1 INTRODUCTION

1.1 Transcription

In the central dogma of molecular biology the genetic information stored in DNA produces RNA, which in turn produces protein. Hence, the genetic information inherited by each individual as DNA (genotype) is converted into protein (phenotype) by first producing RNA. This process of transcription is therefore an essential element in gene expression. Incorrect execution of this step will obviously lead to errors in all following processes and is therefore a critical step for the correct formation of the phenotype.

The central role for transcription in the process of gene expression also makes it an attractive control point for the regulation of cell-type specific expression. Although post-transcriptional regulation mechanisms such as RNA splicing or RNA stability pathways can also play a role in the regulation of gene expression the major control point lies at the level of transcription. Indeed, if a protein is produced only in a particular cell type or at a developmental stage this is achieved by control processes ensuring that the gene is only transcribed in those cells or in response to developmental signals.

The central role for transcription in the process of gene expression and the formation of the phenotype has led to extensive study of this process. At first, only the sequences in the direct vicinity of the transcribed genes were investigated and the proteins binding to these, which are called transcription factors (TFs). It is now clear that the ability of TFs to bind to DNA sequences and thereby to regulate gene expression is highly influenced by the DNA accessibility. In eukaryotic cells DNA is packaged by association with histone proteins into a structure called chromatin. High compaction of the chromatin renders the DNA inaccessible to TF binding and the transcriptional apparatus, silencing the genes in these regions.

Consequently two levels of transcriptional regulation can be distinguished: Factors influencing the compaction of chromatin and thereby rendering sequences more or less accessible. On a second level TF binding to the accessible sequences can recruit or inhibit formation of the transcriptional apparatus. Although this rough separation is commonly used and feasible, these two levels are interdependent: factors that influence the chromatin structure can also recruit TFs and vice versa.

1.1.1 Chromatin Structure and Transcriptional Regulation

1.1.1.1 Chromatin and Histones

Eukaryotic genomic DNA is compacted more than 10,000-fold by highly basic proteins known as histones; the result is a highly structured entity termed chromatin. The fundamental unit of chromatin, the nucleosome core particle consists of 147 bp of superhelical DNA wrapped in 1.75 turns around a histone octamer core[1]. A centrally located histone (H3/H4)2 tetramer is assembled with two histone dimers[2,3]. Consecutive nucleosomes line up, generating a fiber of 11 nm diameter, termed beads-on-a-string[4]. This is further compacted into a 30 nm fiber, at least partially by incorporation of the linker histone H1[5]. The mechanisms governing this process are as yet not fully understood, a recent review by Polo *et al.*[6] summarizes the current knowledge on chromatin assembly.

For a long time after the first isolation of nucleosomes, chromatin was thought to be a static entity providing the scaffolding for DNA. Over the last two decades it has become evident, that chromatin is a highly flexible environment. In a condensed state wherein the DNA is tightly wound around the histone octamer the accessibility for transcription factors and the transcriptional apparatus is limited. Therefore changes between condensed (heterochromatin) and structurally accessible states (euchromatin) are a key factor for the regulation of gene expression.

Histones are a highly conserved family of proteins[7,8] that are rich in the positively charged, basic amino acids lysine (K) and arginine (R). The four histone polypeptides each have a three-α-helix motif called histone-fold[9] in the central region of the protein and an unstructured N-terminal 'tail' that ranges from 16-44 amino acid residues and protrudes out of the nucleosome. These 'histone tails' make up ≈ 25-30% of the mass of individual histones and thus provide an exposed surface for potential interactions with other proteins. These 'histone tails' as well as the carboxy termini of histones H2A, H2B and H1 are susceptible to a wide variety of post-translational modifications: phosphorylation (serine and threonine)[10] , acetylation (K)[11,12] , methylation (K and R)[13,14]; ubiquitination (K)[15,16], sumoylation (K)[17], ADP ribosylation[18], glycosylation[19], biotinylation[20] and carbonylation[21]. A selection of sites of modifications is given in Figure 1-1. To indicate individual modifications in the following the *Brno nomenclature for histone modifications*[22] will be used. The histone is abbreviated by H followed by a number indicating the type of histone. This is followed by the amino acid(s) in one letter notation followed by their position counted from the N-terminus. At the end the type of modification is given. For details and examples please refer to Appendix 11.2 and the original publication. With the implementation of mass spectrometry in histone biology the number of known histone marks is expanding rapidly[23]. These biochemical modifications are referred to as part of the 'epigenetic information'.

