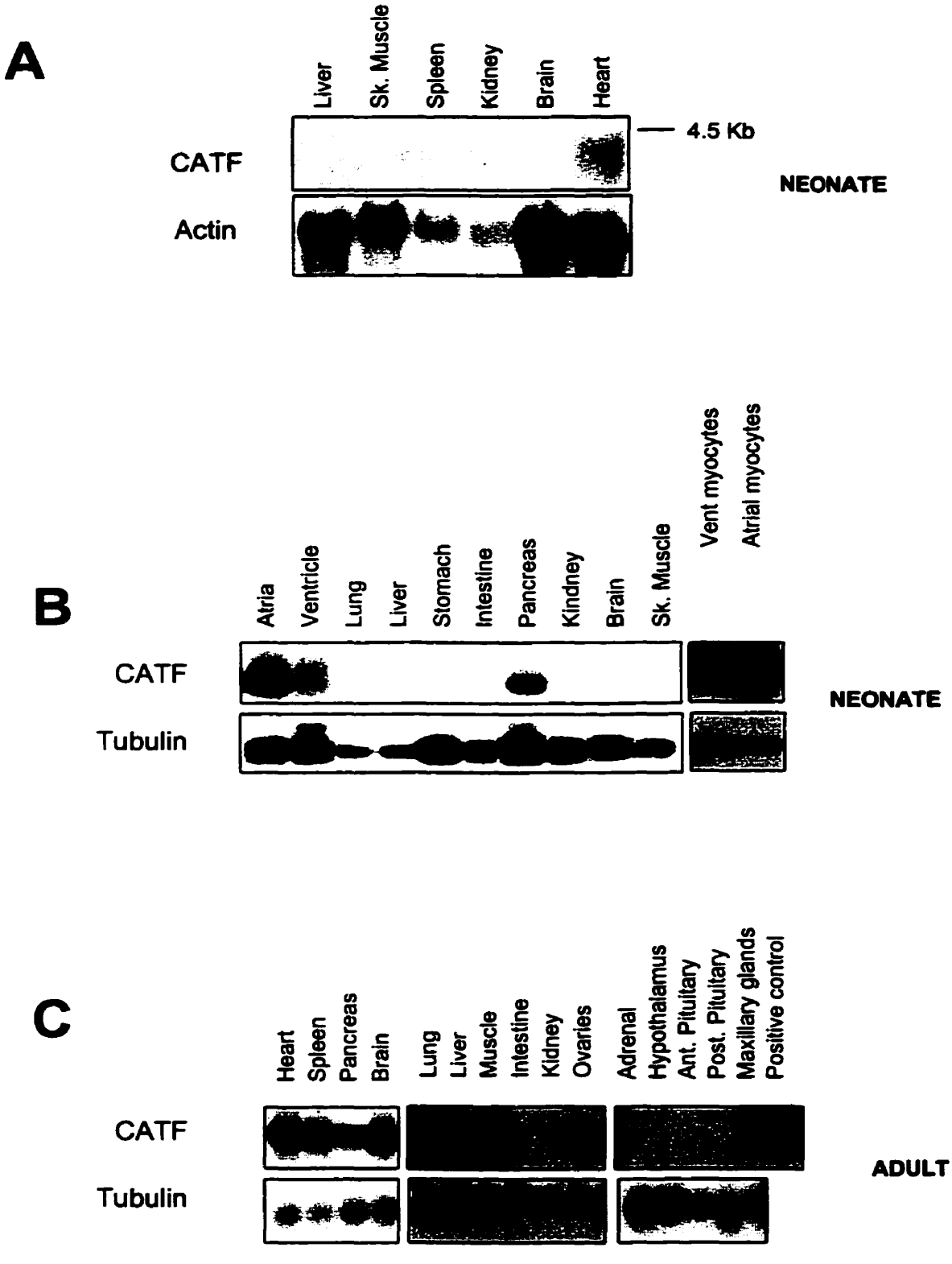
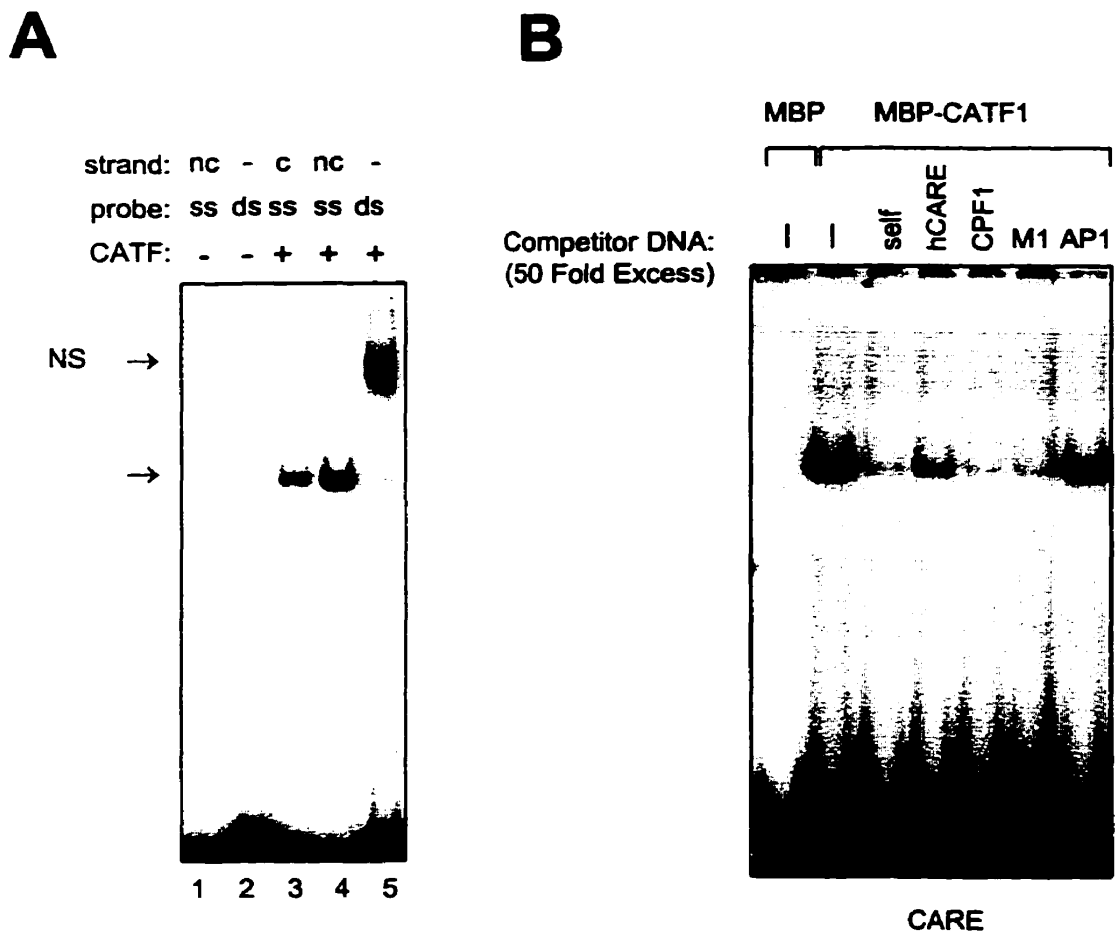

A
Liver
Sk. Muscle
Spleen
Kidney
Brain
Heart
4.5 Kb
CATF
Actin
NEONATE
B
Atria
Ventricle
Lung
Liver
Stomach
Intestine
Pancreas
Kindney
Brain
Sk. Muscle
Vent myocytes
Atrial myocytes
CATF
Tubulin
NEONATE
C
Heart
Spleen
Pancreas
Brain
Lung
Liver
Muscle
Intestine
Kidney
Ovaries
Adrenal
Hypothalamus
Ant. Pituitary
Post. Pituitary
Maxillary glands
Positive control
CATF
Tubulin
ADULT

Figure 4.3.

DNA binding properties of CATF1. (**A**) CATF1 binds preferentially ssDNA. Bacterially produced His-CATF protein was assayed for the binding of various single-stranded or double-stranded rCARE probes by electromobility shift assay (EMSA). NC is for non-coding and c for coding strand. D is double stranded 26 bp rCARE. NS represent non-specific binding. (**B**) CATF1 binds DNA with sequence-specificity. Approximately 50 ng of affinity-purified MBP or MBP-CATF1 protein were assayed for their ability to bind ssCARE oligonucleotide in the presence or absence of a 100-fold excess of unlabelled competitor oligonucleotides. The competitors were the homologous probe rCARE (self), the human CARE element (hCARE), the M1 mutant, the CPF1 element from cTpC (43) and the collagenase AP-1 probe (1).


A
strand: nc - c nc -
probe: ss ds ss ss ds
CATF: - - + + +
NS
1 2 3 4 5
B
MBP
MBP-CATF1
Competitor DNA:
(50 Fold Excess)
self
hCARE
CPF1
M1
AP1
CARE

were with those of the helicase superfamily I (comprising the (+) RNA viral helicase proteins). Koonin (29) had proposed that CATF1 relatives (SEN1 and UPF1) would form a new family of helicases within the superfamily I. Using the new blast algorithm (3) significant homologies were obtained with the SEN1-like helicase proteins (Fig. 4.1B). All these helicases are from the helicase superfamily I and are related to each other at probabilities of non-relatedness $p<1e^{-20}$. Further analysis of sequences within and outside the helicase domain enable the classification of these proteins in 5 subgroups based on their amino acid homologies to which the probability of non-relatedness has been arbitrarily established at a minimum value of $p=1e^{-60}$. The first subgroup, the CATF1 group contains the vertebrate orthologs of CATF1 as well as a member from *S. cerevisiae* (Ykl017c/DIP1) and, interestingly, two members from the archea kingdom: the MJU67467 protein from *Methanococcus jannaschii* (8) and a non-annotated ORF from *Archaeoglobus fulgidus*. This subgroup can be distinguished from the other subgroups by many amino acids outside and inside the helicase motifs which suggest that they might all be orthologs (Fig. 4.1B, C). This subgroup is hereafter named the CATF1 subfamily. The second subgroup comprises putative orthologs of the well-characterized yeast Upf1 protein which is involved in nonsense-mediated mRNA decay (56). The third subgroup is defined by the yeast splicing regulator SEN1 (13). The fourth subgroup is composed of the yeast dna2 protein which is involved in DNA replication (7) and the fifth group comprises only the mouse Mov10/gb110 protein (38). Altogether, based on their sequence homology these subgroups form a new family of helicases, the UPF1 family. A UPF1 helicase signature can be obtained if one combine the signature for type I helicases plus the underlined residues in Fig 4.1B.

CATF1 is a nucleic acid-binding protein. CATF1 and its vertebrate orthologs were cloned by their ability to interact with DNA in a southwestern assay, suggesting already that CATF1 is a DNA-binding protein. To further document the DNA-binding properties of CATF1, we performed gel shift analyses on the full-length and truncated (470-989) recombinant CATF1 protein. As shown in Fig. 4.3A, on the left panel, CATF1 was assayed for its ability to bind either of the two single strands of the CARE element

(lanes 3 and 4) or to the double-stranded CARE element (lanes 5). CATF1 interacted with the two CARE single but not with the dsDNA oligonucleotides. The binding over the CARE probe was specific since it could be efficiently competed following the addition of a 50 fold excess of unlabeled CARE probe but not unrelated CARE sequences (Fig. 4.3A). Intriguingly, the mobility of the CATF1:DNA complex was different depending on the length of the probe used for the EMSA (Fig. 4.3A) suggesting that CATF1 might possess DNA-bending activity.

The DNA-binding specificity of CATF1 was further analyzed by challenging CATF1 DNA-binding activity with different mutant CARE oligonucleotides at a 50 fold excess. As shown in Fig. 4.3B, CATF1 binds specifically the CARE oligonucleotide since the binding of CATF1 could be efficiently competed by the homologous and M1 oligonucleotide and partially with the human CARE sequence whereas it is not competed at all with the unrelated AP-1 binding site. Interestingly, the CPF1 oligonucleotide which corresponds to a cardiac-restricted element homologous to CARE in the cardiac troponin C (cTpC) promoter, also competes efficiently CATF1 binding to DNA.

CATF1 has nucleic acid-stimulated ATPase and helicase activities. The primary structure of CATF1 suggests that it is a putative helicase. Helicases separate DNA strands in an energy-driven mechanism that requires NTP hydrolysis. Thus we tested whether bacterially expressed CATF1 harbored ATPase activity by monitoring the release of inorganic phosphate from γ-labeled ATP (Fig. 4.4A). The release of P_i was detected on PEI-cellulose thin layer chomatography followed by autoradiography. As seen in Fig. 4.4A, in the absence of nucleic acid CATF1 and a CATF1 derivative harboring a mutation in the essential lysine (K281) of the ATP binding site (CATF1 KR), exhibited comparable low levels of ATP hydrolysis (Fig. 4.4A and B). However, in the presence of the nucleic acid poly(dI-dC), only the wild-type fusion protein could stimulate (by more than 10 fold) ATP hydrolysis. Moreover, ATP hydrolysis was nearly linear with time (Fig. 4.4B) and was also proportional to the amount of His-CATF1 added (not shown).

The ATPase activity of the recombinant CATF1 protein could be stimulated by all the nucleic acids tested so-far although higher ATPase activity could be obtained when nucleic acids containing a higher proportion of dG and dC were used (compare dI-dC and dG-dC vs dA-dT); moreover, a bias for deoxyribonucleotides vs ribonucleotides was noted though further studies with polymers of ribonucleotides (e.g poly(G-C)) are required to unequivocally determine the sequence specificity for the ATPase cofactor. The DNA-binding-site for CATF1, CARE, did not stimulate the ATPase activity more than the synthetic polymers (data not shown), implying either that specific DNA-binding is not involved in the stimulation or that the nucleic acids cofactor binds a distinct site in the CATF1 protein. Nonetheless, these data collectively demonstrate that CATF1 is a nucleic acid-stimulated ATPase.

The DNA-binding activity of CATF1 is inhibited by ATP-binding. In the majority of cases, the ss or dsDNA binding activities of helicases are stimulated by either ATP binding or ATP hydrolysis (34). We thus tested whether the DNA-binding activity of the full-length recombinant CATF1 (His-CATF1) or a truncated CATF1 that does not harbor most of the helicase subdomains (MBP-CATF1$_{470-989}$) was affected by the addition of ATP in an EMSA reaction. Surprisingly, as shown in Fig. 4.5A, the addition of 1 mM ATP to the EMSA reaction selectively abolished the binding to DNA of the full-length CATF1 fusion protein but not of the ATPase deficient CATF1 fusion, implying that either ATP binding or ATP hydrolysis was negatively regulating DNA-binding activity. This inhibition could be obtained at physiological concentration of NTPs since titration of the ATP concentration revealed an IC_{50} for the binding of CATF1 to DNA of 27 μM. Interestingly, a DNA-binding activity immunologically related to CATF1 in cardiac whole-cell extracts was also found to be affected by the addition of ATP in the EMSA reaction (Fig. 4.5A) strongly implying that this complex contains endogenous CATF1 protein and that endogenous CATF1 DNA-binding activity is also regulated by ATP.

In order to determine the mechanism by which ATP was affecting the DNA-binding activity of CATF1 we analyzed the effect of ATP, or the non-hydrolysable analog ATPγS on the on-rate and off-rate kinetics of CATF1 DNA-binding activity. The effect on

the off-rate was tested by monitoring the decay of CATF1:DNA complexes after incubation of pre-formed full-length CATF1:DNA complexes with 1 mM ATP. As seen in Fig 4.5C, the half-life of the CATF1:DNA complex was approximately 15 minutes in the absence of ATP. The addition of ATP or ATPγS did not significantly affect the decay of the protein:DNA complex. In the case of the on-rate, ATP or ATPγS was incubated together with CATF1 and the CARE probe for increasing amounts of time; Fig. 4.5D clearly demonstrates that the addition of either ATP or ATPγS drastically affects the formation of the CATF1:CARE complex. Since ATPγS was as efficient as ATP in inhibiting complex formation, the mechanism by which ATP inhibits DNA-binding is likely to be triggered by a conformational change of the CATF1 protein upon ATP binding. This would also be consistent with the fact that it is the on-rate but not the off-rate (which would probably require ATP hydrolysis) that is affected.

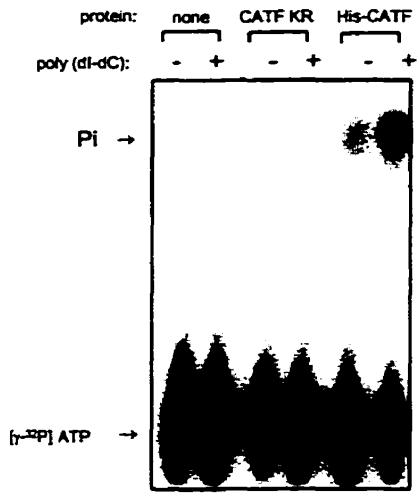
CATF1 is a *bona fide* helicase. The ability of CATF1 to disrupt the structure of DNA duplexes was determined using a strand displacement assay (47). In this assay, the helicase activity is monitored by the disruption of the DNA duplex formed by an annealed ^{32}P-labeled extended nucleotide on M13mp18 ssDNA, in the presence of CATF1 and ATP. As shown in Fig. 4.6, in the absence of CATF1, the DNA substrate was almost completely in the duplex state; however, when recombinant CATF1 was added in the presence of ATP, up to 15% of the oligonucleotide was displaced. This displacement required energy as the removal of ATP totally impaired displacement of the oligonucleotide and the ATPase-deficient CATF1 KR fusion protein resulted in only background displacement activity. The DNA displacement activity also increased in an almost linear fashion with time. These data indicate that CATF1 possesses *bona fide* helicase activity.

CATF1 activates the ANF promoter. As stated earlier, very few members of the helicase family also possess DNA-binding and transactivation properties. In order to examine whether CATF1 is a transcriptional regulator, the full-length CATF1 and ATPase deleted CATF1 fragment (aa 470-989), was inserted in a CMV-driven expression vector and co-transfected with various ANF reporter constructs. Intriguingly, only the truncated

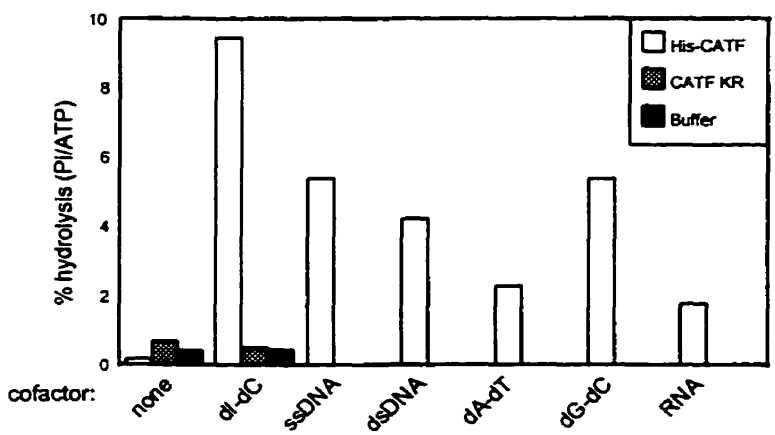
Figure 4.4.

CATF1/Sμbp-2 is a nucleic-acid stimulated ATPase. (**A**) ATPase activity was measured by the release of inorganic phosphate (P_i) from [γ-^{32}P]ATP (as described in Material and Methods). ATPase reactions were carried out in the presence (+) or in absence (-) of poly(dI-dC) as a nucleic acid cofactor. (**B**) Cofactor specificity of the ATPase activity of CATF1. ATPase assays were carried out without (none) or with 1 μg of various nucleic acids as cofactors. Open bars represent assays using 300 ng of His-CATF1, gray bars are assays using the same amount of the ATPase mutant CATF1 KR and black bars are without any CATF1 protein. Cofactors added were poly(dI-dC) [dI-dC], M13mp18 [ssDNA], pBluescript KS [dsDNA], poly(dA-dT) [dA-dT], poly(dG-dC) [dG-dC] and total cellular RNA [RNA]. Results are expressed as percentage of hydrolysis (P_i hydrolyzed/ATP) and represent the mean of the duplicate of a highly reproducible experiment. (**C**) Time course of ATP hydrolysis. The His-CATF1 (CATF1) and His-CATF1-KR mutant (CATF1 KR) were assayed for ATPase activity as described in (A) but the reaction was stopped at different time points. ATPase activity was quantitated on a PhosphorImager.

A


protein:
none
CATF KR
His-CATF
poly (dI-dC):
- + - + - +
Pi
[γ-32P] ATP

B


% hydrolysis (Pi/ATP)
10
8
6
4
2
0
His-CATF
CATF KR
Buffer
cofactor:
none
dI-dC
ssDNA
dsDNA
dA-dT
dG-dC
RNA

C

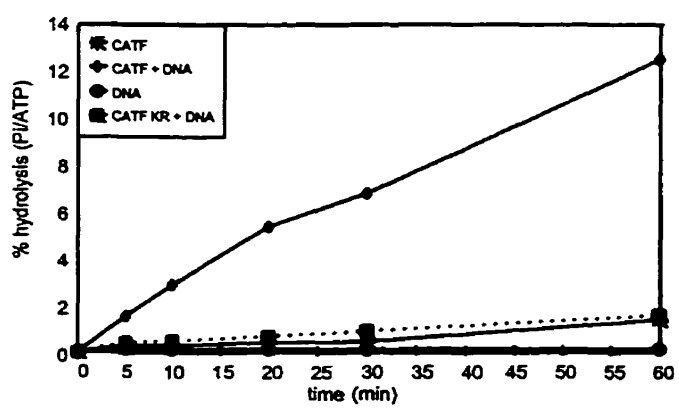
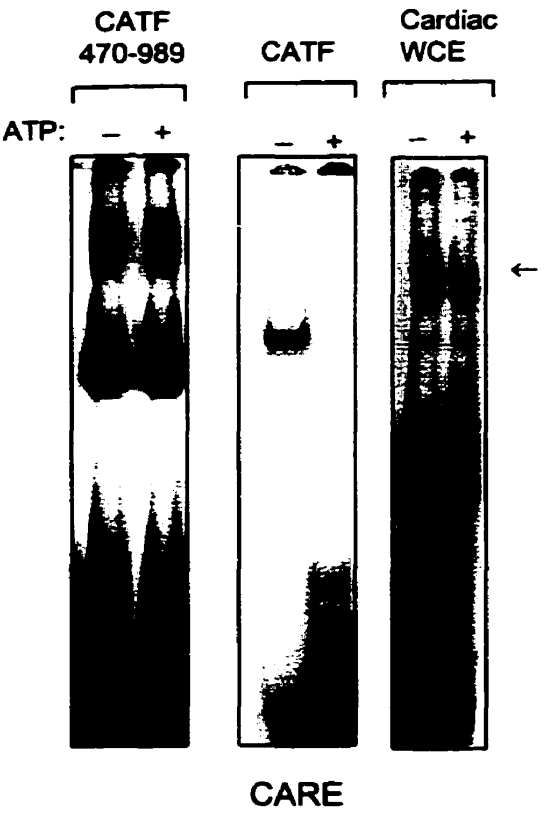

% hydrolysis (Pi/ATP)
14
12
10
8
6
4
2
0
CATF
CATF + DNA
DNA
CATF KR + DNA
0 5 10 15 20 25 30 35 40 45 50 55 60
time (min)

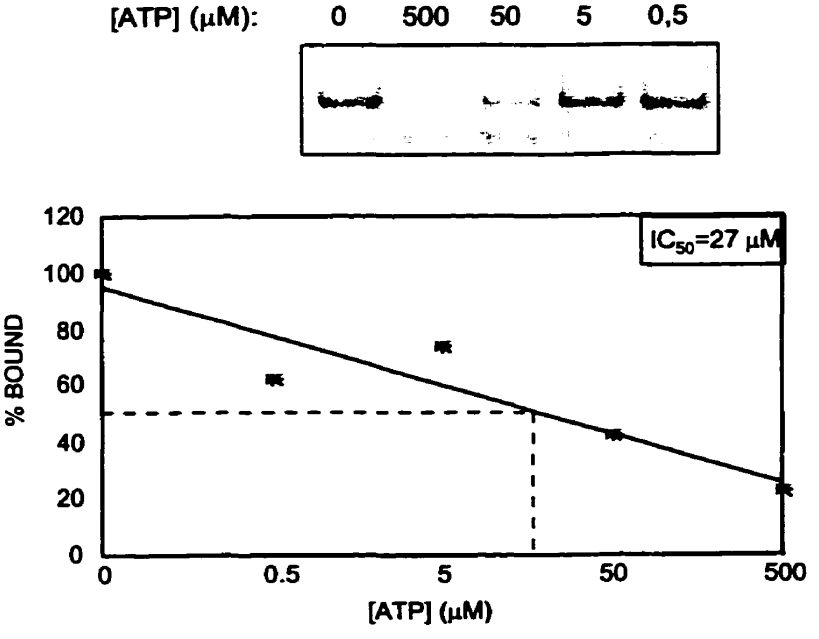
Figure 4.5.

CATF1/Sμbp-2 DNA-binding activity is affected by the binding of ATP. **(A)** Recombinant $CATF1_{470-989}$ (First panel) or CATF1 fusion proteins (second panel) were assayed for their ability to bind the ssCARE element after a preincubation with (+) or without (-) 1 mM ATP. On the far right panel, DNA-binding activity of endogenous CARE-binding proteins present in cardiac whole cell extracts (WCE) were challenged by the pre-incubation of extracts with 1 mM ATP. The lower mobility complex proposed to represent CATF1 or CATF1-related molecules (arrow) and is specifically affected by the addition of ATP. EMSAs were performed as described in Materials and Methods. **(B)** ATP inhibits CATF1 DNA-binding at physiological concentrations. Increasing amounts of ATP were added prior to addition of the ssCARE probe and EMSA was carried out as described in Fig. 4.2. DNA binding activity was quantitated on PhosphorImager and expressed (bottom panel) relative to the binding without ATP (100% bound). The IC_{50} is defined as the concentration required to inhibit 50% of the binding and can be intrapolated at 27 μM of ATP. **(C)** ATP does not affect the off-rate. Off-rate analyses were carried out by assembling an EMSA reaction as described in Material and Methods. ATP or the non-hydrolysable ATP analog ATPγS (1 mM final) was added after 20 min. of incubation to the EMSA reaction mixture. **(D)** ATP binding, but not hydrolysis affects the on-rate of DNA-binding. On-rate reactions were carried out by adding the ATP or ATPγS prior to the addition of the EMSA reaction as described in Material and Methods.

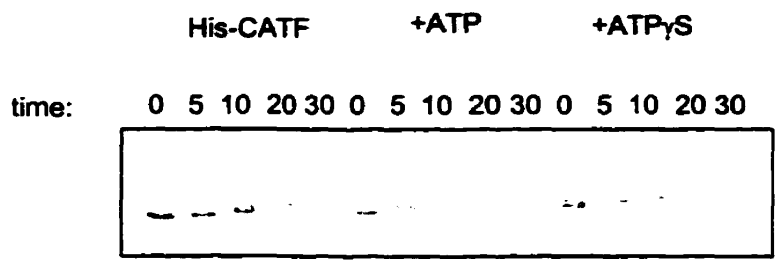
A


CATF
470-989
CATF
Cardiac
WCE
ATP: – + – + – +
←
CARE

B

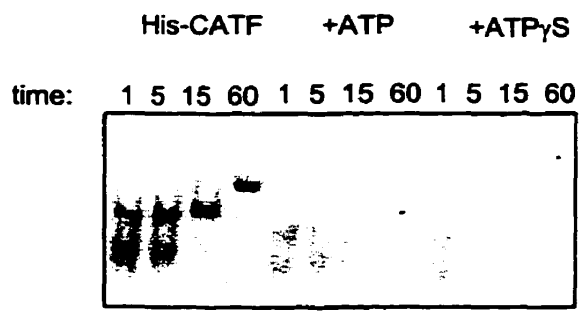

[ATP] (μM): 0 500 50 5 0,5
IC50=27 μM
% BOUND
120
100
80
60
40
20
0
0
0.5
5
50
500
[ATP] (μM)

C



Off-rate

D

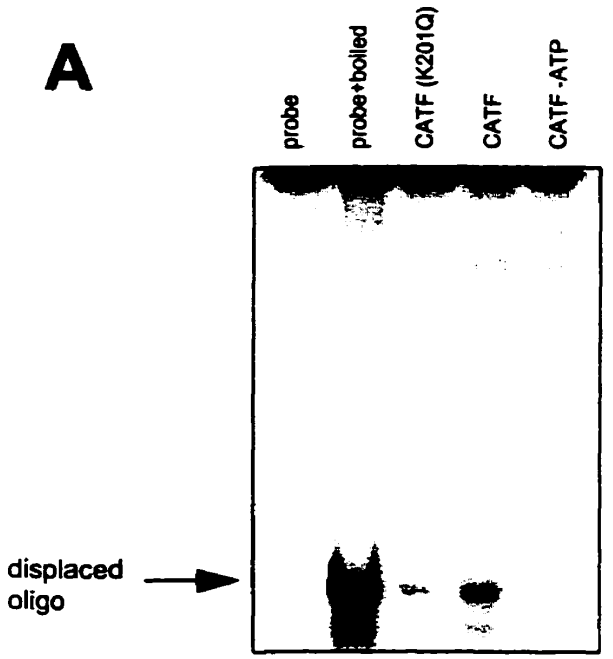


On-rate

Figure 4.6..

CATF1/Sμbp-2 is a helicase. **(A)** Helicase activity was measured by strand displacement assay. Wild-type His-CATF1 or ATPase-deficient CATF1 KR were incubated with the helicase substrate. The substrate used is an oligonucleotide with a 5' single-stranded extension annealed to M13mp18 and was labeled as described in the Materials and Methods section. The probe was boiled as a control for displacement. The assay was carried in the presence or absence of ATP (-ATP). **(B)** Time course of the helicase activity. The His-CATF1 (CATF1) fusion protein was assayed for helicase activity as described in (A). Helicase reactions were run on a denaturing gel, dried and the percentage of displacement was measured on a PhosphorImager (Molecular Dynamics).

A


probe
probe+boiled
CATF (K201Q)
CATF
CATF -ATP
displaced
oligo

B

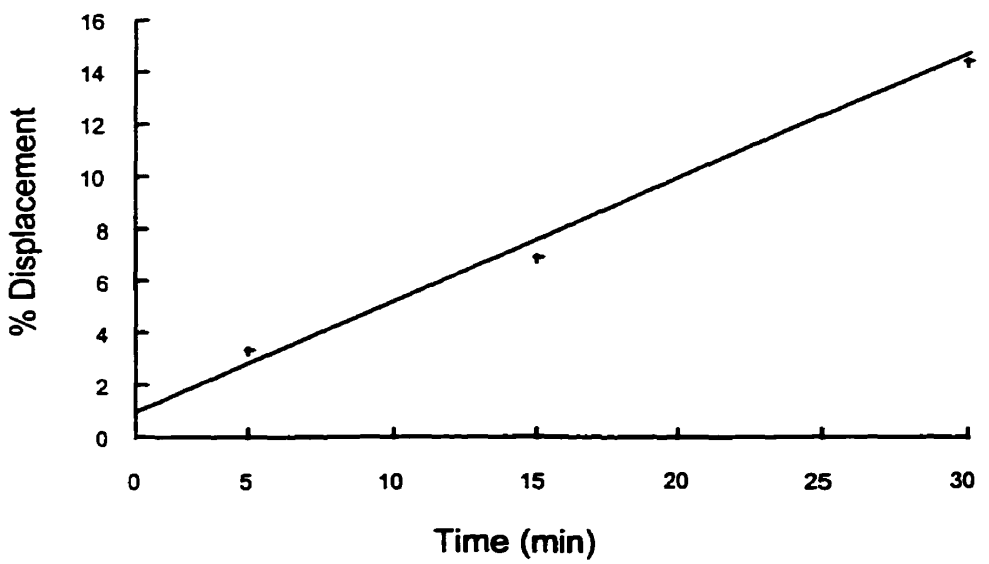
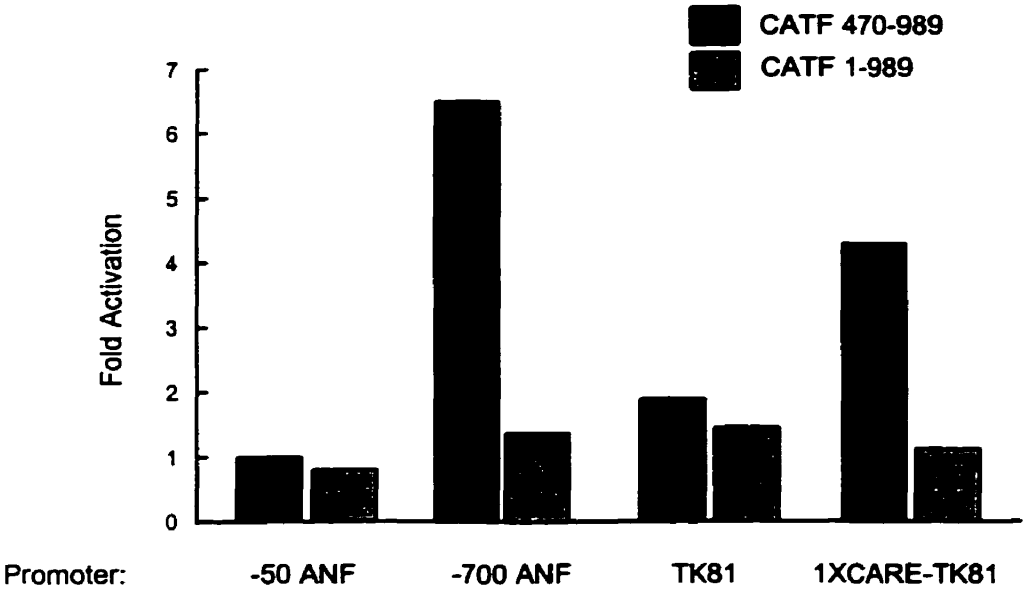

% Displacement
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30
Time (min)

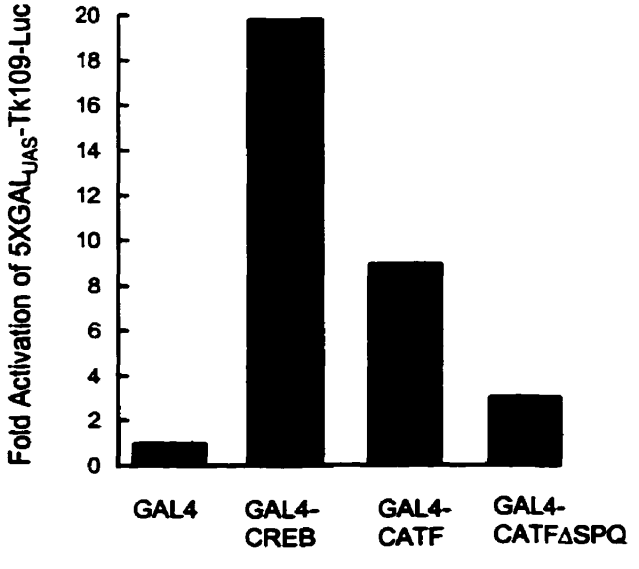
Figure 4.7.

CATF1 possesses transactivating properties. **(A)** CATF1 transactivation is targeted at the CARE element. CMV-driven CATF1 or $CATF1_{470\text{-}989}$ expression vectors were cotransfected transiently in NIH 3T3 cells with either the full-length ANF cardiac promoter (-700), the minimal ANF promoter (-50), TK81-luciferase or 3XCARE-TK81-luciferase reporters. Activation by either truncated $CATF1_{470\text{-}989}$ (black bars) or full-length CATF1 (gray bars) is represented as Fold Activation of the reporter over the CMV backbone vector. Results represent the mean of at least six independent experiments. **(B)** CATF1 contains a transactivation domain in the carboxy-terminus. 3 μg/dish of RSV-driven GAL4-CATF1 containing the C-terminal domain of CATF1 (aa 470-989), a truncated C-terminal domain (aa 470-751) or $GAL4\text{-}CREB_{\Delta KID}$, were cotransfected with 1.5 μg/dish of the reporter $5XGAL_{UAS}$-Tk109-luciferase in NIH 3T3 cells (as described in the Material and Methods section). Data represent the mean of four independent experiments. **(C)** Schematic representation of the CATF1 protein. The black boxes represent the seven helicase domains. The PQ box represents a region rich in proline and glutamine that is not conserved in any other UPF1 family members. **(D)**. A homologous PQ domain is found in another helicase. Advanced BLAST searches (filter off) revealed that the proline and glutamine rich region which is not conserved in other UPF1 family helicases is found on a novel helicase KIA0054 (accession # D29677).

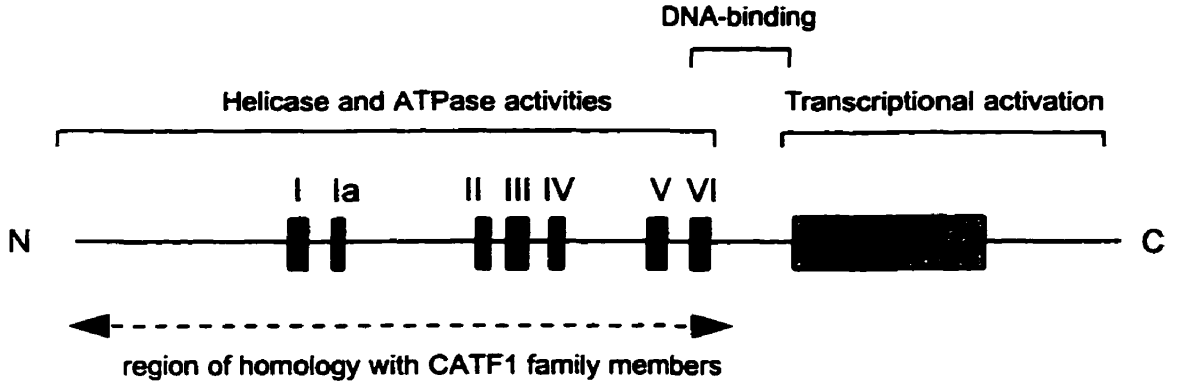
A


CATF 470-989
CATF 1-989
Fold Activation
0
1
2
3
4
5
6
7
Promoter:
-50 ANF
-700 ANF
TK81
1XCARE-TK81

B


Fold Activation of 5XGAL$_{UAS}$-Tk109-Luc
0
2
4
6
8
10
12
14
16
18
20
GAL4
GAL4-CREB
GAL4-CATF
GAL4-CATFΔSPQ

C



D

```
rCATF      793 PQLPSPPSPAQAEPEPLSQQPLGQPHCSTQLDLKALHLQRLQRQQGSQAQPAKAQ 847
               ||||:|| |   :   |:||    |    |:  :   : : | |   | |    |
KIAA0054  1363 PQLPRPPFPIPQQHTLLNQQQNNLPEQPNQIPPQPNQVVQQQSQLNQQPQQPPPQ 1418
```

CATF1 protein activated the -700 bp in a dose-dependent manner (Figure 4.7A and data not shown). The minimal -50 bp ANF promoter was not activated implying that CATF1-dependent activation requires the -50 to -700 bp promoter. This region contains at least one copy of the rCARE element. To examine if $CATF1_{470-989}$ has the ability to transactivate the CARE element, a heterologous promoter consisting of one copy of CARE fused upstream of the minimal thymidine kinase promoter, Tk81, was prepared. CATF1 activated 4.3 fold this heterologous construct. These results indicate that CATF1 possesses transcriptional activation properties that may be modulated by the N-terminal region. The presence of an autonomous activation domain on CATF1 was further confirmed by fusing the coding sequence of CATF1 between amino acids 470 and 989 in frame with the heterologous DNA-binding domain of GAL4. Transcriptional activation by CATF1 protein could be conferred to the heterologous DNA binding domain of GAL4 (Figure 4.7B) confirming that CATF1 is a *bona fide* transcriptional activator. Furthermore this activation domain was localized in the non-conserved carboxy terminal region of CATF1 since a deletion removing the last 200 aa (Δ752-989) severely impaired the ability of GAL4-CATF1 to transactivate the GAL_{UAS} reporter. This region contains a proline/glutamine-rich stretch located between aa 793-847 that shows high homology to another human helicase KIA0054, (accession # D29677) (Fig. 4.7D). Thus, CATF1 harbors properties of transcriptional activators and the ANF gene is likely to be a target for CATF1 in cardiac cells.

DISCUSSION

Helicases play central roles in all aspects of gene expression regulation and nucleic acid metabolism (34). Here we report the isolation of CATF1, a new *bona fide* helicase whose expression is regulated during development and which defines a new family of DNA/RNA helicases that harbor a high degree of sequence homology as well as sharing distinctive

biochemical properties. The CATF1 protein is thus multifunctional, harboring a functional helicase domain and properties of a sequence-specific transcriptional activator.

A new family of DNA/RNA helicases. Based on sequence analysis, three large superfamilies (superfamily I-III) of helicases and nucleic acid-dependent NTPases have been identified to date. CATF1 protein is part of the helicase superfamily I (SF1) and defines a new subgroup of closely related molecules from the archeal and eucaryal kingdoms. Interestingly, the helicase superfamily I is mainly composed of bacterial and viral helicases, except for UPF1 family members that are found in eukaryotes and archea, thus the helicase superfamily I is an ancient, conserved, protein motif. The UPF1 family was first characterized solely based on sequence analyses though classification of proteins in different families is informative only when functional properties are also taken into account. In this study, we provide a molecular basis for the classification of the CATF1 and UPF1 relatives into the same family. They both can bind ssDNA and RNA and they both harbor helicase and nucleic-acid stimulated ATPase activities. More importantly, the distinction between the UPF1 family and the rest of the helicase families is the ability of ATP to inhibit the DNA-binding of CATF1 and Upf1p. This inhibition affects the on-rate of CATF1 binding to DNA and is caused by the binding, and not the hydrolysis, of ATP. This is consistent with the results obtained for Upf1p (55,56), for which it was demonstrated that it is the binding of ATP rather than its hydrolysis that causes ssDNA-binding inhibition since mutants disrupting the helicase activity of Upf1p are still sensitive to the addition of ATP during the assembly of the EMSA reaction. Since UPF1 family members are involved in distinct biological processes, these conserved properties may provide significant insights about the molecular mechanisms by which they exert their activity.

The presence of two putative archeal proteins that are highly conserved with the yeast and mammalian CATF1 suggests that CATF1 plays an important evolutionarily conserved role in cellular physiology. Extensive blast searches in eubacterial genome databases were unable to detect a prokaryotic homolog of CATF1 (or any of the UPF1 family) suggesting that the cellular process (or its regulation) in which CATF1 is involved

evolved after the divergence of the prokaryotic and the archeal-eukaryotic ancestors and lends further support to the hypothesis --based on phylogenetic analyses of rRNA sequences-- that eucarya and archea evolved from a common ancestor. Interestingly, archea seem to be more similar to eukaryotes than to bacteria in all aspects of the control of gene regulation and expression but not in aspects regarding metabolism and its regulation where archea are much more similar to the bacteria (42). This reinforces the possibility that CATF1 is involved in regulating gene expression.

CATF1 is a multifunctional protein. We (Fig. 4.1), and others (28,50) have cloned CATF1 by virtue of its interaction with a labeled oligonucleotide indicating that CATF1 has DNA-binding properties. In this work we presented data supporting the notion that CATF1 can act as a sequence-specific transcriptional activator. Indeed, we observed that a truncated form of CATF1 lacking the helicase domain and containing mainly the divergent C-terminal domain and the nucleic-acid binding domain, was able to transactivate the ANF promoter or a heterologous promoter containing a CATF1 binding site (Fig. 4.7). In addition, we characterized a biochemical activity that might regulate the transcriptional properties in a novel manner i.e. the ATP interference with CATF1 DNA-binding activity. This interesting property of the CATF1 protein may explain why the protein works as an activator solely when the helicase region is deleted. However, it is not clear yet if the helicase region of CATF1 is solely involved in the regulation of the transcriptional activities of the protein or whether CATF1 plays additional roles in cellular physiology. Interestingly, a number of transcriptional regulators have been shown to possess activities other than those associated with the regulation of transcription. These proteins are involved in many aspects of cellular physiology such as DNA repair, replication and even translation. For example, the excision repair pathway is closely linked to transcription and it was shown that the general transcription factor TFIIH plays a preponderant role in this aspect of DNA-repair (48); the protein complex involved in double-strand break repair, DNA-PK is also involved in transcriptional regulation (21). Moreover, the anti-oncogene p53 which is a transcriptional regulator of a number of genes involved in cell cycle control, apoptosis and DNA repair has also been demonstrated to be

directly implicated in DNA-repair by its DNA-DNA annealing activity (41). In addition, a number of transcriptional regulators were also found to be implicated in the initiation of viral and cellular replication; the OCT1 and CTF/NF1 proteins bind their cognate binding sites near the origins of replication where they are thought to play important roles in recruiting the replicating apparatus (39,45). In some cases, translation can also be regulated by transcription factors as in *Drosophila* where the bicoid homeoprotein has recently been shown to directly regulate the translation of the *caudal* messenger RNA by binding directly its 3' UTR (14,46).

The ssDNA binding activity of CATF1 is not incompatible with transcriptional regulation. The transcriptional activation of promoters by single-stranded DNA-binding proteins have recently been demonstrated for the FBP family of proteins that bind and activate transcription through the Far Upstream Element originally described in the c-myc promoter (11,15). Moreover, the hnRNP K, a ssDNA-binding protein can also activate transcription in a transient transfection assay (33,53). It is noteworthy that the mammalian CATF1 contains a C-terminal tail that is not conserved in more distantly related species such as in *S. cerevisiae*, *M. jannaschii* or *A. fulgidus*, which suggests that CATF1 acquired additional or new functions during evolution. This C-terminal tail contains a region rich in prolines and glutamines that is reminiscent of a number of transcriptional activation domains; this region can confer transcriptional activation to a heterologous DNA-binding domain implying it contains an autonomous activation domain. Interestingly, the PQ domain can be found in another putative helicase, the KIA0054 protein of yet unknown function and it would be worthwhile to look if this domain also acts as an activation domain in this protein and might define a new type of multifunctional helicases.

CATF1 is a cardiac-enriched member of the UPF1-related helicases. To date, UPF1 family helicases have been shown to be ubiquitously expressed except for the distantly related UPF1 family member, Mov10 which has a tissue-restricted mRNA distribution. CATF1 transcripts (and protein) are present in cardiac myocytes at all stages of post-natal development. However, in neonatal (Fig. 4.2A) and embryonic tissues (data not shown) the CATF1 mRNA seems to be restricted to the myocardium and

transcriptional activation of the ANF promoter by CATF1 identifies at least one putative target for it in the cardiac muscle. Thus CATF1 seems to be part of a very select set of transcriptional regulators, including GATA-4, Nkx2-5 and MEF2 proteins whose expression is enriched or restricted to the myocardium at some point during development. Given that the ANF promoter is a target for all these proteins, it would be interesting to test whether CATF1 cooperates with or modulates activity of these regulators.

ACKNOWLEDGMENTS

We are grateful to Dr Ariel Arias for preparation of the anti CATF1 antibody, Michel Chamberland for oligonucleotide synthesis, to Mireille Hogue for technical assistance and to Lise Laroche for expert secretarial assistance. We thank Kevin McBride for critical reading of this text and members of the Nemer lab for continuous support. This work was supported by grants from the Medical Research Council (MRC) of Canada and Société de Recherches sur le Cancer Inc. (S.R.C). D.D. is a recipient of the 1967 Fellowship from the Natural Sciences and Engineering Research Council of Canada (NSERC). J.-M.J. was a MRC postdoctoral fellow and M.N is an MRC scientist.

REFERENCES

1. **Angel, P., M. Imagawa, R. Chiu, B. Stein, R. J. Imbra, H. J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin**. 1987. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell **49**:729-739.

2. **Argentin, S., A. Ardati, S. Tremblay, I. Lihrmann, L. Robitaille, J. Drouin, and M. Nemer**. 1994. Developmental stage-specific regulation of atrial natriuretic factor gene transcription in cardiac cells. Mol. Cell. Biol. **14**:777-790.

3. **Atschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman**. 1997. Grapped BLAST and PSI-BLAST: anew generation of protein database search programs. Nucl. Acids. Res. **25**:3389-3402.

4. **Baertschi, A. J., J. M. Adams, and M. P. Sullivan**. 1988. Acute hypoxemia stimulates atrial natriuretic factor secretion in vivo. Am. J. Physiol. **255**:H295-H300.

5. **Bell, J. R., A. Noveen, Y. H. Liu, L. Ma, S. Dobias, R. Kundu, W. Luo, Y. Xia, A. J. Lusis, M. L. Snead, and et al**. 1993. Genomic structure, chromosomal location, and evolution of the mouse Hox 8 gene. Genomics **17**:800

6. **Bode, J., T. Schlake, M. Rios-Ramirez, C. Mielke, M. Stengert, V. Kay, and D. Klehr-Wirth**. 1995. Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. Int. Rev. Cytol. **162A**:389-454.