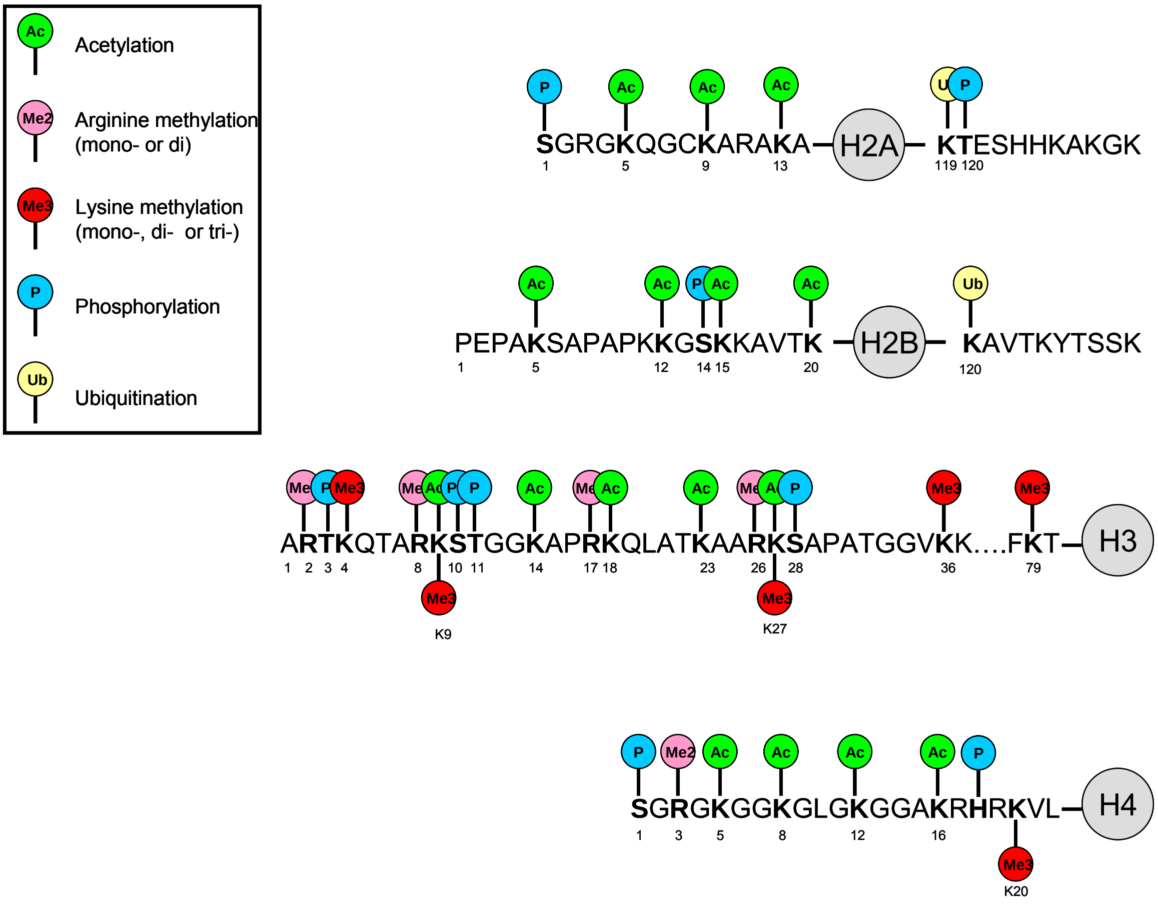
Ac Acetylation
Me2 Arginine methylation (mono- or di)
Me3 Lysine methylation (mono-, di- or tri-)
P Phosphorylation
Ub Ubiquitination
SGRGKQGCKARAKA H2A KTESHHKAKGK
1 5 9 13 119 120
PEPAKSAPAPKKGSKKAVTK H2B KAVTKYTSSK
1 5 12 14 15 20 120
ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKK....FKT H3
1 2 3 4 8 10 11 14 17 18 23 26 28 36 79
K9
K27
SGRGKGGKGLGKGGAKRHRKVL H4
1 3 5 8 12 16
K20

Figure 1-1. Modifications on core histone tails. Modifications associated with active transcription or unknown function are shown above the amino acid sequence, modifications associated with transcriptional repression are shown below.