7. **Budd, M. E. and J. L. Campbell**. 1995. A yeast gene required for DNA replication encodes a protein with homology to DNA helicases. Proc. Natl. Acad. Sci. USA **92**:7642-7646.

8. **Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, and J. C. Venter**. 1996. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science **273**:1058-1073.

9. **Chomczynski, P. and N. Sacchi**. 1987. Single-step method of RNA isolation by acid guanidium thyocyanate-phenol-chloroform-extraction. Anal. Biochem. **162**:156-159.

10. **Dagnino, L., J.-P. Lavigne, and M. Nemer**. 1992. Increased transcripts for B-type natriuretic peptide in spontaneously hypertensive rats. Hypertension **20**:690-700.

11. **Davis-Smyth, T., R. C. Duncan, T. Zheng, G. Michelotti, and D. Levens**. 1996. The far upstream element-binding proteins comprise an ancient family of single-strand DNA-binding transactivators. J. Biol. Chem. **271**:31679-31687.

12. **Delmas, V., D. G. Stokes, and R. P. Perry**. 1993. A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proc. Natl. Acad. Sci. USA **90**:2414-2418.

13. **DeMarini, D. J., M. Winey, D. Ursic, F. Webb, and M. R. Culbertson**. 1992. SEN1, a positive effector of tRNA-splicing endonuclease in Saccharomyces cerevisiae. Mol. Cell. Biol. **12**:2154-2164.

14. **Dubnau, J. and G. Struhl**. 1996. RNA recognition and translational regulation by a homeodomain protein. Nature **379**:694-699.

15. **Duncan, R., L. Bazar, G. Michelotti, T. Tomonaga, H. Krutzsch, M. Avigan, and D. Levens**. 1994. A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. Genes Dev. **8**:465-480.

16. **Durocher, D., C. Y. Chen, A. Ardati, R. J. Schwartz, and M. Nemer**. 1996. The ANF promoter is a downstream target for Nkx-2.5 in the myocardium. Mol. Biol. Cell **16**:4648-4655.

17. **Durocher, D., C. Grépin, and M. Nemer**. 1997. Regulation of gene expression in the endocrine heart. Rec. Prog. Horm. Res. **In press**:

18. **Fanning, E.** 1992. Simian virus 40 large T antigen: the puzzle, the pieces, and the emerging picture. J. Virol. **66**:1289-1293.

19. **Fukita, Y., T. R. Mizuta, M. Shirozu, K. Ozawa, A. Shimizu, and T. Honjo**. 1993. The human S mu bp-2, a DNA-binding protein specific to the single-stranded guanine-rich sequence related to the immunoglobulin mu chain switch region. J. Biol. Chem. **268**:17463-17470.

20. **Genersch, E., C. Eckerskorn, F. Lottspeich, C. Herzog, K. Kuhn, and E. Poschl.** 1995. Purification of the sequence-specific transcription factor CTCBF, involved in the control of human collagen IV genes: subunits with homology to Ku antigen. EMBO J. **14**:791-800.

21. **Giffin, W., H. Torrance, D. J. Rodda, G. G. Prefontaine, L. Pope, and R. J. Hache.** 1996. Sequence-specific DNA binding by Ku autoantigen and its effects on transcription. Nature **380**:265-268.

22. **Gong, X., S. Kaushal, E. Ceccarelli, N. Bogdanova, C. Neville, T. Nguyen, H. Clark, Z. A. Khatib, M. Valentine, A. T. Look, and N. Rosenthal.** 1997. Developmental regulation of Zbu1, a DNA-binding member of the SWI2/SNF2 family. Dev. Biol. **183**:166-182.

23. **Gorbalenya, A. E. and E. V. Koonin.** 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Struct. Biol. **3**:419-429.

24. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov.** 1988. A conserved NTP-motif in putative helicases. Nature **333**:22

25. **Grépin, C., L. Dagnino, L. Robitaille, L. Haberstroh, T. Antakly, and M. Nemer.** 1994. A hormone-encoding gene identifies a pathway for cardiac but not skeletal muscle gene transcription. Mol. Cell. Biol. **14**:3115-3129.

26. **Harlow, E. and D. Lane.** 1988. 5. Immunizations. p. 53-135. In AnonymousAntibodies, a laboratory manual. Cold Spring harbor Laboratory, Cold Spring Harbor.

27. **Hodgman, T. C.** 1988. A new superfamily of replicative proteins. Nature **333**:22-23.

28. **Kerr, D. and K. Khalili.** 1991. A recombinant cDNA derived from human brain encodes a DNA binding protein that stimulates transcription of the human neurotropic virus JCV. J. Biol. Chem. **266**:15876-15881.

29. **Koonin, E. V.** 1992. A new group of putative RNA helicases. Trends Biochem. Sci. **17**:495-497.

30. **Kuhn, A., T. M. Gottlieb, S. P. Jackson, and I. Grummt.** 1995. DNA-dependent protein kinase: a potent inhibitor of transcription by RNA polymerase I. Genes Dev. **9**:193-203.

31. **Kuzuoka, M., T. Takahashi, C. Guron, and R. Raghow**. 1994. Murine homeobox-containing gene, Msx-1: analysis of genomic organization, promoter structure, and potential autoregulatory cis-acting elements. Genomics **21**:85-91.

32. **Laurent, B. C., I. Treich, and M. Carlson**. 1993. The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. Genes Dev. **7**:583-591.

33. **Lee, M. H., S. Mori, and P. Raychaudhuri**. 1996. trans-Activation by the hnRNP K protein involves an increase in RNA synthesis from the reporter genes. J. Biol. Chem. **271**:3420-3427.

34. **Lohman, T. M. and K. P. Bjornson**. 1996. Mechanisms of helicase-catalyzed DNA unwinding. Annu. Rev. Biochem. **65**:169-214.

35. **Matson, S. W. and K. A. Kaiser-Rogers**. 1990. DNA helicases. Annu. Rev. Biochem. **59**:289-329.

36. **McBride, K., L. Robitaille, S. Tremblay, S. Argentin, and M. Nemer**. 1993. Fos/jun repression of cardiac-specific transcription in quiescent and growth stimulated myocytes is targeted at a tissue-specific cis-element. Mol. Cell. Biol. **13**:600-612.

37. **Mizuta, T. R., Y. Fukita, T. Miyoshi, A. Shimizu, and T. Honjo**. 1993. Isolation of cDNA encoding a binding protein specific to 5'-phosphorylated single-stranded DNA with G-rich sequences. Nucleic Acids Res. **21**:1761-1766.

38. **Mooslehner, K., U. Muller, U. Karls, L. Hamann, and K. Harbers**. 1991. Structure and expression of a gene encoding a putative GTP-binding protein identified by provirus integration in a transgenic mouse strain. Mol. Cell. Biol. **11**:886-893.

39. **Mul, Y. M. and P. C. van der Vliet**. 1992. Nuclear factor I enhances adenovirus DNA replication by increasing the stability of a preinitiation complex. EMBO J. **11**:751-760.

40. **Niu, H. and S. T. Jacob**. 1994. Enhancer 1 binding factor (E1BF), a Ku-related protein, is a growth-regulated RNA polymerase I transcription factor: association of a repressor activity with purified E1BF from serum-deprived cells. Proc. Natl. Acad. Sci. U. S. A **91**:9101-9105.

41. **Oberosler, P., P. Hloch, U. Ramsperger, and H. Stahl**. 1993. p53-catalyzed annealing of complementary single-stranded nucleic acids. EMBO J. **12**:2389-2396.

42. **Olsen, G. J. and C. R. Woese**. 1997. Archeal genomics: an overview. Cell **89**:991-994.

43. **Parmacek, M. S., A. J. Vora, T. Shen, E. Barr, F. Jung, and J. M. Leiden**. 1992. Identification and characterization of a cardiac-specific transcriptional regulatory element in the slow/cardiac troponin C gene. Mol. Cell. Biol. **12**:1967-1976.

44. **Pazin, M. J. and J. T. Kadonaga**. 1997. SWI2/SNF2 and related proteins: ATP-driven motors that disrupt protein-DNA interactions? Cell **88**:737-740.

45. **Pruijn, G. J., W. van Driel, and P. C. van der Vliet**. 1986. Nuclear factor III, a novel sequence-specific DNA-binding protein from HeLa cells stimulating adenovirus DNA replication. Nature **322**:656-659.

46. **Rivera-Pomar, R., D. Niessing, U. Schmidt-Ott, W. J. Gehring, and H. Jackle**. 1996. RNA binding and translational suppression by bicoid. Nature **379**:746-749.

47. **Rozen, F., I. Edery, K. Meerovitch, T. E. Dever, W. C. Merrick, and N. Sonenberg**. 1990. Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. Mol. Cell. Biol. **10**:1134-1144.

48. **Schaeffer, L., R. Roy, S. Humbert, V. Moncollin, W. Vermeulen, J. H. Hoeijmakers, P. Chambon, and J. M. Egly**. 1993. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science **260**:58-63.

49. **Sheridan, P. L., M. Schorpp, M. L. Voz, and K. A. Jones**. 1995. Cloning of an SNF2/SWI2-related protein that binds specifically to the SPH motifs of the SV40 enhancer and to the HIV-1 promoter. J. Biol. Chem. **270**:4575-4587.

50. **Shieh, S. Y., C. M. Stellrecht, and M. J. Tsai**. 1995. Molecular characterization of the rat insulin enhancer-binding complex 3b2. Cloning of a binding factor with putative helicase motifs. J. Biol. Chem. **270**:21503-21508.

51. **Singh, H., J. H. LeBowitz, A. S. Baldwin,Jr., and P. A. Sharp**. 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52**:415-423.

52. **Stokes, D. G. and R. P. Perry**. 1995. DNA-binding and chromatin localization properties of CHD1. Mol. Cell. Biol. **15**:2745-2753.

53. **Tomonaga, T. and D. Levens**. 1996. Activating transcription from single stranded DNA. Proc. Natl. Acad. Sci. U. S. A **93**:5830-5835.

54. **Tuteja, N., R. Tuteja, A. Ochem, P. Taneja, N. W. Huang, A. Simoncsits, S. Susic, K. Rahman, L. Marusic, J. Chen, and et al.** 1994. Human DNA helicase II: a novel DNA unwinding enzyme identified as the Ku autoantigen. EMBO J. **13**:4991-5001.

55. **Weng, Y., K. Czaplinski, and S. W. Peltz.** 1996. Genetic and biochemical characterization of mutations in the atpase and helicase regions of the upf1 protein. Mol. Cell. Biol. **16**:5477-90: 11.

56. **Weng, Y., K. Czaplinski, and S. W. Peltz.** 1996. Identification and characterization of mutations in the upf1 gene that affect nonsense suppression and the formation of the upf protein complex but not mrna turnover. Mol. Cell. Biol. **16**:5491-506: 10.

57. **Zhu, J. Y., P. W. Rice, M. Chamberlain, and C. N. Cole.** 1991. Mapping the transcriptional transactivation function of simian virus 40 large T antigen. J. Virol. **65**:2778-2790.

5. GENERAL DISCUSSION

Differentiation, as well as homeostasis, is the result of a combination of extracellular and cell-autonomous signals which trigger the establishment of a new gene expression program. These processes are mainly controlled by the action of transcription factors via their binding to regulatory sequences located in promoter and enhancer regions. It is thus the modulation of both the expression and transcriptional activities of these transcription factors that enables cells to acquire and maintain a terminal differentiated phenotype.

During recent years, a large body of work has led to a paradigm shift toward the notion that cardiac muscle possesses its own genetic program, different from those of skeletal and smooth muscles, even though all three share a large set of differentiation markers (Olson and Srivastava, 1996; Durocher *et al.*, 1997b). Therefore, in order to understand how cardiogenesis is initiated and how the cardiac phenotype is maintained, it is of paramount importance to unravel novel cardiac-restricted transcriptional pathways. The present work describes such an attempt that led to the identification and the characterization of novel transcriptional pathways involved in the regulation of the cardiac-specific transcription of the ANF gene, an ideal marker of the cardiac phenotype. Interestingly, this work also underscores the importance of combinatorial interactions among transcription factors as a mechanism of generating transcriptional diversity.

5.1. CONVERGENCE AND GENETIC INTERACTIONS OF TRANSCRIPTION PATHWAYS ON THE ANF PROMOTER.

The current paradigm on transcription regulation supports the notion that a promoter, through its topology, its *cis*-element disposition, and its chromatin structure, is able to integrate a complex network of transcriptional inputs, carried by the trans-acting factors, to generate a unique transcriptional output. This paradigm has been clearly illustrated by elegant genetic studies on *eve* stripe gene expression in *Drosophila* [reviewed in]. These studies unraveled distinct modular enhancers that act in an independent fashion. Moreover, the unique stripe expression patterns are obtained via the combinatorial

interactions of positive and negative regulators of transcription. For instance, the *bicoid* transcription factor is required for the expression in all stripes, and the generation of specific stripe expression patterns is dependent on the functional interaction of *bcd* with ubiquitous transcription factors (such as the *Drosophila* STAT homologues in the stripe 3 enhancer) and gap repressor genes such as *giant* and *hunchback* that establish boundaries of expression (Small *et al.*, 1992; Arnosti *et al.*, 1996; Stanojevic *et al.*, 1991; Small *et al.*, 1991; Abdel-Malek *et al.*, 1987). Thus, in this mode of regulation, only a small number of transcription factors can generate complex expression patterns such as those established by the *eve* promoter. The identification of transcriptional pathways and their "linear" integration is not sufficient to understand how patterns of gene expression are established. This situation is particularly relevant for the study of myocardial transcription where only a few truly cardiac-enriched transcription factors (GATA, NK2 and MEF2 proteins) have been identified.

5.1.1. The basal cardiac ANF promoter

An earlier study from our laboratory suggested that the ANF promoter is organized in modular domains (Argentin *et al.*, 1994). This study concluded that the cardiac -700 bp promoter is organized in three domains, the A, B and C domains that each possess different cardiac activities. The A domain (-700/-380) is required for adult, ventricular activity; the B domain (-380/-136) for embryonic activity, whereas the C domain (-136/-78) is required for basal cardiac activity (Argentin *et al.*, 1994; Durocher *et al.*, 1996). Interestingly, the C domain is required for A and B domain activity in the context of the promoter (Durocher and Nemer, 1997). The importance of the C-domain in basal, tissue-specific regulation of ANF expression prompted us to further characterize this region. This domain was already known to harbor a GATA element crucial for ANF promoter activity (F. Charron and M. Nemer, unpublished data) and transient transfection assays in myocytes identified a second element required for cardiac basal activity located between nucleotides -78 and -94 (Fig. 3.1). Analysis of this region identified a region of homology with the *in vitro* defined binding site for Nkx2-5 (Chen and Schwartz, 1995). Gel shifts and cotransfection assays in non-cardiac cells confirmed that the NKE was a *bona fide* target for Nkx2-5 in the

myocardium and unpublished results from Richard Harvey's group confirmed our *in vitro* work by showing that ANF expression in the myocardium is under the control of Nkx2-5. Interestingly, these results establish that the C domain contains targets for both Nkx2-5 and GATA-4 and suggest that both proteins might collaborate. Thus, basal cardiac activity of the ANF promoter was proposed to be mediated by the concerted action of the cardiac NK2 and GATA proteins binding the C-domain. Interestingly, we discovered that indeed both proteins were acting cooperatively on the ANF promoter (Chapter 3). This situation is reminiscent of the IFNβ and TCRα promoters and suggests that the basal cardiac promoter might be organized in a topologically defined structure (Tjian and Maniatis, 1994). This structure, the enhanceosome, is created by the non-random distribution of enhancer-bound factors as well as their physical interactions (Tjian and Maniatis, 1994). These physical contacts create a stereo-specific structure that is disturbed when phasing between *cis*-elements is altered. For the ANF C domain, observations other than the functional and physical interaction between Nkx2-5 and GATA-4 suggest a similar nucleoprotein structure. First, the AT-rich region located between the GATA and NKE motifs has been shown to bind MADS box proteins such as SRF (Sprenkle *et al.*, 1995) which itself can physically interact with both Nkx2-5 (Chen and Schwartz, 1996) and GATA-4 (R.J. Schwartz, personal communication). Secondly, the phasing between the NKE and the GATA element is conserved among species even though spacing has been altered during evolution, suggesting an evolutionary pressure to maintain phasing (Fig. 3.1). However, disruption of phasing between GATA and NKE motifs will be required to undoubtedly confirm that the C-domain is a stereo-specific structure.

If such a modular and perhaps stereo-specific arrangement is also found on other cardiac promoters, this would have important implications for the understanding of how cardiac-specificity is established since it provides clues about why some *cis*-elements such as the M-CAT (which is bound by the ubiquitous TEF-1 protein) behave as a muscle-specific element, especially when closely associated with certain neighboring *cis*-elements (Larkin *et al.*, 1996). This concept could be especially important for the understanding of chamber-specific or LR asymmetric gene expression in the heart. Interestingly, the NKE might in fact be an interesting paradigm for this situation since it behaves on the ANF

promoter as an atrial-specific element in primary cultures of postnatal cardiac myocytes, although its intrinsic activity on a minimal promoter is the same in both atrial and ventricular cells (Fig. 2.2). Interestingly, when parts of the C-domain are deleted, the atrial-specificity of the NKE is lost suggesting the presence of a modifier of NKE activity. The nature of this element is still unknown and finer promoter deletions and mutations should be carried out. Interestingly, a recent report (Biben and Harvey, 1997) suggested that Nkx2-5 might be a conveyor of laterality signals in the developing myocardium. In this study, the asymmetric expression of eHAND in the heart was found to be under the control of Nkx2-5 even though the expression of Nkx2-5 itself was not asymmetric. This scenario is strikingly similar to the chamber-specific activity of the NKE. If the eHAND gene is a direct target for Nkx2-5, it would be interesting to search, on the eHAND promoter, for *cis*-elements similar to those of the proximal ANF promoter. This parallel between the eHAND gene and the ANF gene illustrates that Nkx2-5 activity is most likely modified, either directly or indirectly, by other proteins, in order to generate transcriptional diversity.

5.1.2. CAX- and CATF1- dependent pathways converging on the CARE element.

Whereas the C-domain is required to generate a basal cardiac activity, the A- and B-domains dynamically modulate the activity of the ANF promoter during development. Argentin *et al.* (1993) characterized the ANF CArG element which is required for adult ventricular activity of the promoter. The CArG is presumably bound by a yet to be identified cardiac-restricted protein and the molecular cloning of this activity will soon begin in the laboratory. Interestingly, the fact that the NKE acts as an atrial-specific element, further encouraged us to study the establishment of the atrial-specific activity of the promoter. Previous studies from our laboratory indicated that the atrial and embryonic activity of the promoter seems to rely largely on the CARE element (Ali Ardati, thesis, Université de Montréal 1995). The CARE motif was first characterized as a strong cardiac footprint over the ANF promoter. Transfection studies in primary cardiomyocytes revealed that this motif is strongly active in embryonic and neonatal myocytes whereas its activity is restricted to the atrium starting in the first postnatal week. This profile of activity, which is strikingly similar to that of the ANF gene, prompted us to screen an expression cDNA

library with a radiolabeled CARE element in order to clone CARE-binding proteins. We cloned only one cDNA, CATF1 (Chapter 4). CATF1 protein harbors transcriptional activities that are reminiscent of "classical" transcriptional regulators but also possesses other biochemical activities that might regulate the transcriptional properties of CATF1. CATF1 is now being included in the growing family of helicases that harbor transcriptional activities. This functional family comprises the large T antigen, p70 Ku and members of the SWI2 family (Myers *et al.*, 1981; Giffin *et al.*, 1996; Kingston *et al.*, 1996). However, CATF1 seems to harbor unique properties as it binds preferentially ssDNA, which is not surprising for a protein that harbors helicase activity. More interestingly, CATF1 DNA-binding activity is negatively regulated by ATP. It is tempting to speculate that CATF1 could be sensitive to the concentration of ATP pools since the IC_{50} of the ATP-mediated inhibition is 27 μM, a concentration near physiological concentrations. This would perhaps indicate that CATF1 could be an effector of the hypoxic response. Hypoxia, or the lowering of oxygen concentration, depletes ATP pools, and usually results, in the myocardium, in the upregulation of ANF synthesis and secretion (Baertschi *et al.*, 1988; Galipeau *et al.*, 1988). The proposition that CATF1 might be an hypoxia sensor remains highly speculative but is nonetheless worthwhile investigating given the physiological importance of this pathway in cardiac physiology.

The different mobility patterns of CATF1 and CAX complexes as tested by EMSA as well as the widespread distribution of CATF1 mRNA and protein in adulthood supports the notion that CATF1 is not related to CAX and that it might rather constitute a novel regulatory pathway converging on the CARE element. The clear preference for single-stranded DNA as well as the compilation of CATF1 binding sites characterized so far indicate that CATF1 is likely to bind structural determinants rather than a clear sequence motif such as the GCTGG palindrome for CAX (Mizuta *et al.*, 1993; Shieh *et al.*, 1995). Interestingly, the composite nature of the CARE element is reminiscent of the composite nature of the ventricular-specific HF-1 element on the MLC-2 promoter (Navankasattusas *et al.*, 1992). This composite element is composed of a muscle-restricted component, the MEF2 site, and a ubiquitous component, the YB-1 binding site, and it is the combination of both activities that leads to ventricular-specific activity as tested in transgenic mice (Ross *et*

al., 1996). It would be interesting to investigate if CATF1 participates in the generation of the atrial-specific activity of the CARE element and the molecular cloning of CAX should enable us to study a possible CAX/CATF1 functional interaction.

Nevertheless, a question remains as to why CAX proteins were not isolated in our original screen. The most probable explanation is based on the technical drawbacks of the λgt22 expression cloning system. The southwestern-based cloning system is not compatible with proteins that necessitate heterodimerization or that require post-translational modifications in order to dimerize or bind DNA. As previously mentioned, the binding pattern of the CAX complex is reminiscent of a dimeric or oligomeric complex. Additionally, it is possible that the assembly of a putative CAX homodimer could not occur in the phage expression system as a consequence of physico-chemical differences in bacterial cellular environments. Thus, it is important to establish new strategies in order to clone the CAX protein(s). In order to do so, I would perhaps favor a yeast-based screen such as the one-hybrid screen. The main advantages of this screen are that it provides a eukaryotic, *in vivo* environment for protein folding and that it is a genetic based assay. Moreover, unlike the southwestern-based expression cloning strategy, the DNA-binding reaction is carried out in a cellular environment at near physiological concentrations. Alternatively, a biochemical purification approach using bovine atria as starting material could be used to isolate CAX proteins. This approach requires a tremendous amount of starting material but is generally successful if the properties of the CAX protein are well known and characterized. No matter which approach is taken, the potential recombinant CAX proteins should be tested for binding to both rat and human CARE elements to validate the clones. Either of these two approaches should stand better chances in obtaining the endogenous CAX protein(s) and would lead to a better understanding of the complex regulation of the CARE element. The molecular identification of CAX will advance our knowledge on atrial-specific transcription and will enable the study of a potential interaction between CAX and CATF1 pathways.

5.1.3. Cardiac transcription and promoter structure

Interestingly, the importance of a basal cardiac promoter is also observed for BNP

promoter regulation where the first -114 bp are essential for the high cardiac activity of the promoter. This region contains a closely spaced tandem of GATA elements, an M-CAT and YY1 binding motifs (Grépin *et al.*, 1994). This region is modulated by upstream sequences in non-cardiac cell lines where two silencer elements extinguish the activity of the promoter in non-cardiac cells (S. Bhalla and M. Nemer, manuscript in preparation). This lower complexity of the BNP promoter might explain the far less complex expression pattern of the BNP promoter compared to that of ANF. Additionally, work *in vivo* suggests that, for a third promoter, the MLC3 gene promoter, crucial MEF2 and GATA elements organized in a close arrangement within the first 160 bp are necessary to generate basal cardiac activity. Mutation of either site abolished the expression of a reporter gene in cardiac muscle, as tested by direct injection in myocardium (McGrew *et al.*, 1996). Altogether, these data lead us to suggest that a two-step mechanism, the establishment of a basal cardiac activity that is regulated by upstream and downstream elements, is *a modus operandi* in cardiac gene regulation. This mode of regulation is economical since it requires only a limited set of cardiac-specific regulators but can nevertheless establish complex patterns of expression such as those observed in cardiac muscle. This mechanism of combinatorial interactions leading to tissue-specific transcription is not unique to the heart as it has been described in more detail in *Drosophila* and sea urchin (Davidson, 1991).

Interestingly the modular organization of cardiac promoters could perhaps be extended to another level: the regulation of cardiac transcription by chromosomal domains. The ANF and BNP genes, as well as the α/β-MHC genes, are found closely associated in two different loci suggesting that they might be subject to chromosomal regulation (Lompré *et al.*, 1984; Arden *et al.*, 1995). In fact, this would not be surprising since studies on the regulation of tissue-specific transcription carried out in erythroid cells (Grosveld *et al.*, 1987) and in lymphoid cells (Forrester *et al.*, 1994) clearly implicated chromatin organization in the establishment of precise temporal and spatial gene expression. On the globin or immunoglobulin loci, this type of regulation requires chromatin determinants such as locus control regions or MARs that insulate the locus from its surrounding chromatin. Interestingly, the globin LCRs harbor sites for GATA-1 which might indicate that GATA-4 could play a similar role in the regulation of a putative cardiac locus.

Consistent with this suggestion, a GATA site was found in a developmentally regulated hypersensitive site located 1.9 kb upstream of the α-MHC transcription start site (McKnight *et al.*, 1988). Furthermore, the availability of cosmids encompassing these cardiac gene loci should enable investigators to study the behavior of a locus in transgenic mice, and these studies should lead to the identification of putative cardiac MARs or LCRs, thus bringing novel and significant insights on the regulation of transcription in cardiac muscle.

5.2. THE GATA/NK2 INTERACTION.

5.2.1. Protein-protein interaction as a means to impart functional specificity.

GATA-4 and Nkx2-5 gene knockouts clearly demonstrate that these transcription factors are only partially redundant with their respective homologues expressed in the heart (Lyons *et al.*, 1995; Kuo *et al.*, 1997; Molkentin *et al.*, 1997). Moreover, in the case of Nkx2-5, three genes (ANF, MLC-2 and eHAND) are specific (direct or indirect) targets for Nkx2-5 since their expression is severely reduced in the mutant embryos. These genes established a molecular basis for Nkx2-5-dependent transcription in the developing myocardium. However, it is not clear how this molecular specificity is achieved, especially in the case of GATA proteins where it has been very hard to find any sort of *in vitro* specificity (Visvader *et al.*, 1995; Blobel *et al.*, 1995b; Weiss *et al.*, 1997). Three GATA factors are co-expressed in the precardiac mesoderm and in the developing myocardium; they all bind the same consensus sequence and have similar transcriptional activities. Thus, the unique properties of GATA-4 are most likely to come from its interaction with ancillary proteins. The identification of the GATA-4/Nkx2-5 interaction enabled us to investigate the differences among GATA proteins in a functional assay. As seen in Fig. 3.3, both GATA-4 and -5 but not GATA-6 nor GATA-1 are able to cooperate with Nkx2-5. This specificity is also observed at the level of protein-protein interaction since, unlike GATA-4, GATA-6 is unable to contact Nkx2-5 (Fig. 3.7). Thus, the GATA/Nkx2-5 interaction establishes for the first time a molecular basis for GATA specificity and this interaction will enable us to dissect at the amino acid level how the discrimination between GATA-4

and -6 is achieved. Interestingly, in the erythroid system, GATA-1 has recently been shown to interact with an erythroid cofactor known as FOG (Tsang *et al.*, 1997). However, the FOG/GATA interaction does not discriminate among GATA members and it cannot account for the lack of complete redundancy between GATA-1 and GATA-2. In the light of our results, hematopoietic homeoproteins may be interesting regulators of GATA function. In this regard, HOXB4 overexpression in bone marrow leads to the clonal expansion of the hematopoietic stem cell population (Sauvageau *et al.*, 1995), a phenotype that is reminiscent of the proposed function of GATA-2 in stem cells. Putative HOXB4/GATA-2 or -1 interactions may be worth investigating. Indeed, the fact that GATA-4 may cooperate with HOXB3 (Chapter 3) suggests that some HOX genes could functionally, and physically, interact with the hematopoietic GATA proteins.

In the case of NK2 proteins, the situation is less clear. GATA-4 is able to functionally interact with Nkx2-1 and HOXB3, implying that an Antp-type homeodomain is sufficient to support the interaction. The requirement of the homeodomain for the physical interaction is consistent with this observation (Fig. 3.7). However, the physical interaction between Nkx2-1 or HoxB3 and GATA-4 has not yet been unequivocally demonstrated, although preliminary data suggest that Nkx2-1 interacts with GATA-4 as tested in a pull-down assay. Nevertheless, an appealing possibility is that different subsets of NK2-GATA complexes might mediate different transcriptional responses during early cardiac development. Thus, to test this hypothesis, an extensive survey of interactions between GATA and NK2 proteins should be carried out when the reagents become available. Moreover, the observations that the ANF and eHAND genes manifest regional expression patterns that are (at least partly) requiring Nkx2-5, strongly suggest that Nkx2-5 activity is modified by other pathways [(Chapter 2) (Biben and Harvey, 1997)]. This modification could be posttranslational or, alternatively, the activity of the NKE on the eHAND promoter could be modified by upstream or downstream regulatory elements and their DNA-bound molecules. These latter proteins could be activators or repressors and may directly modify Nkx2-5 activity via protein-protein interactions. In this regard, we showed that Nkx2-5 is a rather modest transcriptional activator on its own (Fig 3.7) and we uncovered the presence of an inducible transactivation domain located near the

homeodomain, implying that the Nkx2-5 molecule is likely to be modulated. Consistent with this, the inducible activation domain of Nkx2-5 is derepressed upon the physical binding of GATA-4 (Fig. 3.5 and 3.8). However, the Nkx2-5/GATA-4 interaction cannot explain the asymmetric expression of eHAND, and other promoter-bound factors might interact with Nkx2-5 to generate this spatial specificity. Nevertheless, collectively, these data underscore the likelihood of protein-protein interactions imparting functional specificity to NK2 proteins.

5.2.2. The GATA-NK2 interaction: effector of the BMP/dpp signal?

GATA-4 and Nkx2-5 are the earliest markers of myocardial cell fate and the identification of the upstream regulators of GATA-4 and Nkx2-5 expression in the precardiac mesoderm myocardium should give important insights on the nature of the inducers of cardiac fate. In *Drosophila*, genetic studies have established that the *tinman* gene is directly downstream of the *decapentaplegic* signal from the ectoderm (Staehling-Hampton *et al.*, 1994; Frasch, 1995). This signal is required for cardiac cell formation since null mutations in *dpp* result in an absence of cardiac differentiation. Furthermore, in chicken, there seems to be a molecular conservation of the dpp/BMP signaling since recombinant BMP-2, -4, and -7 can induce Nkx2-5 in anterior lateral mesoderm. This data also reinforces the hypothesis that Nkx2-5 is an ortholog of tinman and that heart structures in *Drosophila* and vertebrates originate from a common ancestor (Harvey, 1996; Bodmer, 1995; Schultheiss *et al.*, 1997). Consistent with this, in *C. elegans*, pharyngeal muscles have been proposed to compose somewhat of a proto-circulatory system and, indeed, an ortholog of Nkx2-5 (CEH-22) is expressed and required for pharyngeal muscle formation (Okkema and Fire, 1994). Comparative embryology and anatomy have proposed that heart muscle arose from the specialization of pulsatile muscular vessels. In fact, in amphioxus, which is thought to represent the most primitive chordate, the circulatory system is organized in pulsatory vessels (Randall and Davie, 1980). It would be interesting to investigate whether a BMP/*dpp* signal is also involved in pharyngeal muscle development or amphioxus muscular vessels. The recent cloning of a homologue of the BMP-4 receptor in C. elegans indicates that this pathway might be present in worms and that it might be

possible to investigate a possible conservation of the BMP/NK2 pathways in worms (Estevez *et al.*, 1993).

Interestingly, GATA-4 expression is also robustly induced in anterior mesoderm explants treated with BMP-2 and -4. This observation would also imply that GATA-4 is a downstream effector of the BMPs. However, when explants are treated with BMP-7, GATA-4 expression is not observed and this correlates with the lesser potency of BMP-7 to induce the expression of terminal cardiac phenotypic markers such as ventricular MHC (vMHC), and this was also true for the formation of beating explants (Schultheiss *et al.*, 1997). These observations are consistent with the gain-of-function studies which demonstrated that neither GATA-4 nor Nkx2-5 could alone initiate cardiogenesis. However, both proteins could potentiate cardiogenesis in committed cells (Grépin *et al.*, 1997; Chen and Fishman, 1996; Cleaver *et al.*, 1996). Together, these data suggest that both GATA-4 and Nkx2-5 are downstream effectors of BMP. Interestingly, GATA-2 has also been shown to be downstream of BMP-2 in *Xenopus*, and the dorsalizing agent chordin can inhibit BMP-2 induction in animal cap explants (Walmsley *et al.*, 1994) which suggests that GATA proteins are effectors of ventralizing signals. Thus, the physical and functional interaction between Nkx2-5 and GATA-4 may provide an explanation for the requirement of both proteins in BMP-mediated cardiogenesis. This functional interaction is able to potentiate the transcriptional activities of both proteins, especially at low concentrations of GATA-4 and Nkx2-5 (Fig. 3.1), which probably occurs in the early moments of cardiac fate induction.

In addition to the possible conservation of *dpp*/BMP signaling in heart development, there are indications that there is also conservation of the GATA/NK2 interaction in *Drosophila* and in *C. elegans*. The characterization of the cardiac enhancer in the D-Mef2 regulatory region established that the *tinman* protein is a direct regulator of D-Mef2 expression. Furthermore, the ectopic expression of *tinman* in fly embryos activates D-Mef2 expression in only a subset of cells, implying that tinman requires a cofactor that could perhaps be a GATA protein since closely GATA motifs are closely linked to the two NKE motifs present on the D-MEF2 cardiac enhancer (Gajewski *et al.*, 1997). On the other hand, in *C. elegans*, the Myo-2 gene promoter is active in pharyngeal muscle and is a direct

target for the NK2 CEH-22 protein. Mutational analyses of the CEH-22 responsive enhancer unraveled also the importance of a GATA motif located just 3' of the NKE. However, the presence of GATA factors in pharyngeal muscle has not yet been detected but a synthetic lethal screen using the GATA-4 cDNA under the control of the Myo-2 promoter could be used to identify GATA-4 like activities in the worm. Nevertheless, these observations suggest that a GATA/NK2 interaction might be a conserved downstream effector of the BMP/*dpp* signal in cardiogenesis.

In mammalian cells, an experimental approach based on the P19 system could be used to further investigate the possibility that the Nkx2-5/GATA-4 complex is required for BMP-dependent cardiogenesis. The main advantages of the P19 system reside in its versatility and the possibility of performing genetic manipulations (which is not the case in the avian system). However, the importance of BMP signaling in P19 cardiac differentiation has not been described as yet. One could first look at the effect of the noggin protein, an antagonist of BMP-2 and BMP-4, when added to the medium of DMSO-treated P19 aggregates. If the addition of the noggin protein results in an inhibition of cardiac differentiation, this would imply that that the BMP pathway is also involved in P19 cardiogenesis. The noggin protein could then be used as a tool to decipher downstream effectors of BMP signaling by carrying out epistatic studies. In such studies, P19 cells stably overexpressing Nkx2-5, GATA-4 or both would be aggregated in the presence of both noggin and DMSO. In any case, if one of these stable cell lines could differentiate into cardiomyocytes in the presence of noggin, this would indicate that GATA-4 or Nkx2-5 are *bona fide* effectors of BMP signaling during cardiac differentiation. However, if GATA-4 and Nkx2-5 have to be expressed together in order to compensate for the loss of BMP signaling, this would strongly argue for the importance of the GATA-4/Nkx2-5 interaction in cardiac mesoderm differentiation and would provide a molecular basis to study the developmental importance of this interaction.

5.2.3. Evolutionary conservation of the Zn-finger/homeodomain interaction.

In addition to our suggestion that the GATA/NK2 interaction might be conserved among species, the recent characterization of a physical and functional interaction between

the Zn-finger protein Ftz-F1 and the homeodomain protein Ftz in *Drosophila* suggests that the Zn-finger protein/homeoprotein could be ancient and relatively widespread. Genetic and physical data demonstrated that the nuclear receptor Ftz-F1, a Zn-finger protein, is a cofactor for Ftz (Yu *et al.*, 1997; Guichet *et al.*, 1997). Mutations in Ftz-F1 gene lead to defects similar to the pair-rule phenotype observed in Ftz mutant flies, implying that these two genes are in the same epistatic group (Guichet *et al.*, 1997). Ftz-F1 was shown to physically bind Ftz in two-hybrid assays (Yu *et al.*, 1997) as well as in co-immunoprecipitations and seemed to be required for efficient binding of Ftz on DNA (Yu *et al.*, 1997; Guichet *et al.*, 1997). Interestingly, the deletion of the Ftz-F1/Ftz interaction domain on Ftz is sufficient to phenocopy the loss of Ftz-F1 (Guichet *et al.*, 1997). Thus, Ftz-F1 is a *bona fide* cofactor of the Ftz homeoprotein and collectively, these data demonstrate that the use of physical interaction between homeoproteins and Zn-finger proteins is a common and evolutionarily conserved mean to modulate their transcriptional properties.

Interestingly, a novel class of homeoproteins further supports the notion that the Zn-finger/homeoprotein interaction is ancient and widespread. The atypical class of ZFH homeobox genes encodes transcriptional regulators that harbor both Zn-finger and homeodomain motifs (Lai *et al.*, 1991). These transcriptional regulators can be seen as an extreme representation of the Zn-finger/homeodomain interaction. The number, the nature and the position of the Zn-finger and homeodomain motifs vary enormously among members of the ZFH class and can range from 2 to 16 Zn-fingers and up to 4 homeodomains (Postigo and Dean, 1997). This observation suggests that this class of proteins arose from multiple, stabilized, gene fusion events which occurred independently during evolution. This situation resembles that of Pax genes which can harbor two DNA-binding domains, the paired domain and the homeodomain, acquired by a gene fusion event (Wilson and Desplan, 1995). Some members of the Pax family, such as Pax-6, harbor both domains whereas certain others such as Pax-5 and Mix.1 possess only a paired or homeodomain motif respectively [reviewed in (Wilson and Desplan, 1995)]. DNA-binding studies established that the paired domain and the homeodomain bind DNA independently and cooperatively and this cooperation greatly extends the specificity of the target sequence

(Jun and Desplan, 1996; Wilson and Desplan, 1995). Thus, both the gene fusion events strongly argue that cooperative interaction between DNA-binding proteins is widely used for imparting further functional specificity.

5.3. CATF1

5.3.1. Potential non-transcriptional roles of CATF1

The present work characterized the helicase and the transcriptional activities of CATF1. The novel properties of this transcription factor are discussed in greater details in Chapter 4 and in section 5.1.2. Intriguingly, the presence of putative CATF1 orthologues in the eukaryal and archeal kingdoms strongly suggests a conserved role for the CATF1 helicase activity. This activity is probably not associated with transcription since CATF1 proteins in lower organisms do not harbor the PQ motif required for the transcriptional activation properties of mammalian CATF1. However, as stated in a recent review on archeal genomics, the great majority, if not all, of genes conserved between archea and eukaryotes have a role in regulating some aspects of gene expression or genome replication and repair (Olsen and Woese, 1997). Thus, it is tempting to speculate on the possible functions of CATF1 in lower organisms. Roles in the regulation of transcription, RNA metabolism or even DNA replication would have to be envisaged.

A role for CATF1 in RNA metabolism is likely since two out of three CATF1 relatives are involved in this cellular process. SEN1 is an mRNA or tRNA splicing regulator (Steinmetz and Brow, 1996; DeMarini *et al.*, 1992) and UPF1 is involved in the nonsense-mediated mRNA decay pathway in yeast and probably in vertebrates (Ruiz-Echevarria *et al.*, 1996; Weng *et al.*, 1996a). Interestingly, even Dna2p, (a replicative helicase) could be proposed to harbor RNA:DNA helicase activity since it seems to be involved in the metabolism of Okazaki fragments which are primed with short stretches of RNA (Budd *et al.*, 1995; Budd and Campbell, 1997). Moreover, most of the helicase superfamily I members are proteins involved in RNA metabolism (Gorbalenya and Koonin, 1993). We have preliminary data showing that CATF1 can indeed bind RNA although further studies are required to show that CATF1 is also an RNA helicase. Interestingly,

Upf1p is both a DNA and RNA helicase and a DNA and RNA-binding protein as well (Weng *et al.*, 1996b). In addition, we also have preliminary data showing that CATF1 is also present in the cytoplasm as well as in the nucleus which suggests that the protein might have a role in both compartments. However, further biochemical studies will be necessary to unequivocally assess a role for CATF1 in RNA metabolism. Alternatively, another function that is not exclusive with the latter could be proposed for CATF1, based on a recent review. This review underscored the role of some members of helicase SF1 as cofactors of nucleases (Anderson and Parker, 1996). Consistent with this observation, many members of the UPF1 family seem to interact genetically and physically with nucleases for which they act as ancillary proteins to unwind the target nucleic acid. Upf1p is thought to accelerate mRNA decay through its potential association with the xrn1p nuclease whereas Sen1p might regulate tRNA splicing via its interaction with the Sen2p/Sen34p/Sen15p/Sen54p endonuclease complex (Trotta *et al.*, 1997). Interestingly, Dna2p has been recently demonstrated to be associated with the FEN1 nuclease, a nuclease involved in the processing of Okazaki fragments (Budd and Campbell, 1997). The possibility that CATF1 would be associated with a nuclease when performing a potential non-transcriptional function could be worthwhile investigating, but remains at this point highly speculative.

In order to characterize the evolutionary conserved functions of CATF1, if any, a genetic approach would be favored. In yeast, the ORF ykl017c harbors a high degree of similarity with CATF1. We, and others, have already disrupted ykl017c alleles in order to find a cellular process in which a functional CATF1 is essential in budding yeast. Unfortunately, the ykl017c gene is not essential for viability. However, it would be possible to perform a synthetic lethal screen in yeast to uncover genes that are linked genetically to CATF1. The synthetic lethal screen would be performed using a yeast plasmid carrying a mammalian CATF1 cDNA under the control of a yeast promoter. The plasmid would be transformed in a Δykl017c strain which would then be chemically mutagenized. Mutants would then be replica plated on plates containing media that would either keep, or chase, the plasmid encoding the mammalian CATF1 cDNA. Mutants that are only growing when the CATF1-encoding plasmid is present are synthetic lethal

mutants. Using this approach, it would be possible to find which cellular processes require a functional CATF1 protein and, since we would be using a mammalian CATF1, this screen would also imply that this function is evolutionary conserved and would enable testing of this function in higher eukaryotes.

If this approach is successful, the accomplishment of structure-function studies in yeast using mammalian, yeast or archeal CATF1 proteins would be amenable. The assay would be based on the suppression of the synthetic lethal phenotype. By these means, it would be possible to dissect the role of the non-conserved C-terminal tail of the mammalian CATF1, as well as the role of the ATPase and helicase activities. Moreover, this work would have a direct impact on the study of cellular functions in archea if the archeal CATF1 protein is functional in yeast. In the long term, this work might enable us to identify new gene regulatory networks that are conserved both in the eukaryal and archeal kingdoms.

5.3.3. CATF1 and embryonic development

As seen in Fig 4.2, the expression of CATF1 is developmentally regulated and in embryos and in neonates, the expression of CATF1 is largely restricted to cardiac tissue. In order to investigate the role of CATF1 protein in cardiac development, we initiated a knock-down experiment in P19 embryocarcinoma cells (this work is carried out by Dr. Ariel Arias in the laboratory). P19 cells are pluripotent and can be induced to differentiate in a number of cell lineages upon aggregation (McBurney *et al.*, 1982; Rudnicki and McBurney, 1987). During P19 *in vitro* differentiation, CATF1 expression is normally specifically restricted to the cardiac and skeletal muscle differentiation pathways; implying that CATF1 might have a role in lateral and axial mesoderm differentiation. Disruption of CATF1 expression by an antisense strategy has dramatic effects on both cardiac and skeletal muscle differentiation, which are characterized by a dramatic reduction of fully differentiated myocytes and by the induction of ectopic neuronal differentiation. Further, molecular analysis of the markers affected by the disruption of CATF1 expression revealed that the defect seems to lie at the mesodermal induction stage, a result that is consistent with the upregulation of neuronal (ectodermal) differentiation often seen in mesodermal

induction defects (Hemmati-Brivanlou and Melton, 1997). Nevertheless, gene inactivation of CATF1 alleles in mice will have to be performed in order to unequivocally assess the role of CATF1 in mesodermal and/or cardiac differentiation. Furthermore, these studies could be coupled to the Cre-Lox system in order to generate a cardiac-specific knockout of the CATF1 gene. Additionally, the use of the P19 system would also enable us to genetically dissect the functions of the CATF1 protein. Using retroviruses expressing various CATF1 proteins, we could assess if yeast or archeal CATF1 proteins can perform the developmental functions of CATF1 in this system. Thus, it would be in theory possible to dissect the cellular processes that require a transcriptionally active CATF1 protein from those which only require the helicase activity.

Collectively the studies carried out in P19 cells, and the observation that CATF1 is the only UPF1 family member that is conserved in both archea and eukaryal kingdoms, strongly suggest that it may play a critical role in the regulation of gene expression. It will thus be important to pursue further studies of CATF1 function both at the genetic and molecular levels.

5.4. FUTURE PROSPECTS.

The characterization of two additional transcriptional pathways acting to establish the cardiac-specific activity of the ANF promoter (the NKE and the CATF1 pathway), as well as the unraveling of the molecular interaction between NK2 and GATA proteins, leads to a clearer picture of the molecular networking involved in cardiac-specific transcription of the ANF promoter. The ANF paradigm suggests that chamber and temporal specificity of cardiac promoters might be dictated by combinatorial interactions between (as well as within) modular regulatory domains. As a direct consequence of this work, future work should be directed at the study of the molecular architecture of as many cardiac promoters as possible. These promoters should not only be those of contractile protein genes but also those of other classes of genes involved in cardiac physiology such as ion channel genes and cardiac transcription factor genes. In the long term, this work should identify common sets of molecular interactions leading to discrete patterns of gene expression and might perhaps establish a molecular code for regional and temporal cardiac transcriptional

activities.

Additionally, for the next decade, the major focus of investigators in the cardiac field should be directed toward the genetic and molecular decipherment of the events occurring during early cardiogenesis. Events such as the initiation of the cardiogenic field from lateral plate mesoderm, terminal differentiation of cardioblasts, as well as the processes controlling cardiac morphogenesis, septation and trabeculation, should be particularly emphasized. The zebrafish genetic model will be an extraordinary tool to identify new gene networks involved in cardiac development and the possibility of manipulating the mouse genome will enable the targeted disruption of candidate cardiac developmental genes. Furthermore, the identification of the rapidly growing number of cardiac transcriptional regulators will allow the study of molecular interactions among them and will enable us to understand how these interactions are imparting further specificity and diversity in their transcriptional activities

In conclusion, the study of cardiac transcription is a useful means to understand how cardiogenesis and cardiac homeostasis are controlled. This, in turn, will undoubtedly help us understand and perhaps manipulate the profound genetic and cellular changes that are associated with cardiac diseases such as those involving myocyte loss and cardiac hypertrophy.

5. REFERENCES

Abdel-Malek,Z.A., Swope,V.B., Amornsiripanitch,N. and Nordlund,J.J. (1987) In vitro modulation of proliferation and melanization of S91 melanoma cells by prostaglandins. *Cancer Res.* **47**, 3141-3146.

Adachi,Y., Kas,E. and Laemmli,U.K. (1989) Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. *EMBO J.* **8**, 3997-4006.

Adams,C.C. and Workman,J.L. (1993) Nucleosome displacement in transcription. *Cell*, **72**, 305-8: 26.

Aird,W.C., Parvin,J.D., Sharp,P.A. and Rosenberg,R.D. (1994) The interaction of GATA-binding proteins and basal transcription factors with GATA box-containing core promoters. a model of tissue-specific gene expression. *J. Biol. Chem.* **269**, 883-889.

Allison,L.A., Wong,J.K., Fitzpatrick,V.D., Moyle,M. and Ingles,C.J. (1988) The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. *Mol. Cell. Biol.* **8**, 321-329.

Anderson,J.S. and Parker,R. (1996) RNA turnover: the helicase story unwinds. *Curr. Biol.* **6**, 780-782.

Arceci,R.J., King,A.A., Simon,M.C., Orkin,S.H. and Wilson,D.B. (1993) Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol. Cell. Biol.* **13**, 2235-2246.