Positions of Nucleosomes

To analyze the role of histone modifications in transcriptional regulation several elegant studies have been performed using yeast as model organism. Recently, a study by Segal *et al.*[24] has demonstrated that the positioning of nucleosomes is highly sequence-dependent. A model was developed, explaining ≈ 50% of the *in vivo* nucleosome positions. In the promoters of actively transcribed yeast genes, histone occupancy is reduced by approximately 20%[25], thereby making the DNA accessible for the transcriptional apparatus. These nucleosome-depleted regions (NDRs) are mainly associated with poly(dA-dT)stretches, which have been shown to destabilize nucleosome formation *in vitro.*[26].

Histone Acetylation

Acetylation of histones is a highly dynamic process with a half-life of as little as 15 to 300 min in eukaryotic cells[27,28]. More than three decades ago, Allfrey and colleagues found a correlation between increased histone acetylation and increased transcription[29]. Since then, using both genetic and biochemical approaches, several mechanisms by which histone acetylation and deacetylation regulate gene activity have been elucidated. The earliest findings demonstrated that the nucleosome is a repressor of transcription initiation *in vitro*[30] and *in vivo*[31]. This nucleosomal repression could be alleviated by acetylation of histone H4[32].

The acetylation is mediated by histone acetyl transferases (HATs)[33]. These enzymes are divided into three main families: GNAT, MYST, and CPB/p300. In general, they are capable of modifying more than one lysine residue, but a limited specificity is detected for some HATs. The reversal of acetylation is effected by histone deacetylases (HDACs)[33]. There are three distinct classes of these enzymes: the class I and II HDACs, and the class III NAD-dependent enzymes of the Sir family. Apparently the enzymes have no specificity for particular acetyl groups.

It is generally accepted, that acetylation of lysine residues on histone tails decreases the positive charge of the nucleosomes and thereby weakens the attraction between the negatively charged DNA backbone and the positive nucleosomes (Figure 1-2). It is thought that by this mechanism chromatin compaction is decreased[34] and the DNA is made accessible for transcription.

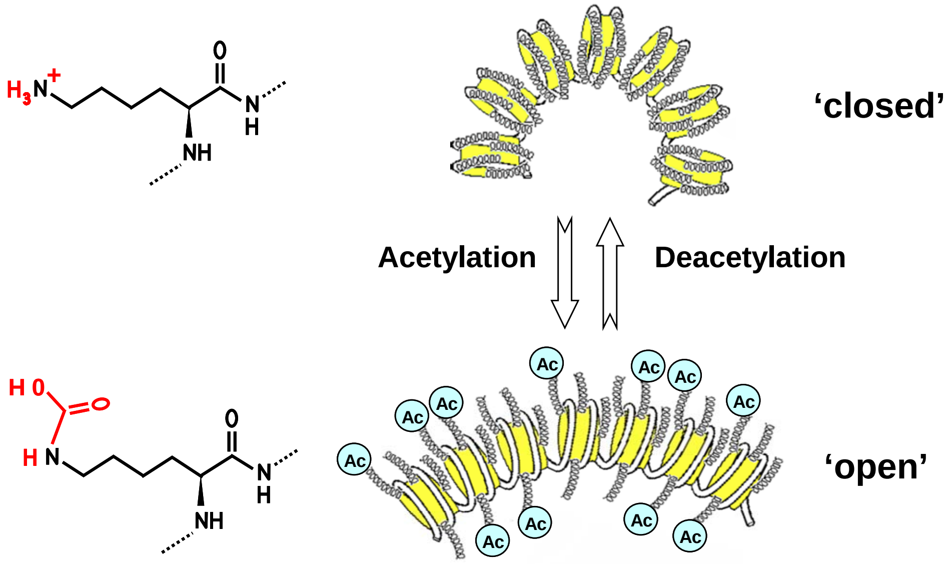

Figure 1-2. Histone tail acetylation diminishes the interaction between the histones and the DNA leading to an open chromatin structure. Enzymes acetylating histones are called histone acetyl transferases (HATs). The process can be reversed by histone deacetylases (HDACs).