Archambault,J., Lacroute,F., Ruet,A. and Friesen,J.D. (1992) Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. *Mol. Cell. Biol.* **12**, 4142-4152.

Arden,K.C., Viars,C.S., Weiss,S., Argentin,S. and Nemer,M. (1995) Localization of the human B-type natriuretic peptide precursor (NPPB) gene to chromosome 1p36. *Genomics*, **26**, 385-389.

Argentin,S., Ardati,A., Tremblay,S., Lihrmann,I., Robitaille,L., Drouin,J. and Nemer,M. (1994) Developmental stage-specific regulation of atrial natriuretic factor gene transcription in cardiac cells. *Mol. Cell. Biol.* **14**, 777-790.

Arnosti,D.N., Barolo,S., Levine,M. and Small,S. (1996) The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development*, **122**, 205-14:28.

Aso,T., Conaway,J.W. and Conaway,R.C. (1995a) The RNA polymerase II elongation complex. *FASEB J.* **9**, 1419-1428.

Aso,T., Lane,W.S., Conaway,J.W. and Conaway,R.C. (1995b) Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science*, **269**, 1439-1443.

Aso,T., Shilatifard,A., Conaway,J.W. and Conaway,R.C. (1996) Transcription syndromes and the role of RNA polymerase II general transcription factors in human disease. *J. Clin. Invest.* **97**, 1561-1569.

Azpiazu,N., Lawrence,P.A., Vincent,J.P. and Frasch,M. (1996) Segmentation and specification of the Drosophila mesoderm. *Genes Dev.* **10**, 3183-3194.

Azpiazu,N. and Frasch,M. (1993) Tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of drosophila. *Genes Dev.* **7**, 1325-1340.

Baertschi,A.J., Adams,J.M. and Sullivan,M.P. (1988) Acute hypoxemia stimulates atrial natriuretic factor secretion in vivo. *Am. J. Physiol.* **255**, H295-H300.

Baeuerle,P.A. and Baltimore,D. (1988) IkappaB: a specific inhibitor of the NF-kappaB transcription factor. *Science*, **242**, 540-546.

Bannister,A.J. and Kouzarides,T. (1995) CBP-induced stimulation of c-Fos activity is abrogated by E1A. *EMBO J.* **14**, 4758-4762.

Bannister,A.J. and Kouzarides,T. (1996) The CBP co-activator is a histone acetyltransferase. *Nature*, **384**, 641-643.

Barberis,A., Pearlberg,J., Simkovich,N., Farrell,S., Reinagel,P., Bamdad,C., Sigal,G. and Ptashne,M. (1995) Contact with a component of the polymerase ii holoenzyme suffices for gene activation. *Cell*, **81**, 359-68: 15.

Barlow,P., Owen,D.A. and Graham,C. (1972) DNA synthesis in the preimplantation mouse embryo. *J. Embryol. Exp. Morphol.* **27**, 431-445.

Baumann,P., Qureshi,S.A. and Jackson,S.P. (1995) Transcription: new insights from studies on archaea. *Trends Genet.* **11**, 279-83: 10.

Baylies,M.K., Martinez Arias,A. and Bate,M. (1995) wingless is required for the formation of a subset of muscle founder cells during Drosophila embryogenesis. *Development*, **121**, 3829-3837.

Beaulieu,M., Goldman,M.E., Miyazaki,K., Frey,E.A., Eskay,R.L., Kebabian,J.W. and Côté,T.E. (1984) Bromocriptine-induced changes in the biochemistry, physiology, and histology of the intermediate lobe of the rat pituitary gland. *Endocrinology*, **114**, 1871-1884.

Beiman,M., Shilo,B.Z. and Volk,T. (1996) Heartless, a Drosophila FGF receptor homolog, is essential for cell migration and establishment of several mesodermal lineages. *Genes Dev.* **10**, 2993-3002.

Benezra,R., Davis,R.L., Lockshon,D., Turner,D.L. and Weintraub,H. (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell*, **61**, 49-59.

Bestor,T.H. and Verdine,G.L. (1994) Dna methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380-9: 19.

Biben,C. and Harvey,R.P. (1997) Homeodomain factor Nkx2-5 controls left/right asymmetric expression of bHLH gene eHand during murine heart development. *Genes Dev.* **11**, 1357-1369.

Bielinska,M. and Wilson,D.B. (1995) Regulation of J6 gene expression by transcription factor GATA-4. *Biochem. J.* **307**, 183-189.

Bienroth,S., Wahle,E., Suter-Crazzolara,C. and Keller,W. (1991) Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J. Biol. Chem.* **266**, 19768-19776.

Bird,A.P. (1993) Functions for dna methylation in vertebrates. *CSH Symp. Quant. Biol.* **58**, 281-5: 22.

Bisaha,J.G. and Bader,D. (1991) Identification and characterization of a ventricular-specific avian myosin heavy chain, VMHC1: expression in differentiating cardiac and skeletal muscle. *Dev. Biol.* **148**, 355-364.

Blau,J., Xiao,H., McCracken,S., O'Hare,P., Greenblatt,J. and Bentley,D. (1996) Three functional classes of transcriptional activation domain. *Mol. Cell. Biol.* **16**, 2044-55: 7.

Blobel,G.A., Sieff,C.A. and Orkin,S.H. (1995a) Ligand-dependent repression of the erythroid transcription factor gata-1 by the estrogen receptor. *Mol. Cell. Biol.* **15**, 3147-53: 4.

Blobel,G.A., Simon,M.C. and Orkin,S.H. (1995b) Rescue of GATA-1-deficient embryonic stem cells by heterologous GATA-binding proteins. *Mol. Cell. Biol.* **15**, 626-633.

Bode,J., Schlake,T., Rios-Ramirez,M., Mielke,C., Stengert,M., Kay,V. and Klehr-Wirth,D. (1995) Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int. Rev. Cytol.* **162A**, 389-454.

Bodmer,R. (1993) The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. *Development*, **118**, 719-729.

Bodmer,R. (1995) Heart development in *drosophila* and its relationship to vertebrates. *Trends Cardiovasc. Med.* **5**, 21-28.

Bour,B.A., O'Brien,M.A., Lockwood,W.L., Goldstein,E.S., Bodmer,R., Taghert,P.H., Abmayr,S.M. and Nguyen,H.T. (1995) Drosophila MEF2, a transcription factor that is essential for myogenesis. *Genes Dev.* **9**, 730-741.

Boxer,L.M., Prywes,R., Roeder,R.G. and Kedes,L. (1989) The sarcomeric actin CArG-binding factor is indistinguishable from the c-fos serum response factor. *Mol. Cell. Biol.* **9**, 515-522.

Boyes,J. and Bird,A. (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell*, **64**, 1123-1134.

Boyes,J. and Bird,A. (1992) Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* **11**, 327-333.

Braun,T., Rudnicki,M.A., Arnold,H.H. and Jaenisch,R. (1992) Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. *Cell*, **71**, 369-382.

Breitbart,R.E., Liang,C.S., Smoot,L.B., Laheru,D.A., Mahdavi,V. and Nadal-Ginard,B. (1993) A fourth human MEF2 transcription factor, hMEF2D, is an early marker of the myogenic lineage. *Development*, **118**, 1095-1106.

Briegel,K., Limi,K.C., Plank,C., Beug,H., Engel,J.D. and Zenke,M. (1993) Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. *Genes Dev.* **7**, 1097-1109.

Brody,E. and Abelson,J. (1985) The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science*, **228**, 963-967.

Brown,C.B., Boyer,A.S., Runyan,R.B. and Barnett,J.V. (1996) Antibodies to the Type II TGFbeta receptor block cell activation and migration during atrioventricular cushion transformation in the heart. *Dev. Biol.* **174**, 248-257.

Brueckner,M., D'eustachio,P. and Horwich,A.L. (1989) Linkage mapping of a mouse gene, iv, that controls left-right asymmetry of the heart and viscera. *Proc. Natl. Acad. Sci. USA*, **86**, 5035-5038.

Buchberger,A., Pabst,O., Brand,T., Seidl,K. and Arnold,H.H. (1996) Chick Nkx-2.3 represents a novel family member of vertebrate homologues to the drosophila homeobox gene tinman - differential expression of CNkx-2.3 and CNkx-2.5 during heart and gut development. *Mech. Dev.* **56**, 151-163.

Budd,M.E., Choe,W.C. and Campbell,J.L. (1995) DNA2 encodes a DNA helicase essential for replication of eukaryotic chromosomes. *J. Biol. Chem.* **270**, 26766-26769.

Budd,M.E. and Campbell,J.L. (1997) A yeast replicative helicase, Dna2 helicase, interacts with yeast FEN-1 nuclease in carrying out its essential function. *Mol. Cell. Biol.* **17**, 2136-2142.

Buratowski,S., Hahn,S., Sharp,P.A. and Guarente,L. (1988) Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature*, **334**, 37-42.

Buratowski,S., Hahn,S., Guarente,L. and Sharp,P.A. (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, **56**, 549-561.

Burke,T.W. and Kadonaga,J.T. (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.* **10**, 711-724.

Cadena,D.L. and Dahmus,M.E. (1987) Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *J. Biol. Chem.* **262**, 12468-12474.

Cairns,B.R., Lorch,Y., Li,Y., Zhang,M., Lacomis,L., Erdjument-Bromage,H., Tempst,P., Du,J., Laurent,B. and Kornberg,R.D. (1996) RSC, an essential, abundant chromatin-remodeling complex. *Cell*, **87**, 1249-1260.

Cal,H. and Levine,M. (1995) Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. *Nature*, **376**, 533-536.

Carlson,M. and Laurent,B.C. (1994) The SNF/SWI family of global transcriptional activators. *Curr. Opin. Cell Biol.* **6**, 396-402.

Catron,K.M., Zhang,H., Marshall,S.C., Inostroza,J.A., Wilson,J.M. and Abate,C. (1995) Transcriptional repression by Msx-1 does not require homeodomain DNA-binding sites. *Mol. Cell. Biol.* **15**, 861-871.

Cavallini,B., Huet,J., Plassat,J.-L., Sentenac,A., Egly,J.-M. and Chambon,P. (1988) A yeast activity can substitute for the HeLa cell TATA box factor. *Nature*, **334**, 77-80.

Chabot,B. (1996) Directing alternative splicing: cast and scenarios. *Trends Genet.* **12**, 472-8: 4.

Chambers,A.E., Logan,M., Kotecha,S., Towers,N., Sparrow,D. and Mohun,T.J. (1994) The RSRF/MEF2 protein SL1 regulates cardiac muscle-specific transcription of a myosin light-chain gene in Xenopus embryos. *Genes Dev.* **8**, 1324-1334.

Chan,S.K., Jaffe,L., Capovilla,M., Botas,J. and Mann,R.S. (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, **78**, 603-615.

Chan,S.K., Popperl,H., Krumlauf,R. and Mann,R.S. (1996) An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J.* **15**, 2476-2487.

Charron,J., Malynn,B.A., Fisher,P., Stewart,V., Jeannotte,L., Goff,S.P., Robertson,E.J. and Alt,F.W. (1992) Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. *Genes Dev.* **6**, 2248-2257.

Chen,C.Y., Croissant,J., Majesky,M., Topouzis,S., McQuinn,T., Frankovsky,M.J. and Schwartz,R.J. (1996) Activation of the cardiac alpha-actin promoter depends upon serum response factor, Tinman homologue, Nkx-2.5, and intact serum response elements. *Dev. Genet.* **19**, 119-130.

Chen,C.Y. and Schwartz,R.J. (1995) Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. *J. Biol. Chem.* **270**, 15628-15633.

Chen,C.Y. and Schwartz,R.J. (1996) Recruitment of the tinman homolog Nkx-2.5 by serum response factor activates cardiac α-actin gene transcription. *Mol. Biol. Cell*, **16**, 6372-6384.

Chen,J.L., Attardi,L.D., Verrijzer,C.P., Yokomori,K. and Tjian,R. (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, **79**, 93-105.

Chen,J.N., Haffter,P., Odenthal,J., Vogelsang,E., Brand,M., van Eeden,F.J., Furutani-Seiki,M., Granato,M., Hammerschmidt,M., Heisenberg,C.P., Jiang,Y.J., Kane,D.A., Kelsh,R.N., Mullins,M.C. and Nusslein-Volhard,C. (1996) Mutations affecting the cardiovascular system and other internal organs in zebrafish. *Development*, **123**, 293-302.

Chen,J.N. and Fishman,M.C. (1996) Zebrafish *tinman* homolog demarcates the heart field and initiates myocardial differentiation. *Development*, **122**, 3809-3816.

Chen,Y. and Solursh,M. (1995) Mirror-image duplication of the primary axis and heart in Xenopus embryos by the overexpression of Msx-1 gene. *J. Exp. Zool.* **273**, 170-174.

Chen,Z., Friedrich,G.A. and Soriano,P. (1994) Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.* **8**, 2293-2301.

Chisaka,O. and Capecchi,M.R. (1991) Regionally restricted developmental defects resulting from the targetted disruption of the mouse homeobox gene hox 1.5. *Nature*, **350**, 473-479.

Choy,B., Roberts,S.G., Griffin,L.A. and Green,M.R. (1993) How eukaryotic transcription activators increase assembly of preinitiation complexes. *CSH Symp. Quant. Biol.* **58**, 199-203.

Choy,B. and Green,M.R. (1993) Eukaryotic activators function during multiple steps of preinitiation complex assembly. *Nature*, **366**, 531-536.

Clark-Adams,C.D., Norris,D., Osley,M.A., Fassler,J.S. and Winston,F. (1988) Changes in histone gene dosage alter transcription in yeast. *Genes Dev.* **2**, 150-159.

Claycomb,W.C. (1983) Cardiac muscle cell proliferation and cell differentiation in vivo and in vitro. [Review]. *Adv. Exp. Med. Biol.* **161**, 249-265.

Claycomb,W.C. (1992) Control of cardiac muscle cell division. *Trends Cardiovasc. Med.* **2**, 231-236.

Cleaver,O.B., Patterson,K.D. and Krieg,P.A. (1996) Overexpression of the tinman-related genes XNKX-2.5 and XNKX-2.3 in xenopus embryos results in myocardial hyperplasia. *Development*, **122**, 3549-3556.

Collignon,J., Varlet,I. and Robertson,E.J. (1996) Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature*, **381**, 155-158.

Cooper,T.A. and Ordahl,C.P. (1985) A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. *J. Biol. Chem.* **260**, 11140-11148.

Cooper,T.A. and Ordahl,C.P. (1989) Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. *Nucleic Acids Res.* **17**, 7905-7921.

Corsetti,M.T., Briata,P., Sanseverino,L., Daga,A., Airoldi,I., Simeone,A., Palmisano,G., Angelini,C., Boncinelli,E. and Corte,G. (1992) Differential DNA binding properties of three human homeodomain proteins. *Nucleic Acids Res.* **20**, 4465-4472.

Cote,J., Quinn,J., Workman,J.L. and Peterson,C.L. (1994) Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science*, **265**, 53-60.

Cross,J.C., Flannery,M.L., Blanar,M.A., Steingrimsson,E., Jenkins,N.A., Copeland,N.G., Rutter,W.J. and Werb,Z. (1995) Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. *Development*, **121**, 2513-2523.

Croston,G.E. and Kadonaga,J.T. (1993) Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Curr. Opin. Cell Biol.* **5**, 417-423.

Cserjesi,P., Brown,D., Lyons,G.E. and Olson,E.N. (1995) Expression of the novel basic helix-loop-helix gene eHAND in neural crest derivatives and extraembryonic membranes during mouse development. *Dev. Biol.* **170**, 664-678.

Cui,Y., Hagan,K.W., Zhang,S. and Peltz,S.W. (1995) Identification and characterization of genes that are required for the accelerated degradation of mrnas containing a premature translational termination codon. *Genes Dev.* **9**, 423-36: 18.

Czaplinski,K., Weng,Y., Hagan,K.W. and Peltz,S.W. (1995) Purification and characterization of the upf1 protein: a factor involved in translation and mrna degradation. *RNA*, **1**, 610-23: 15.

Dagnino,L., Drouin,J. and Nemer,M. (1991) Differential expression of natriuretic peptide genes in cardiac and extracardiac tissues. *Mol. Endocrinol.* **5**, 1292-1300.

Damante,G., Pellizzari,L., Esposito,G., Fogolari,F., Viglino,P., Fabbro,D., Tell,G., Formisano,S. and Dilauro,R. (1996) A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J.* **15**, 4992-5000.

Davidson,E.H. (1991) Spatial mechanisms of gene regulation in metazoan embryos. [Review]. *Development*, **113**, 1-26.

de Jong,F., Opthof,T., Wilde,A.A., Janse,M.J., Charles,R., Lamers,W.H. and Moorman,A.F. (1992) Persisting zones of slow impulse conduction in developing chicken hearts. *Circ. Res.* **71**, 240-250.

Decker,C.J. and Parker,R. (1995) Diversity of cytoplasmic functions for the 3' untranslated region of eukaryotic transcripts. *Curr. Opin. Cell Biol.* **7**, 386-392.

Delorme,B., Dahl,E., Jarry-Guichard,T., Marics,I., Briand,J.P., Willecke,K., Gros,D. and Theveniau-Ruissy,M. (1995) Developmental regulation of connexin 40 gene expression in mouse heart correlates with the differentiation of the conduction system. *Dev. Dynamics*, **204**, 358-371.

DeMarini,D.J., Winey,M., Ursic,D., Webb,F. and Culbertson,M.R. (1992) SEN1, a positive effector of tRNA-splicing endonuclease in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **12**, 2154-2164.

Dessain,S., Gross,C.T., Kuziora,M.A. and McGinnis,W. (1992) Antp-type homeodomains have distinct DNA binding specificities that correlate with their different regulatory functions in embryos. *EMBO J.* **11**, 991-1002.

Dikstein,R., Ruppert,S. and Tjian,R. (1996) TAFII250 is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. *Cell*, **84**, 781-790.

Donovan,M.J., Hahn,R., Tessarollo,L. and Hempstead,B.L. (1996) Identification of an essential nonneuronal function of neurotrophin 3 in mammalian cardiac development. *Nature Genetics*, **14**, 210-213.

Driever,W., Solnica-Krezel,L., Schier,A.F., Neuhauss,S.C., Malicki,J., Stemple,D.L., Stainier,D.Y., Zwartkruis,F., Abdelilah,S., Rangini,Z., Belak,J. and Boggs,C. (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development*, **123**, 37-46.

Duan,D.R., Pause,A., Burgess,W.H., Aso,T., Chen,D.Y., Garrett,K.P., Conaway,R.C., Conaway,J.W., Linehan,W.M. and Klausner,R.D. (1995) Inhibition of transcription elongation by the VHL tumor suppressor protein. *Science*, **269**, 1402-1406.

Durocher,D., Chen,C.Y., Ardati,A., Schwartz,R.J. and Nemer,M. (1996) The ANF promoter is a downstream target for Nkx-2.5 in the myocardium. *Mol. Biol. Cell*, **16**, 4648-4655.

Durocher,D., Charron,F., Warren,R., Schwartz,R.J. and Nemer,M. (1997a) The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J.* **16**, 5687-5696.

Durocher,D., Grépin,C. and Nemer,M. (1997b) Regulation of gene expression in the endocrine heart. *Rec. Prog. Horm. Res.* **In press,**

Durocher,D. and Nemer,M. (1997) Combinatorial interactions regulating cardiac transcription. *Dev. Genet.* **submitted,**

Dynlacht,B.D., Hoey,T. and Tjian,R. (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell*, **66**, 563-576.

Eden,S. and Cedar,H. (1994) Role of DNA methylation in the regulation of transcription. *Curr. Genetics Dev.* **4**, 255-259.

Edmondson,D.G., Lyons,G.E., Martin,J.F. and Olson,E.N. (1994) *Mef2* gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development*, **120**, 1251-1263.

Edwards,A.M., Kane,C.M., Young,R.A. and Kornberg,R.D. (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J. Biol. Chem.* **266**, 71-75.

Eisenberg,C.A. and Bader,D. (1995) QCE-6: a clonal cell line with cardiac myogenic and endothelial cell potentials. *Dev. Biol.* **167**, 469-481.

Eisenberg,L.M. and Markwald,R.R. (1995) Molecular regulation of atrioventricular valvuloseptal morphogenesis. *Circ. Res.* **77**, 1-6.

Epstein,D.J., Vekemans,M. and Gros,P. (1991) Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell*, **67**, 767-774.

Ernst,H., Walsh,K., Harrison,C.A. and Rosenthal,N. (1991) The myosin light chain enhancer and the skeletal actin promoter share a binding site for factors involved in muscle-specific gene expression. *Mol. Cell. Biol.* **11**, 3735-3744.

Estevez,M., Attisano,L., Wrana,J.L., Albert,P.S., Massague,J. and Riddle,D.L. (1993) The daf-4 gene encodes a bone morphogenetic protein receptor controlling c. elegans dauer larva development. *Nature*, **365**, 644

Evans,S.M., Tai,L.J., Tan,V.P., Newton,C.B. and Chien,K.R. (1994) Heterokaryons of cardiac myocytes and fibroblasts reveal the lack of dominance of the cardiac muscle phenotype. *Mol. Cell. Biol.* **14**, 4269-4279.

Evans,S.M., Yan,W., Murillo,M.P., Ponce,J. and Papalopulu,N. (1995) Tinman, a drosophila homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates - xnkx-2.3, a second vertebrate homologue of tinman. *Development*, **121**, 3889-3899.

Evans,S.M. and O'Brien,T.X. (1993) Expression of the helix-loop-helix factor Id during mouse embryonic development. *Dev. Biol.* **159**, 485-499.

Evans,T. and Felsenfeld,G. (1989) The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell*, **58**, 877-885.

Farrance,I.K., Mar,J.H. and Ordahl,C.P. (1992) M-CAT binding factor is related to the SV40 enhancer binding factor, TEF-1. *J. Biol. Chem.* **267**, 17234-17240.

Field,L.J. (1988) Atrial natriuretic factor-SV40 T antigen transgenes produce tumors and cardiac arrythmias in mice. *Science*, **239**, 1029-1033.

Fishman,M.C. and Chien,K.R. (1997) Fashioning the vertebrate heart: earliest embryonic decisions. *Development*, **124**, 2099-2117.

Flanagan,P.M., Kelleher,R.J.,3d, Sayre,M.H., Tschochner,H. and Kornberg,R.D. (1991) A mediator required for activation of RNA polymerase II transcription in vitro. *Nature*, **350**, 436-438.

Fong,T.C. and Emerson,B.M. (1992) The erythroid-specific protein cGATA-1 mediates distal enhancer activity through a specialized beta-globin TATA box. *Genes Dev.* **6**, 521-532.

Forman,B.M., Umesono,K., Chen,J. and Evans,R.M. (1995) Unique response pathways are established by allosteric interactions among nuclear hormone receptors. *Cell*, **81**, 541-550.

Forrester,W.C., Thompson,C., Elder,J.T. and Groudine,M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc. Natl. Acad. Sci. USA*, **83**, 1359-1363.

Forrester,W.C., van Genderen,C., Jenuwein,T. and Grosschedl,R. (1994) Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. *Science*, **265**, 1221-1225.

Frasch,M. (1995) Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early Drosophila embryo. *Nature*, **374**, 464-467.

Frasch,M. and Levine,M. (1987) Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. *Genes Dev.* **1**, 981-995.

Fu,Y.C. and Izumo,S. (1995) Cardiac myogenesis - overexpression of xcsx2 or xmef2a in whole xenopus embryos induces the precocious expression of xmhc-alpha gene. *Rouxs Arch. Dev. Biol.* **205**, 198-202.

Fujiwara,Y., Browne,C.P., Cunniff,K., Goff,S.C. and Orkin,S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA*, **93**, 12355-12358.

Gajewski,K., Kim,Y., Lee,Y.M., Olson,E.N. and Schulz,R.A. (1997) D-mef2 is a target for Tinman activation during Drosophila heart development. *EMBO J.* **16**, 515-522.

Gajewski,T.F. and Thompson,C.B. (1996) Apoptosis meets signal transduction: elimination of a BAD influence. [Minireview]. *Cell*, **87**, 589-592.

Galipeau,J., Nemer,M. and Drouin,J. (1988) Ventricular activation of the atrial natriuretic factor gene in acute myocardial infarction. *N. Engl. J. Med.* **319**, 654-655.

Garcia-Martinez,V. and Schoenwolf,G.C. (1993) Primitive-streak origin of the cardiovascular system in avian embryos. *Dev. Biol.* **159**, 706-719.

Gasser,S.M., Laroche,T., Falquet,J., Boy de la Tour,E. and Laemmli,U.K. (1986) Metaphase chromosome structure. Involvement of topoisomerase II. *J. Mol. Biol.* **188**, 613-629.

Gasser,S.M. and Laemmli,U.K. (1986) Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. *Cell*, **46**, 521-530.

Gassmann,M., Casagranda,F., Orioli,D., Simon,H., Lai,C., Klein,R. and Lemke,G. (1995) Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature*, **378**, 390-4:19.

Ge,H. and Manley,J.L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell*, **62**, 25-34.

Gehring,W.J., Affolter,M. and Burglin,T. (1994a) Homeodomain proteins. [Review]. *Annu. Rev. Biochem.* **63**, 487-526.

Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubo-Tokunaga,K., Schier,A.F., Resendez-Perez,D., Affolter,M., Otting,G. and Wüthrich,K. (1994b) Homeodomain-DNA recognition. *Cell*, **78**, 211-223.

Geiduschek,E.P. (1992) Two prokaryotic transcriptional enhancer systems. [Review]. *Prog. Nucleic Acid. Res. & . Mol. Biol.* **43**, 109-133.

Gerber,H.P., Hagmann,M., Seipel,K., Georgiev,O., West,M.A., Litingtung,Y., Schaffner,W. and Corden,J.L. (1995) RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature*, **374**, 660-662.

Ghosh,S. and Baltimore,D. (1990) Activation in vitro of NF-kappaB by phosphorylation of its inhibitor IkappaB. *Nature*, **344**, 678-682.

Giffin,W., Torrance,H., Rodda,D.J., Prefontaine,G.G., Pope,L. and Hache,R.J. (1996) Sequence-specific DNA binding by Ku autoantigen and its effects on transcription. *Nature*, **380**, 265-268.

Gilbert, S.F. 1994. *Developmental Biology*, Ed.4th Sinauer Associates, Inc. Sunderland, Massachusetts:pp. 1-894.

Gill,G. and Tjian,R. (1991) A highly conserved domain of tfiid displays species specificity in vivo. *Cell*, **65**, 333-40: 17.

Gilmartin,G.M. and Nevins,J.R. (1989) An ordered pathway of assembly of components required for polyadenylation site recognition and processing. *Genes Dev.* **3**, 2180-2190.

Gisselbrecht,S., Skeath,J.B., Doe,C.Q. and Michelson,A.M. (1996) heartless encodes a fibroblast growth factor receptor (DFR1/DFGF-R2) involved in the directional migration of early mesodermal cells in the Drosophila embryo. *Genes Dev.* **10**, 3003-3017.

Gopal-Srivastava,R., Haynes,J.I.,2nd and Piatigorsky,J. (1995) Regulation of the murine alpha b-crystallin/small heat shock protein gene in cardiac muscle. *Mol. Cell. Biol.* **15**, 7081-7090.

Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Struct. Biol.* **3**, 419-429.

Gossett,L.A., Kelvin,D.J., Sternberg,E.A. and Olson,E.N. (1989) A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol. Cell. Bioi.* **9**, 5022-5033.

Gourdie,R.G., Green,C.R., Severs,N.J., Anderson,R.H. and Thompson,R.P. (1993) Evidence for a distinct gap-junctional phenotype in ventricular conduction tissues of the developing and mature avian heart. *Circ. Res.* **72**, 278-289.

Gourdie,R.G., Mima,T., Thompson,R.P. and Mikawa,T. (1995) Terminal diversification of the myocyte lineage generates Purkinje fibers of the cardiac conduction system. *Development*, **121**, 1423-1431.

Gove,C., Walmsley,M., Nijjar,S., Bertwistle,D., Guille,M., Partington,G., Bomford,A. and Patient,R. (1997) Over-expression of GATA-6 in Xenopus embryos blocks differentiation of heart precursors. *EMBO J.* **16**, 355-368.

Grépin,C., Dagnino,L., Robitaille,L., Haberstroh,L., Antakly,T. and Nemer,M. (1994) A hormone-encoding gene identifies a pathway for cardiac but not skeletal muscle gene transcription. *Mol. Cell. Biol.* **14**, 3115-3129.

Grépin,C., Durocher,D. and Nemer,M. (1995a) Le coeur : un programme unique de transcription et de différenciation musculaire. *M/S*, **11**, 395-405.

Grépin,C., Robitaille,L., Antakly,T. and Nemer,M. (1995b) Inhibition of transcription factor GATA-4 expression blocks *in vitro* cardiac muscle differentiation. *Mol. Cell. Biol.* **15**, 4095-4102.

Grépin,C., Nemer,G. and Nemer,M. (1997) Enhanced cardiogenesis in embryonic stem cells overexpressing the GATA-4 transcription factor. *Development*, **124**, 2387-2395.

Grosschedl,R. (1995) Higher-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr. Opin. Cell Biol.* **7**, 362-370.

Grosveld,F., Blom van Assendelft,G., Greaves,D.R. and Kollias,G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-985.

Grueneberg,D.A., Natesan,S., Alexandre,C. and Gilman,M.Z. (1992) Human and Drosophila homeodomain proteins that enhance the DNA-binding activity of serum response factor. *Science*, **257**, 1089-1095.

Gruschus,J.M., Tsao,D.H., Wang,L.H., Nirenberg,M. and Ferretti,J.A. (1997) Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry*, **36**, 5372-5380.

Gualberto,A., Lepage,D., Pons,G., Mader,S.L., Park,K., Atchison,M.L. and Walsh,K. (1992) Functional antagonism between YY1 and the serum response factor. *Mol. Cell. Biol.* **12**, 4209-4214.

Guazzi,S., Price,M., De Felice,M., Damante,G., Mattei,M.G. and Di Lauro,R. (1990) Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J.* **9**, 3631-3639.

Guichet,A., Copeland,J.W., Erdélyl,M., Hlousek,D., Zàvorszky,P., Ho,J., Brown,S., Percival-Smith,A., Krause,H.M. and Ephrussi,A. (1997) The nuclear receptor homologue Ftz-F1 and the homeodomain protein Ftz are mutually dependent cofactors. *Nature*, **385**, 548-552.

Guillemin,K., Groppe,J., Ducker,K., Treisman,R., Hafen,E., Affolter,M. and Krasnow,M.A. (1996) The pruned gene encodes the Drosophila serum response factor and regulates cytoplasmic outgrowth during terminal branching of the tracheal system. *Development*, **122**, 1353-1362.

Gupta,M.P., Gupta,M. and Zak,R. (1994) An E-box/M-CAT hybrid motif and cognate binding protein(s) regulate the basal muscle-specific and camp-inducible expression of the rat cardiac alpha-myosin heavy chain gene. *J. Biol. Chem.* **269**, 29677-29687.

Haffter,P., Granato,M., Brand,M., Mullins,M.C., Hammerschmidt,M., Kane,D.A., Odenthal,J., van Eeden,F.J., Jiang,Y.J., Heisenberg,C.P., Kelsh,R.N., Furutani-Seiki,M., Vogelsang,E., Beuchle,D., Schach,U., Fabian,C. and Nusslein-Volhard,C. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development*, **123**, 1-36.

Hahn,S., Buratowski,S., Sharp,P.A. and Guarente,L. (1989) Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions. *Cell*, **58**, 1173-1181.

Halaban,R., Ghosh,S. and Baird,A. (1987) bFGF is the putative natural growth factor for human melanocytes. *In Vitro Cell. Dev. Biol. An.* **23**, 47-52.

Hanna-Rose,W. and Hansen,U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* **12**, 229-34: 4.

Harvey,R.P. (1996) NK-2 homeobox genes and heart development [Review]. *Dev. Biol.* **178**, 203-216.

Hasty,P., Bradley,A., Hsi Morris,J., Edmondson,D.G., Venuti,J.M., Olson,E.N. and Klein,W.H. (1993) Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature*, **364**, 501-506.

Hecht,A., Laroche,T., Strahl-Bolsinger,S., Gasser,S.M. and Grunstein,M. (1995) Histone h3 and h4 n-termini interact with sir3 and sir4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell*, **80**, 583-92:19.

Hedley,M.L. and Maniatis,T. (1991) Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell*, **65**, 579-586.

Heikinheimo,M., Scandrett,J.M. and Wilson,D.B. (1994) Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. *Dev. Biol.* **164**, 361-373.

Hemmati-Brivanlou,A. and Melton,D. (1997) Vertebrate embryonic cells will become nerve cells unless told otherwise. *Cell*, **88**, 13-17.

Herschbach,B.M. and Johnson,A.D. (1993) Transcriptional repression in eukaryotes. *Annu. Rev. Cell Biol.* **9**, 479-509:2.

Hershey,J.W. (1989) Protein phosphorylation controls translation rates. [Review] [59 refs]. *J. Biol. Chem.* **264**, 20823-20826.

Hershey,J.W.B. (1991) Translational control in mammalian cells. *Annu. Rev. Biochem.* **60**, 717-755.

Hirschhorn,J.N., Brown,S.A., Clark,C.D. and Winston,F. (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* **6**, 2288-2298.

Hoffman,A., Sinn,E., Yamamoto,T., Wang,J., Roy,A., Horikoshi,M. and Roeder,R.G. (1990) Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). *Nature*, **346**, 387-390.

Hollenberg,S.M., Sternglanz,R., Cheng,P.F. and Weintraub,H. (1995) Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol. Cell. Biol.* **15**, 3813-3822.

Horikoshi,M., Wangk,C.K., Fujii,H., Cromlish,J.A., Weil,P.A. and Roeder,R.G. (1989) Cloning and structure of a yeast gene encoding a general transcription initiation factor TFIID that binds to the TATA box. *Nature*, **341**, 299-303.

Icardo, J.M. and Manasek, F.J. (1996) Cardiogenesis: development mechanisms and embryology. In Fozzard,H.A. and et al. *The heart and cardiovascular system* - Raven Press, Ltd. New York, pp. 1563-1586.

Iguchi-Ariga,S.M. and Schaffner,W. (1989) Cpg methylation of the camp-responsive enhancer/promoter sequence tgacgtca abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* **3**, 612-9: 20.

Ince,T.A. and Scotto,K.W. (1995) A conserved downstream element defines a new class of RNA polymerase II promoters. *J. Biol. Chem.* **270**, 30249-30252.

Ip,H.S., Wilson,D.B., Heikinheimo,M., Tang,Z., Ting,C.N., Simon,M.C., Leiden,J.M. and Parmacek,M.S. (1994) The GATA-4 transcription factor transactivates the cardiac muscle-specific troponin C promoter-enhancer in nonmuscle cells. *Mol. Cell. Biol.* **14**, 7517-7526.

Isaac,A., Sargent,M.G. and Cooke,J. (1997) Control of vertebrate left-right asymmetry by a snail-related zinc finger gene. *Science*, **275**, 1301-4:27.

Izumo,S., Nadal-Ginard,B. and Mahdavi,V. (1988) Protooncogene induction and reprogramming of cardiac gene expression produced by pressure overload. *Proc. Natl. Acad. Sci. USA*, **85**, 339-343.

Jaber,M., Koch,W.J., Rockman,H., Smith,B., Bond,R.A., Sulik,K.K., Ross,J.,Jr., Lefkowitz,R.J., Caron,M.G. and Giros,B. (1996) Essential role of beta-adrenergic receptor kinase 1 in cardiac development and function. *Proc. Natl. Acad. Sci. USA*, **93**, 12974-12979.

Jacobson,A. and Peltz,S.W. (1996) Interrelationships of the pathways of mrna decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* **65**, 693-739: 6.

Jacobson,A.G. and Sater,A.K. (1988) Features of embryonic induction. [Review]. *Development*, **104**, 341-359.

Jacquemin,P., Hwang,J.J., Martial,J.A., Dolle,P. and Davidson,I. (1996) A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS DNA binding domain. *J. Biol. Chem.* **271**, 21775-21785.

Javahery,R., Khachi,A., Lo,K., Zenzie-Gregory,B. and Smale,S.T. (1994) DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* **14**, 116-127.

Jiang,Y.M. and Evans,T. (1996) The Xenopus GATA-4/5/6 genes are associated with cardiac specification and can regulate cardiac-specific transcription during embryogenesis. *Dev. Biol.* **174**, 258-270.

Jimenez,F., Martin-Morris,L.E., Velasco,L., Chu,H., Sierra,J., Rosen,D.R. and White,K. (1995) vnd, a gene required for early neurogenesis of Drosophila, encodes a homeodomain protein. *EMBO J.* **14**, 3487-3495.

John,S., Robbins,C.M. and Leonard,W.J. (1996) An IL-2 response element in the human IL-2 receptor α chain promoter is a composite element that binds Stat5, Elf-1, HMG-I(Y) and GATA family protein. *EMBO J.* **15**, 5627-5635.

Johnson,A.D. (1995) The price of repression. *Cell*, **81**, 655-8:22.

Johnson,R.C., Manasek,F.J., Vinson,W.C. and Seyer,J.M. (1974) The biochemical and ultrastructural demonstration of collagen during early heart development. *Dev. Biol.* **36**, 252-271.

Jonat,C., Rahmsdorf,H.J., Park,K.-K., Cato,A.C.B., Gebel,S., Ponta,H. and Herrlich,P. (1990) Antitumor promotion and antiinflammation: down-modulation of AP-1 (fos/jun) activity by glucocorticoid hormone. *Cell*, **62**, 1189-1204.

Joulin,V., Bories,D., Eleouet,J.F., Labastie,M.C., Chretien,S., Mattei,M.G. and Romeo,P.H. (1991) A T-cell specific TCR delta DNA binding protein is a member of the human GATA family. *EMBO J.* **10**, 1809-1816.

Jun,S. and Desplan,C. (1996) Cooperative interactions between paired domain and homeodomain. *Development*, **122**, 2639-2650.

Kadosh,D. and Struhl,K. (1997) Repression by ume6 involves recruitment of a complex containing sin3 corepressor and rpd3 histone deacetylase to target promoters. *Cell*, **89**, 365-371:

Kamei,Y., Xu,L., Heinzel,T., Torchia,J., Kurokawa,R., Gloss,B., Lin,S.C., Heyman,R.A., Rose,D.W., Glass,C.K. and Rosenfeld,M.G. (1996) A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell*, **85**, 403-414.

Kao,C.C., Lieberman,P.M., Schmidt,M.C., Zhou,Q., Pei,R. and Berk,A.J. (1990) Cloning of a transcriptionally active human TATA binding factor. *Science*, **248**, 1646-1650.

Karns,L.R., Kariya,K. and Simpson,P.C. (1995) M-CAT, CArG, and Sp1 elements are required for alpha 1-adrenergic induction of the skeletal alpha-actin promoter during cardiac myocyte hypertrophy. Transcriptional enhancer factor-1 and protein kinase C as conserved transducers of the fetal program in cardiac growth. *J. Biol. Chem.* **270**, 410-417.

Katz,E.B., Steinhelper,M.E., Delcarpio,J.B., Daud,A.I., Claycomb,W.C. and Field,L.J. (1992) Cardiomyocyte proliferation in mice expressing alpha-cardiac myosin heavy chain-SV40 T-antigen transgenes. *Am. J. Physiol.* **262**, H1867-H1876.

Kaufmann,J. and Smale,S.T. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* **8**, 821-829.

Kaushal,S., Schneider,J.W., Nadal-Ginard,B. and Mahdavi,V. (1994) Activation of the myogenic lineage by mef2a, a factor that induces and cooperates with myod. *Science*, **266**, 1236-40:

Kawana,M., Lee,M.E., Quertermous,E.E. and Quertermous,T. (1995) Cooperative interaction of gata-2 and ap1 regulates transcription of the endothelin-1 gene. *Mol. Cell. Biol.* **15**, 4225-31: 3.

Kelleher,R.J.,3d, Flanagan,P.M., Chasman,D.I., Ponticelli,A.S., Struhl,K. and Kornberg,R.D. (1992) Yeast and human tfiids are interchangeable for the response to acidic transcriptional activators in vitro. *Genes Dev.* **6**, 296-303: 5.

Kennison,J.A. (1995) The Polycomb and trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. *Annu. Rev. Genet.* **29**, 289-303.

Kim,T.K., Hashimoto,S., Kelleher,R.J.,3rd, Flanagan,P.M., Kornberg,R.D., Horikoshi,M. and Roeder,R.G. (1994) Effects of activation-defective TBP mutations on transcription initiation in yeast. *Nature*, **369**, 252-255.

Kim,Y.J., Bjorklund,S., Li,Y., Sayre,M.H. and Kornberg,R.D. (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, **77**, 599-608.

Kingston,R.E., Bunker,C.A. and Imbalzano,A.N. (1996) Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev.* **10**, 905-920.

Kissinger,C.R., Liu,B.S., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**, 579-590.

Klein,C. and Struhl,K. (1994) Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains in vivo. *Science*, **266**, 280-282.

Knoepfler,P.S. and Kamps,M.P. (1995) The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. *Mol. Cell. Biol.* **15**, 5811-5819.

Ko,L.J. and Engel,J.D. (1993) DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.* **13**, 4011-4022.

Koleske,A.J. and Young,R.A. (1995) The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* **20**, 113-116.

Kramer,A. (1996) The structure and function of proteins involved in mammalian pre- mrna splicing. *Annu. Rev. Biochem.* **65**, 367-409: 7.

Kreidberg,J.A., Sariola,H., Loring,J.M., Maeda,M., Pelletier,J., Housman,D. and Jaenisch,R. (193) WT-1 is required for early kidney development. *Cell*, **74**, 679-691.

Kruger,W., Peterson,C.L., Sil,A., Coburn,C., Arents,G., Moudrianakis,E.N. and Herskowitz,I. (1995) Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev.* **9**, 2770-2779.

Krumm,A. and Groudine,M. (1995) Tumor suppression and transcription elongation: the dire consequences of changing partners. *Science*, **269**, 1400-1401.

Kuchin,S., Yeghiayan,P. and Carlson,M. (1995) Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast. *Proc. Natl. Acad. Sci. USA*, **92**, 4006-4010.

Kuo,C.T., Morrisey,E.E., Anandappa,R., Sigrist,K., Lu,M.M., Parmacek,M.S., Soudais,C. and Leiden,J.M. (1997) GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048-1060.

Kuo,M.H., Brownell,J.E., Sobel,R.E., Ranalli,T.A., Cook,R.G., Edmondson,D.G., Roth,S.Y. and Allis,C.D. (1996) Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature*, **383**, 269-272.

Kurihara,Y., Kurihara,H., Oda,H., Maemura,K., Nagai,R., Ishikawa,T. and Yazaki,Y. (1995) Aortic arch malformations and ventricular septal defect in mice deficient in endothelin-1. *J. Clin. Invest.* **96**, 293-300.

Kwee,L., Baldwin,H.S., Shen,H.M., Stewart,C.L., Buck,C., Buck,C.A. and Labow,M.A. (1995) Defective development of the embryonic and extraembryonic circulatory systems in vascular cell adhesion molecule (VCAM-1) deficient mice. *Development*, **121**, 489-503.

Kwok,R.P.S., Lundblad,J.R., Chrivia,J.C., Richards,J.P., Bächinger,H.P., Brennan,R.G., Roberts,S.G.E., Green,M.R. and Goodman,R.H. (1994) Nuclear protein CBP is a coactivator for the transcription factor CREB. *Nature*, **370**, 223-226.

Laherty,C.D., Yang,W.M., Sun,J.M., Davie,J.R., Seto,E. and Eisenman,R.N. (1997) Histone deacetylases associated with the msin3 corepressor mediate mad transcriptional repression. *Cell*, **89**, 349-356: 7.

Lai,Z.C., Fortini,M.E. and Rubin,G.M. (1991) The embryonic expression patterns of zfh-1 and zfh-2, two Drosophila genes encoding novel zinc-finger homeodomain proteins. *Mech. Dev.* **34**, 123-134.

Laoide,B.M., Foulkes,N.S., Schlotter,F. and Sassone-Corsi,P. (1993) The functional versatility of crem is determined by its modular structure. *EMBO J.* **12**, 1179-91:

Larkin,S.B., Farrance,I.K. and Ordahl,C.P. (1996) Flanking sequences modulate the cell specificity of M-CAT elements. *Mol. Cell. Biol.* **16**, 3742-3755.

Laurent,B.C., Treitel,M.A. and Carlson,M. (1991) Functional interdependence of the yeast snf2, snf5, and snf6 proteins in transcriptional activation. *Proc. Natl. Acad. Sci. USA*, **88**, 2687-91: 15.

Laurent,B.C., Treich,I. and Carlson,M. (1993) The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev.* **7**, 583-591.

Laverriere,A.C., MacNeill,C., Mueller,C., Poelmann,R.E., Burch,J.B. and Evans,T. (1994) GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut. *J. Biol. Chem.* **269**, 23177-23184.

Lawson,K.A., Meneses,J.J. and Pedersen,R.A. (1991) Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, **113**, 891-911.

Lawson,M.A., Whyte,D.B. and Mellon,P.L. (1996) GATA factors are essential for activity of the neuron-specific enhancer of the gonadotropin-releasing hormone gene. *Mol. Cell. Biol.* **16**, 3596-3605.

Laybourn,P.J. and Dahmus,M.E. (1990) Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription. *J. Biol. Chem.* **265**, 13165-13173.

Lee,K.F., Simon,H., Chen,H., Bates,B., Hung,M.C. and Hauser,C. (1995) Requirement for neuregulin receptor erbB2 in neural and cardiac development [see comments]. *Nature*, **378**, 394-398.

Lee,K.H., Xu,Q. and Breitbart,R.E. (1996) A new *tinman*-related gene, *nkx2.7*, anticipates the expression of *nkx2.5* and *nkx2.3* in Zebrafish heart and pharyngeal endocerm. *Dev. Biol.* **180**, 722-731.

Lee,R.K., Stainier,D.Y., Weinstein,B.M. and Fishman,M.C. (1994) Cardiovascular development in the zebrafish. II. Endocardial progenitors are sequestered within the heart field. *Development*, **120**, 3361-3366.

Lee,T.-C., Shi,Y. and Schwartz,R.J. (1992) Displacement of BrdUrd-induced YY1 by serum response factor activates skeletal alpha-actin transcription in embryonic myoblasts. *Proc. Natl. Acad. Sci. USA*, **89**, 9814-9818.

Lewis,J.D., Gunderson,S.I. and Mattaj,I.W. (1995) The influence of 5' and 3' end structures on pre-mRNA metabolism. *J. Cell Sci. -. Supp.* **19**, 13-19.

Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262-269.

Liao,S.M., Zhang,J., Jeffery,D.A., Koleske,A.J., Thompson,C.M., Chao,D.M., Viljoen,M., van Vuuren,H.J. and Young,R.A. (1995) A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature*, **374**, 193-196.

Lilly,B., Zhao,B., Ranganayakulu,G., Paterson,B.M., Schulz,R.A. and Olson,E.N. (1995) Requirement of MADS domain transcription factor D-MEF2 for muscle formation in Drosophila. *Science*, **267**, 688-693.

Lin,Q., Schwarz,J., Bucana,C. and Olson,E.N. (1997) Control of mouse cardiac morphogenesis and myogenesis by transcription factor mef2c. *Science*, **276**, 1404-1407:2.

Lin,T.A., Kong,X., Haystead,T.A., Pause,A., Belsham,G., Sonenberg,N. and Lawrence,J.C.,Jr. (1994) Phas-i as a link between mitogen-activated protein kinase and translation initiation. *Science*, **266**, 653-6: 16.

Linask,K.K. and Lash,J.W. (1993) Early heart development: dynamics of endocardial cell sorting suggests a common origin with cardiomyocytes. *Dev. Dynamics*, **196**, 62-69.

Lints,T.J., Parsons,L.M., Hartley,L., Lyons,I. and Harvey,R.P. (1993) Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development*, **119**, 419-431.

Litvin,J., Montgomery,M., Gonzalez-Sanchez,A., Bisaha,J.G. and Bader,D. (1992) Commitment and differentiation of cardiac myocytes. *Trends Cardiovasc. Med.* **2**, 27-32.

Lohr,D., Venkov,P. and Zlatanova,J. (1995) Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *FASEB J.* **9**, 777-787.

Lompré,A.-M., Nadal-Ginard,B. and Mahdavi,V. (1984) Expression of the cardiac ventricular alph- and beta-myosin heavy chain genes is developmentally and hormonally regulated. *J. Biol. Chem.* **259**, 6437-6446.

Loo,S. and Rine,J. (1995) Silencing and heritable domains of gene expression. *Annu. Rev. Cell Dev. Biol.* **11**, 519-48:16.

Lough,J., Barron,M., Brogley,M., Sugi,Y., Bolender,D.L. and Zhu,X. (1996) Combined bmp-2 and fgf-4, but neither factor alone, induces cardiogenesis in non-precardiac embryonic mesoderm. *Dev. Biol.* **178**, 198-202:21.

Lowe,L.A., Supp,D.M., Sampath,K., Yokoyama,T., Wright,C.V., Potter,S.S., Overbeek,P. and Kuehn,M.R. (1996) Conserved left-right asymmetry of nodal expression and alterations in murine situs inversus. *Nature*, **381**, 158-61:25.

Lu,H., Zawel,L., Fisher,L., Egly,J.M. and Reinberg,D. (1992) Human general transcription factor IIH phosphorylates the C- terminal domain of RNA polymerase II. *Nature*, **358**, 641-645.

Lyons,G.E. (1994) In situ analysis of the cardiac muscle gene program during embryogenesis. *Trends Cardiovasc. Med.* **4**, 70-77.

Lyons,G.E. (1996) Vertebrate heart development. *Curr. Genetics Dev.* **6**, 454-460.

Lyons,I., Parsons,L.M., Hartley,L., Li,R., Andrews,J.E., Robb,L. and Harvey,R.P. (1995) Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev.* **9**, 1654-1666.

MacDougall,C., Harbison,D. and Bownes,M. (1995) The developmental consequences of alternate splicing in sex determination and differentiation in drosophila. *Dev. Biol.* **172**, 353-76: 14.

Manasek,F.J. (1968) Mitosis in developing cardiac muscle. *J. Cell Biol.* **37**, 191-196.

Manasek,F.J., Burnside,M.B. and Waterman,R.E. (1972) Myocardial cell shape change as a mechanism of embryonic heart looping. *Dev. Biol.* **29**, 349-371.

Manley,J.L. and Tacke,R. (1996) Sr proteins and splicing control. *Genes Dev.* **10**, 1569-79: 12.

Manuelidis,L. (1990) A view of interphase chromosomes. *Science*, **250**, 1533-1540.

Mar,J.H. and Ordahl,C.P. (1988) A conserved CATTCCT motif is required for skeletal muscle-specific activity of the cardiac troponin T gene promoter. *Proc. Natl. Acad. Sci. USA*, **85**, 6404-6408.

Marais,R., Wynne,J. and Treisman,R. (1993) The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell*, **73**, 381-393.

Marshall,N.F. and Price,D.H. (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.* **270**, 12335-12338.

Martin,D.I., Zon,L.I., Mutter,G. and Orkin,S.H. (1990) Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. *Nature*, **344**, 444-447.

Martin,D.I., Fiering,S. and Groudine,M. (1996) Regulation of beta-globin gene expression: straightening out the locus. *Curr. Genetics Dev.* **6**, 488-95: 5.

Martin,J.F., Schwarz,J.J. and Olson,E.N. (1993) Myocyte enhancer factor (MEF) 2C: a tissue-restricted member of the MEF-2 family of transcription factors. *Proc. Natl. Acad. Sci. USA*, **90**, 5282-5286.

Martin,J.F., Miano,J.M., Hustad,C.M., Copeland,N.G., Jenkins,N.A. and Olson,E.N. (1994) A MEF2 gene that generates a muscle-specific isoform via alternative mRNA splicing. *Mol. Cell. Biol.* **14**, 1647-1656.

Mattaj,I.W. (1994) RNA processing. Splicing in space. *Nature*, **372**, 727-728.

Maxon,M.E., Goodrich,J.A. and Tjian,R. (1994) Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev.* **8**, 515-524.

McBride,K., Robitaille,L., Tremblay,S., Argentin,S. and Nemer,M. (1993) Fos/jun repression of cardiac-specific transcription in quiescent and growth stimulated myocytes is targeted at a tissue-specific cis-element. *Mol. Cell. Biol.* **13**, 600-612.

McBurney,M.W., Jones-Villeneuve,E.M., Edwards,M.K. and Anderson,P.J. (1982) Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. *Nature*, **299**, 165-167.

McGrew,M.J., Bogdanova,N., Hasegawa,K., Hughes,S.H., Kitsis,R.N. and Rosenthal,N. (1996) Distinct gene expression patterns in skeletal and cardiac muscle are dependent on common regulatory sequences in the MLC1/3 locus. *Mol. Cell. Biol.* **16**, 4524-4538.

McKnight,G.S., Clegg,C.H., Uhler,M.D., Chrivia,J.C., Cadd,G.G., Correll,L.A. and Otten,A.D. (1988) Analysis of the cAMP-dependent protein kinase system using molecular genetic approaches. *Rec. Prog. Horm. Res.* **44**, 307-335.

Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L. and Bird,A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, **58**, 499-507.

Meno,C., Saijoh,Y., Fujii,H., Ikeda,M., Yokoyama,T., Yokoyama,M., Toyoda,Y. and Hamada,H. (1996) Left-right asymmetric expression of the TGF beta-family member lefty in mouse embryos. *Nature*, **381**, 151-155.

Merika,M. and Orkin,S.H. (1993) DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.* **13**, 3999-4010.

Merika,M. and Orkin,S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor gata-1 with the kruppel family proteins sp1 and eklf. *Mol. Cell. Biol.* **15**, 2437-47: 5.

Meurs,E., Chong,K., Galabru,J., Thomas,N.S., Kerr,I.M., Williams,B.R. and Hovanessian,A.G. (1990) Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell*, **62**, 379-390.

Meyer,D. and Birchmeier,C. (1995) Multiple essential functions of neuregulin in development [see comments]. *Nature*, **378**, 386-90:18.

Mikawa,T., Borisov,A., Brown,A.M. and Fischman,D.A. (1992a) Clonal analysis of cardiac morphogenesis in the chicken embryo using a replication-defective retrovirus: i. formation of the ventricular myocardium. *Dev. Dynamics*, **193**, 11-23:7.

Mikawa,T., Cohen-Gould,L. and Fischman,D.A. (1992b) Clonal analysis of cardiac morphogenesis in the chicken embryo using a replication-defective retrovirus. III: Polyclonal origin of adjacent ventricular myocytes. *Dev. Dynamics*, **195**, 133-141.

Mikawa,T. and Fischman,D.A. (1996) The polyclonal origin of myocyte lineages. [review] [76 refs]. *Annu. Rev. Physiol.* **58**, 509-21:20.

Mima,T., Ueno,H., Fischman,D.A., Williams,L.T. and Mikawa,T. (1995) Fibroblast growth factor receptor is required for in vivo cardiac myocyte proliferation at early embryonic stages of heart development. *Proc. Natl. Acad. Sci. USA*, **92**, 467-471.

Mirkovitch,J., Mirault,M.E. and Laemmli,U.K. (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, **39**, 223-232.

Mizuta,T.R., Fukita,Y., Miyoshi,T., Shimizu,A. and Honjo,T. (1993) Isolation of cDNA encoding a binding protein specific to 5'-phosphorylated single-stranded DNA with G-rich sequences. *Nucleic Acids Res.* **21**, 1761-1766.

Mizzen,C.A., Yang,X.J., Kokubo,T., Brownell,J.E., Bannister,A.J., Owen-Hughes,T., Workman,J., Wang,L., Berger,S.L., Kouzarides,T., Nakatani,Y. and Allis,C.D. (1996) The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*, **87**, 1261-1270.

Moehrle,A. and Paro,R. (1994) Spreading the silence: epigenetic transcriptional regulation during drosophila development. *Dev. Genet.* **15**, 478-84: 16.

Mohun,T.J., Taylor,M.B., Garrett,N. and Gurdon,J.B. (1989) The CArG promoter sequence is necessary for muscle-specific transcription of the cardiac actin gene in Xenopus embryos. *EMBO J.* **8**, 1153-1161.

Mohun,T.J., Chambers,A.E., Towers,N. and Taylor,M.V. (1991) Expression of genes encoding the transcription factor SRF during early development of Xenopus laevis: identification of a CArG box- binding activity as SRF. *EMBO J.* **10**, 933-940.

Molkentin,J.D., Brogan,R.S., Jobe,S.M. and Markham,B.E. (1993) Expression of the alpha-myosin heavy chain gene in the heart is regulated in part by an E-box-dependent mechanism. *J. Biol. Chem.* **268**, 2602-2609.

Molkentin,J.D., Kalvakolanu,D.V. and Markham,B.E. (1994) Transcription factor GATA-4 regulates cardiac muscle-specific expression of the α-myosin heavy-chain gene. *Mol. Cell. Biol.* **14**, 4947-4957.

Molkentin,J.D., Black,B.L., Martin,J.F. and Olson,E.N. (1995) Cooperative activation of muscle gene expression by mef2 and myogenic bhlh proteins. *Cell*, **83**, 1125-1136: 2.

Molkentin,J.D., Firulli,A.B., Black,B.L., Martin,J.F., Hustad,C.M., Copeland,N., Jenkins,N., Lyons,G. and Olson,E.N. (1996) MEF2B is a potent transactivator expressed in early myogenic lineages. *Mol. Cell. Biol.* **16**, 3814-3824.

Molkentin,J.D., Lin,Q., Duncan,S.A. and Olson,E.N. (1997) Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061-1072.

Molkentin,J.D. and Markham,B.E. (1993) Myocyte-specific enhancer-binding factor (MEF-2) regulates alpha-cardiac myosin heavy chain gene expression in vitro and in vivo. *J. Biol. Chem.* **268**, 19512-19520.

Molkentin,J.D. and Markham,B.E. (1994) An M-CAT binding factor and an RSRF-related A-rich binding factor positively regulate expression of the alpha-cardiac myosin heavy-chain gene in vivo. *Mol. Cell. Biol.* **14**, 5056-5065.

Monkley,S.J., Delaney,S.J., Pennisi,D.J., Christiansen,J.H. and Wainwright,B.J. (1996) Targeted disruption of the Wnt2 gene results in placentation defects. *Development*, **122**, 3343-3353.

Montgomery,M.O., Litvin,J., Gonzalez-Sanchez,A. and Bader,D. (1994) Staging of commitment and differentiation of avian cardiac myocytes. *Dev. Biol.* **164**, 63-71.

Moqtaderi,Z., Bai,Y., Poon,D., Weil,P.A. and Struhl,K. (1996) TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature*, **383**, 188-191.

Morgan,H.E. and Baker,K.M. (1991) Cardiac hypertrophy. Mechanical, neural, and endocrine dependence. *Circulation*, **83**, 13-25.

Morrisey,E.E., Ip,H.S., Lu,M.M. and Parmacek,M.S. (1996) GATA-6 - a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Dev. Biol.* **177**, 309-322.

Morrisey,E.E., Ip,H.S., Tang,Z. and Parmacek,M.S. (1997a) GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. *J. Biol. Chem.* **272**, 8515-8524.

Morrisey,E.E., Ip,H.S., Tang,Z.H., Lu,M.M. and Parmacek,M.S. (1997b) GATA-5 - a transcriptional activator expressed in a novel temporally and spatially-restricted pattern during embryonic development. *Dev. Biol.* **183**, 21-36.

Moss,J.B., McQuinn,T.C. and Schwartz,R.J. (1994) The avian cardiac alpha-actin promoter is regulated through a pair of complex elements composed of E boxes and serum response elements that bind both positive- and negative-acting factors. *J. Biol. Chem.* **269**, 12731-12740.

Murphy,A.M., Thompson,W.R., Peng,L.F. and Jones,L.,2nd. (1997) Regulation of the rat cardiac troponin I gene by the transcription factor GATA-4. *Biochem. J.* **322**, 393-401.

Murray,E.J. and Grosveld,F. (1987) Sites specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation mediated suppression. *EMBO J.* **6**, 2329-2335.

Myers,R.M., Rio,D.C., Robbins,A.K. and Tjian,R. (1981) SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. *Cell*, **25**, 373-384.

Nagy,L., Kao,H.Y., Chakravarti,D., Lin,R.J., Hassig,C.A., Ayer,D.E., Schreiber,S.L. and Evans,R.M. (1997) Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell*, **89**, 373-380.

Naidu,P.S., Ludolph,D.C., To,R.Q., Hinterberger,T.J. and Konieczny,S.F. (1995) Myogenin and MEF2 function synergistically to activate the MRF4 promoter during myogenesis. *Mol. Cell. Biol.* **15**, 2707-2718.

Nakajima,T., Uchida,C., Anderson,S.F., Parvin,J.D. and Montminy,M. (1997) Analysis of a cAMP-responsive activator reveals a two-component mechanism for transcriptional induction via signal-dependent factors. *Genes Dev.* **11**, 738-747.

Navankasattusas,S., Zhu,H., Garcia,A.V., Evans,S.M. and Chien,K.R. (1992) A ubiquitous factor (HF-1a) and a distinct muscle factor (HF-1b/MEF-2) form an E-box-independent pathway for cardiac muscle gene expression. *Mol. Cell. Biol.* **12**, 1469-1479.

Navankasattusas,S., Sawadogo,M., van Bilsen,M., Dang,C.V. and Chien,K.R. (1994) The basic helix-loop-helix protein upstream stimulating factor regulates the cardiac ventricular myosin light-chain 2 gene via independent cis regulatory elements. *Mol. Cell. Biol.* **14**, 7331-7339.

Nemer,M., Chamberland,M., Sirois,D., Argentin,S., Drouin,J., Dixon,R.A., Zivin,R.A. and Condra,J.H. (1984) Gene structure of human cardiac hormone precursor, pro-natriodilatin. *Nature*, **312**, 654-656.

Newman,A.J. (1994) Pre-mrna splicing. *Curr. Genetics Dev.* **4**, 298-304: 20.

Newman,C.S., Grow,M.W., Cleaver,O., Chia,F. and Krieg,P. (1997) Xbap, a vertebrate gene related to bagpipe, is expressed in developing craniofacial structures and in anterior gut muscle. *Dev. Biol.* **181**, 223-233.

Nikolov,D.B., Hu,S.H., Lin,J., Gasch,A., Hoffmann,A., Horikoshi,M., Chua,N.H., Roeder,R.G. and Burley,S.K. (1992) Crystal structure of tfiid tata-box binding protein. *Nature*, **360**, 40-6: 6.

Nonet,M., Sweetser,D. and Young,R.A. (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell*, **50**, 909-915.

Norman,C., Runswick,M., Pollock,R. and Treisman,R. (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell*, **55**, 989-1003.

Nudler,E., Goldfarb,A. and Kashlev,M. (1994) Discontinuous mechanism of transcription elongation. *Science*, **265**, 793-796.

Okkema,P.G. and Fire,A. (1994) The Caenorhabditis elegans NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. *Development*, **120**, 2175-2186.

Olsen,G.J. and Woese,C.R. (1997) Archeal genomics: an overview. *Cell*, **89**, 991-994.

Olson,E.N. (1992) Regulatory mechanisms for skeletal muscle differentiation and their relevance to gene expression in the heart. *Trends Cardiovasc. Med.* **2**, 163-170.

Olson,E.N. and Srivastava,D. (1996) Molecular pathways controlling heart development. *Science*, **272**, 671-676.

Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J. and Gronenborn,A.M. (1993a) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438-446.

Omichinski,J.G., Trainor,C., Evans,T., Gronenborn,A.M., Clore,G.M. and Felsenfeld,G. (1993b) A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. *Proc. Natl. Acad. Sci. USA*, **90**, 1676-1680.

Orkin,S.H. (1990) Globin gene regulation and switching: circa 199. *Cell*, **63**, 665-72: 19.

Osada,H., Grutz,G., Axelson,H., Forster,A. and Rabbitts,T.H. (1995) Association of erythroid transcription factors: complexes involving the lim protein rbtn2 and the zinc-finger protein gata1. *Proc. Natl. Acad. Sci. USA*, **92**, 9585-9: 2.

Ossipow,V., Tassan,J.P., Nigg,E.A. and Schibler,U. (1995) A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell*, **83**, 137-146.

Otting,G., Qian,Y.Q., Billeter,M., Muller,M., Affolter,M., Gehring,W.J. and Wuthrich,K. (1990) Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085-3092.

Pabst,O., Schneider,A., Brand,T. and Arnold,H.H. (1997) The mouse Nkx2-3 homeodomain gene is expressed in gut mesenchyme during pre- and postnatal mouse development. *Dev. Dynamics*, **209**, 29-35.

Padgett,R.A., Konarska,M.M., Grabowski,P.J., Hardy,S.F. and Sharp,P.A. (1984) Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, **225**, 898-903.

Pandolfi,P.P., Roth,M.E., Karis,A., Leonard,M.W., Dzierzak,E., Grosveld,F.G., Engel,J.D. and Lindenbaum,M.H. (1995) Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis [see comments]. *Nature Genetics*, **11**, 40-44.

Pari,G., Jardine,K. and McBurney,M.W. (1991) Multiple CArG boxes in the human cardiac actin gene promoter required for expression in embryonic cardiac muscle cells developing in vitro from embryonal carcinoma cells. *Mol. Cell. Biol.* **11**, 4796-4803.

Park,M., Wu,X., Golden,K., Axelrod,J.D. and Bodmer,R. (1996) The wingless signaling pathway is directly involved in drosophila heart development. *Dev. Biol.* **177**, 104-16:22.

Parker,C.S. and Topol,J. (1984) A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. *Cell*, **36**, 357-369.

Parmacek,M.S., Vora,A.J., Shen,T., Barr,E., Jung,F. and Leiden,J.M. (1992) Identification and characterization of a cardiac-specific transcriptional regulatory element in the slow/cardiac troponin C gene. *Mol. Cell. Biol.* **12**, 1967-1976.

Paulson,J.R. and Laemmli,U.K. (1977) The structure of histone-depleted metaphase chromosomes. *Cell*, **12**, 817-828.

Pause,A., Belsham,G.J., Gingras,A.C., Donze,O., Lin,T.A., Lawrence,J.C.,Jr. and Sonenberg,N. (1994a) Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature*, **371**, 762-7: 18.

Pause,A., Methot,N., Svitkin,Y., Merrick,W.C. and Sonenberg,N. (1994b) Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation. *EMBO J.* **13**, 1205-1215.

Pazin,M.J. and Kadonaga,J.T. (1997) Whats up and down with histone deacetylation and transcription. *Cell*, **89**, 325-328.

Pedersen,R.A., Wu,K. and Balakier,H. (1986) Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection. *Dev. Biol.* **117**, 581-595.

Peltenburg,L.T.C. and Murre,C. (1997) Specific residues in the PBX homeodomain differentially modulate the DNA-binding activity of Hox and engrailed proteins. *Development*, **124**, 1089-1098.

Pennica,D., King,K.L., Shaw,K.J., Luis,E., Rullamas,J., Luoh,S.M., Darbonne,W.C., Knutzon,D.S., Yen,R., Chien,K.R. and et al. (1995) Expression cloning of cardiotrophin 1, a cytokine that induces cardiac myocyte hypertrophy. *Proc. Natl. Acad. Sci. USA*, **92**, 1142-1146.

Peterson,C.L., Dingwall,A. and Scott,M.P. (1994) Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement [see comments]. *Proc. Natl. Acad. Sci. USA*, **91**, 2905-2908.

Peterson,C.L. and Herskowitz,I. (1992) Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*, **68**, 573-583.

Peterson,C.L. and Tamkun,J.W. (1995) The SWI-SNF complex: a chromatin remodeling machine?. [Review]. *Trends Biochem. Sci.* **20**, 143-146.

Peterson,M.G., Inostroza,J., Maxon,M.E., Flores,O., Admon,A., Reinberg,D. and Tjian,R. (1991) Structure and functional properties of human general transcription factor iie. *Nature*, **354**, 369-73: 14.

Pevny,L., Simon,M.C., Robertson,E., Klein,W.H., Tsai,S.F., D'Agati,V., Orkin,S.H. and Costantini,F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257-260.

Pevny,L., Lin,C.S., D'Agati,V., Simon,M.C., Orkin,S.H. and Costantini,F. (1995) Development of hematopoietic cells lacking transcription factor GATA-1. *Development*, **121**, 163-172.

Peyrieras,N., Hyafil,F., Louvard,D., Ploegh,H.L. and Jacob,F. (1983) Uvomorulin: a nonintegral membrane protein of early mouse embryo. *Proc. Natl. Acad. Sci. USA*, **80**, 6274-6277.

Phelan,M.L., Rambaldi,I. and Featherstone,M.S. (1995) Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell. Biol.* **15**, 3989-3997.

Pikielny,C.W., Rymond,B.C. and Rosbash,M. (1986) Electrophoresis of ribonucleoproteins reveals an ordered assembly pathway of yeast splicing complexes. *Nature*, **324**, 341-345.

Pollock,R. and Treisman,R. (1991) Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev.* **5**, 2327-2341.

Postigo,A.A. and Dean,D.C. (1997) ZEB, a vertebrate homolog of Drosophila Zfh-1, is a negative regulator of muscle differentiation. *EMBO J.* **16**, 3935-3943.

Price,M., Lazzaro,D., Pohl,T., Mattei,M.G., Ruther,U., Olivo,J.C., Duboule,D. and Di Lauro,R. (1992) Regional expression of the homeobox gene Nkx-2.2 in the developing mammalian forebrain. *Neuron*, **8**, 241-255.

Ptashne,M. (1988) How eukaryotic transcriptional activators work. *Nature*, **335**, 683-689.

Ptashne,M. and Gann,A. (1997) Transcriptional activation by recruitment. *Nature*, **386**, 569-77: 2.

Puelles,L. and Rubenstein,J.L.R. (1993) Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization. *Trends Neurosci.* **16**, 472-479.

Pugh,B.F. and Tjian,R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**, 1935-1945.

Qureshi,S.A., Khoo,B., Baumann,P. and Jackson,S.P. (1995) Molecular cloning of the transcription factor tfiib homolog from sulfolobus shibatae. *Proc. Natl. Acad. Sci. USA*, **92**, 6077-81: 13.

Randall, D.J. and Davie, P.S. (1980) The hearts of urochordates and cephalochordates. In Bourne,G.H. *Hearts and heart-like organs* - Academic Press, New York, pp. 41-58.

Reaume,A.G., de Sousa,P.A., Kulkarni,S., Langille,B.L., Zhu,D., Davies,T.C., Juneja,S.C., Kidder,G.M. and Rossant,J. (1995) Cardiac malformation in neonatal mice lacking connexin43. *Science*, **267**, 1831-1834.

Reinberg,D., Horikoshi,M. and Roeder,R.G. (1987) Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *J. Biol. Chem.* **262**, 3322-3330.

Reinberg,D. and Roeder,R.G. (1987) Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. *J. Biol. Chem.* **262**, 3310-3321.

Reines,D., Chamberlin,M.J. and Kane,C.M. (1989) Transcription elongation factor SII (TFIIS) enables RNA polymerase II to elongate through a block to transcription in a human gene in vitro. *J. Biol. Chem.* **264**, 10799-10809.

Reines,D., Ghanouni,P., Li,Q.Q. and Mote,J.,Jr. (1992) The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *J. Biol. Chem.* **267**, 15516-15522.

Reines,D., Conaway,J.W. and Conaway,R.C. (1996) The RNA polymerase II general elongation factors. *Trends Biochem. Sci.* **21**, 351-355.

Reynolds,K., Zimmer,A.M. and Zimmer,A. (1996) Regulation of RAR beta 2 mRNA expression: evidence for an inhibitory peptide encoded in the 5'-untranslated region. *J. Cell Biol.* **134**, 827-835.

Rio,D.C. (1993) Splicing of pre-mrna: mechanism, regulation and role in development. *Curr. Genetics Dev.* **3**, 574-84: 25.

Rivier,D.H. and Pillus,L. (1994) Silencing speaks up. *Cell*, **76**, 963-6:11.

Robb,L., Lyons,I., Li,R., Hartley,L., Kontgen,F., Harvey,R.P., Metcalf,D. and Begley,C.G. (1995) Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc. Natl. Acad. Sci. USA*, **92**, 7075-7079.

Roberts,S.G., Ha,I., Maldonado,E., Reinberg,D. and Green,M.R. (1993) Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. *Nature*, **363**, 741-744.

Rosenquist,G.C. (1966) A radioautographic study of labeled grafts in the chick blastoderm. Development from primitive-streak stages to stage 12. *Carnegie Inst. Wash. Contrib. Embryol.* **38**, 71-110.

Rosenquist,G.C. (1972) Endoderm movements in the chick embryo between the early short streak and head process stages. *J. Exp. Zool.* **180**, 95-103.

Rosenzweig,A. and Seidman,C.E. (1991) Atrial natriuretic factor and related peptide hormones. *Annu. Rev. Biochem.* **60**, 229-255.

Ross,R.S., Navankasattusas,S., Harvey,R.P. and Chien,K.R. (1996) An HF-1a/HF1b/MEF-2 combinatorial element confers cardiac ventricular specificity and established an anterior-posterior gradient of expression. *Development*, **122**, 1799-1809.

Rossant,J. (1996) Mouse mutants and cardiac development - New molecular insights into cardiogenesis [Review]. *Circ. Res.* **78**, 349-353.

Rowlands,T., Baumann,P. and Jackson,S.P. (1994) The tata-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science*, **264**, 1326-9: 17.

Roy,R., Adamczewski,J.P., Seroz,T., Vermeulen,W., Tassan,J.P., Schaeffer,L., Nigg,E.A., Hoeijmakers,J.H. and Egly,J.M. (1994) The mo15 cell cycle kinase is associated with the tfiih transcription-dna repair factor. *Cell*, **79**, 1093-101: 18.

Rudnicki,M.A., Braun,T., Hinuma,S. and Jaenisch,R. (1992) Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell*, **71**, 383-390.

Rudnicki, M.A. and McBurney, M.W. 1987. *Teratocarcinomas and embryonic stem cells, a practical approach*, IRL Press, Oxford:pp. 19-49.

Ruiz-Echevarria,M.J., Czaplinski,K. and Peltz,S.W. (1996) Making sense of nonsense in yeast. *Trends Biochem. Sci.* **21**, 433-8: 2.

Ruppert,S., Wang,E.H. and Tjian,R. (1993) Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature*, **362**, 175-179.

Ruskin,B., Krainer,A.R., Maniatis,T. and Green,M.R. (1984) Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, **38**, 317-331.

Sachs,A.B., Sarnow,P. and Hentze,M.W. (1997) Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell*, **13**, 831-838.

Sadoshima,J., Xu,Y., Slayter,H.S. and Izumo,S. (1993) Autocrine release of angiotensin II mediates stretch-induced hypertrophy of cardiac myocytes in vitro. *Cell*, **75**, 977-984.

Samuel,C.E. (1993) The eIF-2 alpha protein kinases, regulators of translation in eukaryotes from yeasts to humans. *J. Biol. Chem.* **268**, 7603-7606.

Sartorelli,V., Webster,K.A. and Kedes,L. (1990) Muscle-specific expression of the cardiac alpha-acting gene requires MyoD1, CArG-box binding factor, and Sp1. *Genes Dev.* **4**, 1811-1822.

Sater,A.K. and Jacobson,A.G. (1990) The restriction of the heart morphogenetic field in xenopus laevis. *Dev. Biol.* **140**, 328-36:3.

Sauer,F., Wassarman,D.A., Rubin,G.M. and Tjian,R. (1996) TAF(II)s mediate activation of transcription in the Drosophila embryo. *Cell*, **87**, 1271-1284.

Sauvageau,G., Thorsteinsdottir,U., Eaves,C.J., Lawrence,H.J., Largman,C., Lansdorp,P.M. and Humphries,R.K. (1995) Overexpression of hoxb4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. *Genes Dev.* **9**, 1753-65: 2.

Schatteman,G.C., Motley,S.T., Effmann,E.L. and Bowen-Pope,D.F. (1995) Platelet-derived growth factor receptor alpha subunit deleted Patch mouse exhibits severe cardiovascular dysmorphogenesis. *Teratology*, **51**, 351-366.

Schultheiss,T.M., Burch,J.B. and Lassar,A.B. (1997) A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* **11**, 451-62:26.

Schumacher,A. and Magnuson,T. (1997) Murine Polycomb- and trithorax-group genes regulate homeotic pathways and beyond. *Trends Genet.* **13**, 167-170.

Scott,M.P., Tamkun,J.W. and Hartzell,G.W.I. (1989) The structure and function of the homeodomain. *Biochem. Biophys. Acta*, **989**, 25-48.

Seidman,C.E., Schmidt,E.V. and Seidman,J.G. (1991) cis-Dominance or rat atrial natriuretic factor gene regulatory sequences in transgenic mice. *Can. J. Physiol. Pharmacol.* **69**, 1486-1492.

Shieh,S.Y., Stellrecht,C.M. and Tsai,M.J. (1995) Molecular characterization of the rat insulin enhancer-binding complex 3b2. Cloning of a binding factor with putative helicase motifs. *J. Biol. Chem.* **270**, 21503-21508.

Shield,M.A., Haugen,H.S., Clegg,C.H. and Hauschka,S.D. (1996) E-box sites and a proximal regulatory region of the muscle creatine kinase gene differentially regulate expression in diverse skeletal muscles and cardiac muscle of transgenic mice. *Mol. Cell. Biol.* **16**, 5058-5068.

Shilatifard,A., Lane,W.S., Jackson,K.W., Conaway,R.C. and Conaway,J.W. (1996) An RNA polymerase II elongation factor encoded by the human ELL gene. *Science*, **271**, 1873-1876.

Shimell,M.J., Simon,J., Bender,W. and O'Connor,M.B. (1994) Enhancer point mutation results in a homeotic transformation in Drosophila. *Science*, **264**, 968-971.

Shore,P. and Sharrocks,A.D. (1995) The MADS-box family of transcription factors. *Eur. J. Biochem.* **229**, 1-13.

Shyu,A.B., Belasco,J.G. and Greenberg,M.E. (1991) Two distinct destabilizing elements in the c-fos message trigger deadenylation as a first step in rapid mRNA decay. *Genes Dev.* **5**, 221-231.

Simon,M.C. (1995) Gotta have GATA [news; comment]. [Review]. *Nature Genetics*, **11**, 9-11.

Simpson,P. (1985) Stimulation of hypertrophy of cultured neonatal rat heart cells through an alpha1-adrenergic receptor and induction of beating through an alpha1- and beta1-adrenergic receptor interaction: evidence for independent regulation of growth and beating. *Circ. Res.* **56**, 884-894.

SivaRaman,L., Reines,D. and Kane,C.M. (1990) Purified elongation factor SII is sufficient to promote read-through by purified RNA polymerase II at specific termination sites in the human histone H3.3 gene. *J. Biol. Chem.* **265**, 14554-14560.

Small,S., Kraut,R., Hoey,T., Warrior,R. and Levine,M. (1991) Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes Dev.* **5**, 827-839.

Small,S., Blair,A. and Levine,M. (1992) Regulation of even-skipped stripe 2 in the drosophila embryo. *EMBO J.* **11**, 4047-4057.

Smith,J.L., Gesteland,K.M. and Schoenwolf,G.C. (1994) Prospective fate map of the mouse primitive streak at 7.5 days of gestation. *Dev. Dynamics*, **201**, 279-289.

Sonenberg,N. (1993) Translation factors as effectors of cell growth and tumorigenesis. *Curr. Opin. Cell Biol.* **5**, 955-60: 23.

Soudais,C., Wilson,D.B., Simon,M.C., Ip,H.S., Parmacek,M.S. and Leiden,J.M. (1994) Targeted disruption of the GATA-4 gene in embryonic stem cells abolishes cardiac myocyte development in differentiating embryoid bodies. *Circulation*, **90**, I78

Sprenkle,A.B., Murray,S.F. and Glembotski,C.C. (1995) Involvement of multiple cis elements in basal- and alpha-adrenergic agonist-inducible atrial natriuretic factor transcription - roles for serum response elements and an sp-1-like element. *Circ. Res.* **77**, 1060-1069:

Springhorn,J.P., Ellingsen,O., Berger,H.J., Kelly,R.A. and Smith,T.W. (1992) Transcriptional regulation in cardiac muscle. Coordinate expression of Id with a neonatal

phenotype during development and following a hypertrophic stimulus in adult rat ventricular myocytes in vitro. *J. Biol. Chem.* **267**, 14360-14365.

Srivastava,D., Cserjesi,P. and Olson,E.N. (1995) A subclass of bHLH proteins required for cardiac morphogenesis. *Science*, **270**, 1995-1999.

Srivastava,D., Thomas,T., Lin,Q., Kirby,M.L., Brown,D. and Olson,E.N. (1997) Regulation of cardiac mesodermal and neural crest development by the bhlh transcription factor, dhand. *Nature Genetics*, **16**, 154-160: 3.

Staehling-Hampton,K., Hoffmann,F.M., Baylies,M.K., Rushton,E. and Bate,M. (1994) dpp induces mesodermal gene expression in Drosophila. *Nature*, **372**, 783-786.

Stainier,D.Y., Weinstein,B.M., Detrich,H.W.,3rd, Zon,L.I. and Fishman,M.C. (1995) Cloche, an early acting zebrafish gene, is required by both the endothelial and hematopoietic lineages. *Development*, **121**, 3141-3150.

Stainier,D.Y., Fouquet,B., Chen,J.N., Warren,K.S., Weinstein,B.M., Meiler,S.E., Mohideen,M.A., Neuhauss,S.C., Solnica-Krezel,L., Schier,A.F., Zwartkruis,F., Stemple,D.L., Malicki,J., Driever,W. and Fishman,M.C. (1996) Mutations affecting the formation and function of the cardiovascular system in the zebrafish embryo. *Development*, **123**, 285-292.

Stainier,D.Y.R., Lee,R.K. and Fishman,M.C. (1993) Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. *Development*, **119**, 31-40.

Stainier,D.Y.R. and Fishman,M.C. (1992) Patterning the zebrafish heart tube: acquisition of anteroposterior polarity. *Dev. Biol.* **153**, 91-101.

Stainier,D.Y.R. and Fishman,M.C. (1994) The zebrafish as a model system to study cardiovascular development. *Trends Cardiovasc. Med.* **4**, 207-212: 3.

Stanojevic,D., Small,S. and Levine,M. (1991) Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science*, **254**, 1385-1387.

Stargell,L.A. and Struhl,K. (1995) The TBP-TFIIA interaction in the response to acidic activators in vivo. *Science*, **269**, 75-78.

Steinmetz,E.J. (1997) Pre-mrna processing and the ctd of rna polymerase ii - the tail that wags the dog. *Cell*, **89**, 491-494: 6.

Steinmetz,E.J. and Brow,D.A. (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.* **16**, 6993-7003.

Stephens,R.M. and Schneider,T.D. (1992) Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.* **228**, 1124-1136.

Stewart,A.F.R., Larkin,S.B., Farrance,I.K.G., Mar,J.H., Hall,D.E. and Ordahl,C.P. (1994) Muscle-enriched TEF-1 isoforms bind M-CAT elements from muscle-specific promoters and differentially activate transcription. *J. Biol. Chem.* **269**, 3147-3150.

Stief,A., Winter,D.M., Stratling,W.H. and Sippel,A.E. (1989) A nuclear dna attachment element mediates elevated and position-independent gene activity. *Nature*, **341**, 343-5:

Struhl,K. (1994) Duality of TBP, the universal transcription factor. *Science*, **263**, 1103-1104.

Struhl,K. (1996) Chromatin structure and RNA polymerase II connection - Implications for transcription [Review]. *Cell*, **84**, 179-182.

Subramaniam,A., Jones,W.K., Gulick,J., Wert,S., Neumann,J. and Robbins,J. (1991) Tissue-specific regulation of the alpha-myosin heavy chain gene promoter in transgenic mice. *J. Biol. Chem.* **266**, 24613-24620.

Sucov,H.M., Dyson,E., Gumeringer,C.L., Price,J., Chien,K.R. and Evans,R.M. (1994) RXRα mutant mice establish a genetic basis for vitamin A signaling in heart morphogenesis. *Genes Dev.* **8**, 1007-1018.

Svaren,J. and Horz,W. (1993) Histones, nucleosomes and transcription. [Review] [73 refs]. *Curr. Genetics Dev.* **3**, 219-225.

Svejstrup,J.Q., Vichi,P. and Egly,J.M. (1996) The multiple roles of transcription/repair factor tfiih. *Trends Biochem. Sci.* **21**, 346-50: 3.

Tam,P.P.L., Parameswaran,M., Kinder,S.J. and Weinberger,R.P. (1997) The allocation of epiblast cells to the embryonic heart and other mesodermal lineages - the role of ingression and tissue movement during gastrulation. *Development*, **124**, 1631-1642: 4.

Tamkun,J.W., Deuring,R., Scott,M.P., Kissinger,M., Pattatucci,A.M., Kaufman,T.C. and Kennison,J.A. (1992) brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell*, **68**, 561-572.

Tate,P.H. and Bird,A.P. (1993) Effects of dna methylation on dna-binding proteins and gene expression. *Curr. Genetics Dev.* **3**, 226-31: 26.

Taylor,M.V. (1991) A family of muscle gene promoter element (CArG) binding activities in Xenopus embryos: CArG/SRE discrimination and distribution during myogenesis. *Nucleic Acids Res.* **19**, 2669-2675.

Thanos,D. (1996) Mechanisms of transcriptional synergism of eukaryotic genes. The interferon-beta paradigm. *Hypertension*, **27**, 1025-1029.

Thanos,D. and Maniatis,T. (1995) Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell*, **83**, 1091-1100.

Thompson,C.M., Koleske,A.J., Chao,D.M. and Young,R.A. (1993) A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell*, **73**, 1361-1375.

Thompson,W.R., Nadal-Ginard,B. and Mahdavi,V. (1991) A MyoD1-independent muscle-specific enhancer controls the expression of the β-myosin heavy chain gene in skeletal and cardiac muscle cells. *J. Biol. Chem.* **266**, 22678-22688.

Thuerauf,D.J. and Glembotski,C.C. (1997) Differential effects of protein kinase C, Ras, and Raf-1 kinase on the induction of the cardiac B-type natriuretic peptide gene through a critical promoter-proximal M-CAT element. *J. Biol. Chem.* **272**, 7464-7472.

Ting,C.N., Olson,M.C., Barton,K.P. and Leiden,J.M. (1996) Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature*, **384**, 474-478.

Tjian,R. and Maniatis,T. (1994) Transcriptional activation: a complex puzzle with few easy pieces. [Review]. *Cell*, **77**, 5-8.

Tonissen,K.F., Drysdale,T.A., Lints,T.J., Harvey,R.P. and Krieg,P.A. (1994) XNkx-2.5, a Xenopus gene related to Nkx-2.5 and tinman: evidence for a conserved role in cardiac development. *Dev. Biol.* **162**, 325-328.

Tournamille,C., Colin,Y., Cartron,J.P. and Le Van Kim,C. (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nature Genetics*, **10**, 224-228.

Trainor,C.D., Omichinski,J.G., Vandergon,T.L., Gronenborn,A.M., Clore,G.M. and Felsenfeld,G. (1996) A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol. Cell. Biol.* **16**, 2238-2247.

Trotta,C.R., Miao,F., Arn,E.A., Stevens,S.W., Ho,C.K., Rauhut,R. and Ableson,J.N. (1997) The yeast tRNA splicing endonuclease: A tetrameric enzyme with two active sites subunits homologous to the archeal tRNA endonuclease. *Cell*, **89**, 849-858.

Tsai,F.Y., Keller,G., Kuo,F.C., Weiss,M., Chen,J., Rosenblatt,M., Alt,F.W. and Orkin,S.H. (1994) An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, **371**, 221-226.

Tsai,S.F., Martin,D.I., Zon,L.I., D'Andrea,A.D., WOng,G.G. and Orkin,S.H. (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, **339**, 446-451.

Tsang,A.P., Visvader,J.E., Turner,C.A., Fujiwara,Y., Yu,C., Weiss,M.J., Crosslely,M. and Orkin,S.H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*, **20**, 109-119.

Tsao,D.H., Gruschus,J.M., Wang,L.H., Nirenberg,M. and Ferretti,J.A. (1995) The three-dimensional solution structure of the NK-2 homeodomain from Drosophila. *J. Mol. Biol.* **251**, 297-307.

Tsuda,T., Philp,N., Zile,M.H. and Linask,K.K. (1996) Left-right asymmetric localization of flectin in the extracellular matrix during heart looping. *Dev. Biol.* **173**, 39-50.

Tsukiyama,T., Daniel,C., Tamkun,J. and Wu,C. (1995) ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell*, **83**, 1021-1026.

Tsukiyama,T. and Wu,C. (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell*, **83**, 1011-1020.

Van Dijk,M.A. and Murre,C. (1994) extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell*, **78**, 617-624.

van Holde,K.E., Lohr,D.E. and Robert,C. (1992) What happens to nucleosomes during transcription? *J. Biol. Chem.* **267**, 2837-40: 4.

Verrijzer,C.P., Chen,J.L., Yokomori,K. and Tjian,R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115-1125.

Vincent,C.K., Gualberto,A., Patel,C.V. and Walsh,K. (1993) Different regulatory sequences control creatine kinase-M gene expression in directly injected skeletal and cardiac muscle. *Mol. Cell. Biol.* **13**, 1264-1272.

Visvader,J.E., Crossley,M., Hill,J., Orkin,S.H. and Adams,J.M. (1995) The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Mol. Cell. Biol.* **15**, 634-641.

Wahle,E. and Keller,W. (1996) The biochemistry of polyadenylation. *Trends Biochem. Sci.* **21**, 247-50: 9.

Walker,S.S., Reese,J.C., Apone,L.M. and Green,M.R. (1996) Transcription activation in cells lacking TAFIIS. *Nature*, **383**, 185-188.

Walmsley,M.E., Guille,M.J., Bertwistle,D., Smith,J.C., Pizzey,J.A. and Patient,R.K. (1994) Negative control of Xenopus GATA-2 by activin and noggin with eventual expression in precursors of the ventral blood islands. *Development*, **120**, 2519-2529.

Wang,C. and Song,B. (1996) Cell-type-specific expression of the platelet-derived growth factor alpha receptor: a role for GATA-binding protein. *Mol. Cell. Biol.* **16**, 712-723.

Wang,C.Y., Mayo,M.W. and Baldwin Jr.,A.S. (1996) TNF-and cancer therapy-induced apoptosis: potentiation by inhibition of NF-kB. *Science*, **274**, 784-787.

Weinzierl,R.O., Dynlacht,B.D. and Tjian,R. (1993) Largest subunit of Drosophila transcription factor IID directs assembly of a complex containing TBP and a coactivator. *Nature*, **362**, 511-517.

Weis,L. and Reinberg,D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexe. *FASEB J.* **6**, 3300-3309.

Weiss,M.J., Keller,G. and Orkin,S.H. (1994) Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev.* **8**, 1184-1197.

Weiss,M.J., Yu,C.N. and Orkin,S.H. (1997) Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol.* **17**, 1642-1651.

Weiss,M.J. and Orkin,S.H. (1995) Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc. Natl. Acad. Sci. USA*, **92**, 9623-9627.

Weng,Y., Czaplinski,K. and Peltz,S.W. (1996a) Identification and characterization of mutations in the upf1 gene that affect nonsense suppression and the formation of the upf protein complex but not mrna turnover. *Mol. Cell. Biol.* **16**, 5491-506: 10.

Weng,Y., Czaplinski,K. and Peltz,S.W. (1996b) Genetic and biochemical characterization of mutations in the atpase and helicase regions of the upf1 protein. *Mol. Cell. Biol.* **16**, 5477-90: 11.

Whyatt,D.J., deBoer,E. and Grosveld,F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *EMBO J.* **12**, 4993-5005.

Wilson,C.J., Chao,D.M., Imbalzano,A.N., Schnitzler,G.R., Kingston,R.E. and Young,R.A. (1996) RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell*, **84**, 235-244.

Wilson,D.B., Dorfman,D.M. and Orkin,S.H. (1990) A nonerythroid GATA-binding protein is required for function of the human preproendothelin-1 promoter in endothelial cells. *Mol. Cell. Biol.* **10**, 4854-4862.

Wilson,D.S. and Desplan,C. (1995) Homeodomain proteins. Cooperating to be different. *Curr. Biol.* **5**, 32-34.

Wilson,T. and Treisman,R. (1988) Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. *Nature*, **336**, 396-399.

Winston,F. and Carlson,M. (1992) Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. [Review]. *Trends Genet.* **8**, 387-391.

Wood,W.G. (1996) The complexities of beta globin gene regulation. *Trends Genet.* **12**, 204-6: 8.

Wu,X., Golden,K. and Bodmer,R. (1995) Heart development in Drosophila requires the segment polarity gene wingless. *Dev. Biol.* **169**, 619-628.

Wu,Y., Reece,R.J. and Ptashne,M. (1996) Quantitation of putative activator-target affinities predicts transcriptional activating potentials. *EMBO J.* **15**, 3951-63: 5.

Xie,X., Kokubo,T., Cohen,S.L., Mirza,U.A., Hoffmann,A., Chait,B.T., Roeder,R.G., Nakatani,Y. and Burley,S.K. (1996) Structural similarity between tafs and the heterotetrameric core of the histone octamer. *Nature*, **380**, 316-22: 9.

Yamagata,T., Nishida,J., Sakai,R., Tanaka,T., Honda,H., Hirano,N., Mano,H., Yazaki,Y. and Hirai,H. (1995) Of the GATA-binding proteins, only GATA-4 selectively regulates the human interleukin-5 gene promoter in interleukin-5-producing cells which express multiple GATA-binding proteins. *Mol. Cell. Biol.* **15**, 3830-3839.

Yamashita,S., Wada,K., Horikoshi,M., Gong,D.W., Kokubo,T., Hisatake,K., Yokotani,N., Malik,S., Roeder,R.G. and Nakatani,Y. (1992) Isolation and characterization of a cdna encoding drosophila transcription factor tfiib. *Proc. Natl. Acad. Sci. USA*, **89**, 2839-43: 8.

Yang,H.Y. and Evans,T. (1992) Distinct roles for the two cGATA-1 finger domains. *Mol. Cell. Biol.* **12**, 4562-4570.

Yankulov,K., Blau,J., Purton,T., Roberts,S. and Bentley,D.L. (1994) Transcriptional elongation by rna polymerase ii is stimulated by transactivators. *Cell*, **77**, 749-59:10.

Yokoyama,T., Copeland,N.G., Jenkins,N.A., Montgomery,C.A., Elder,F.F. and Overbeek,P.A. (1993) Reversal of left-right asymmetry: a situs inversus mutation. *Science*, **260**, 679-682.

Yost,H.J. (1992) Regulation of vertebrate left-right asymmetries by extracellular matrix. *Nature*, **357**, 158-161.

Yost,H.J. (1995) Vertebrate left-right development. *Cell*, **82**, 689-692.

Young,R.A. (1991) RNA polymerase II. *Annu. Rev. Biochem.* **60**, 689-715.

Yu,Y., Li,W., Su,K., Yussa,M., Han,W., Perrimon,N. and Pick,L. (1997) The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. *Nature*, **385**, 552-555.

Yu,Y.T., Breitbart,R.E., Smoot,L.B., Lee,Y., Mahdavi,V. and Nadal-Ginard,B. (1992) Human myocyte-specific enhancer factor 2 comprises a group of tissue-restricted MADS box transcription factors. *Genes Dev.* **6**, 1783-1798.

Yuan,S., Darnell,D.K. and Schoenwolf,G.C. (1995) Mesodermal patterning during avian gastrulation and neurulation: experimental induction of notochord from non-notochordal precursor cells. *Dev. Genet.* **17**, 38-54.

Yutzey,K.E., Rhee,J.T. and Bader,D. (1994) Expression of the atrial-specific myosin heavy chain AMHC1 and the establishment of anteroposterior polarity in the developing chicken heart. *Development*, **120**, 871-883.

Yutzey,K.E. and Bader,D. (1995) Diversification of cardiomyogenic cell lineages during early heart development. *Circ. Res.* **77**, 216-219.

Zamore,P.D., Patton,J.G. and Green,M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355**, 609-614.

Zawel,L. and Reinberg,D. (1993) Initiation of transcription by RNA polymerase II: a multi-step process. *Prog. Nucleic Acid. Res. & . Mol. Biol.* **44**, 67-108.

Zhang,H., Catron,K.M. and Abate-Shen,C. (1996) A role for the Msx-1 homeodomain in transcriptional regulation: residues in the N-terminal arm mediate TATA binding protein interaction and transcriptional repression. *Proc. Natl. Acad. Sci. USA*, **93**, 1764-1769.

Zhang,H. and Bradley,A. (1996) Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development*, **122**, 2977-2986.

Zhang,Y., Iratni,R., Erdjumentbromage,H., Tempst,P. and Reinberg,D. (1997) Histone deacetylases and sap18, a novel polypeptide, are components of a human sin3 complex. *Cell*, **89**, 357-364: 8.

Zhou,Q., Boyer,T.G. and Berk,A.J. (1993) Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. *Genes Dev.* **7**, 180-187.

Zhu,H., Garcia,A.V., Ross,R.S., Evans,S.M. and Chien,K.R. (1991) A conserved 28-base-pair element (HF-1) in the rat cardiac myosin light-chain-2 gene confers cardiac-specific and alpha-adrenergic-inducible expression in cultured neonatal rat myocardial cells. *Mol. Cell. Biol.* **11**, 2273-2281.