Histone Methylation

Consistent with transient versus long-term epigenetic memory some histone modifications (e.g. phosphorylation and acetylation) are highly dynamic[27,28], whereas others (e.g. methylation) are more stable[35]. For a long time methylation of histones was thought to be different from acetylation or phosphorylation in that it was not reversible due to the high thermodynamic stability of the N-CH_3 bond. The methyl mark was considered to be static, and only to be removed by cleavage of histone N-termini[36], exchange with histone variants[37], or destabilization by oxidation or radical attack[38]. In yeast H3K4me3 persists over an hour after transcription ceases and therefore remains stable over several cell cycles[39]. In fact, only recently enzymes actively demethylating histones at specific residues could be identified[40-42].

Unlike acetylation which seems to be clearly linked to euchromatin, methylation provides a more complex picture. Depending on the position in the 'histone tail' methylation can function as activating or repressive mark (Figure 1-1). Furthermore, residues can be mono-, di- or trimethylated adding a further level of complexity. None of these three degrees of methylation will alter lysine's positive charge under conditions of physiological pH. As a result, it is unlikely that charge interactions are modulated by methylation. These marks appear to function mainly through methyl-lysine-binding proteins. Consistent with this view there are several lysine residues whose methylation has been described to be associated with repression (e.g. H3K9[43-50], H4K20[51] and possibly H3K27[52]) or with activation (e.g. H3K4[25], H3K4me2[53,54] H3K36[55,56] and H3K79[57]). In accordance with the plethora of functionalities associated with histone methylation marks an enormous number of enzymes controlling the placement (histone methyl transferases, HMTs) and removal (histone demethylases, HDMs) of methylation marks have been identified. A comprehensive overview has recently been published by Kouzarides[33].

1.1.1.2 Characteristics of Investigated Histone Modifications

H3ac & H4ac

First investigations into the functions of hyperacetylations of the tails of histone 3 (H3ac) and histone 4 (H4ac) were conducted using yeast as a model organism (Figure 1-3). In a first study[53] analyzing 95% of the yeast genome it was found that genes where these modifications occur near the transcription start sites (TSSs) generally have higher expression levels. The correlations between hyperacetylation and transcriptional activity were 21.1% in case of H3 and 13.4% in case of H4. Acetylations within coding regions appeared to be less influential. In a second study[58] achieving a higher resolution the major sites of acetylation could be mapped to be within 500 bp of the start codon ATG. This was subsequently confirmed in a study covering the entire yeast genome at nucleosomal resolution[25]; furthermore, the relative amounts of acetylation were found to reflect the transcription rates.

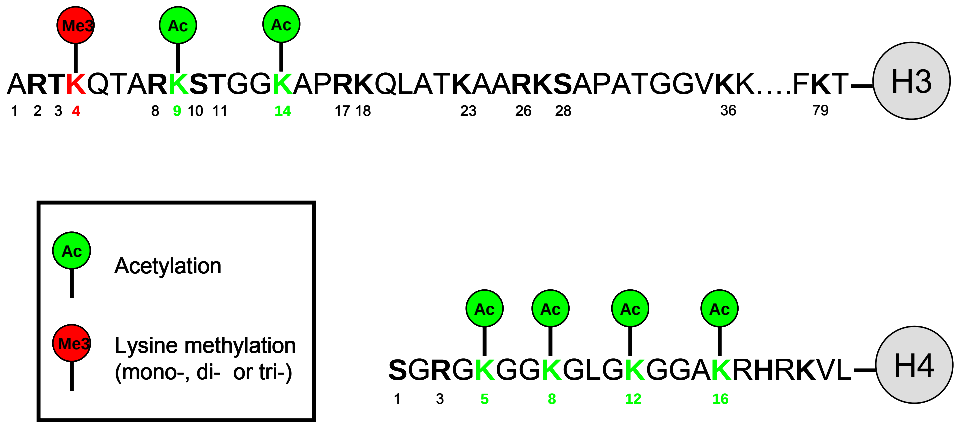

Figure 1-3. Shows histone tail modifications investigated in this study. This figure is the same as Figure 1-1 but only shows the investigated modifications.

A genome-wide approach investigating H3 and H4 acetylation in *Drosophila melanogaster* also reported these modifications to be associated with active genes. In human resting and activated T-cells a genome-wide sequencing approach revealed that H3K9K14ac occurs in gene-rich regions[59]. As proteins specifically recognizing different methylated residues are being discovered, a further function of acetylations as signaling marks is becoming likely.

H3K4me2 & H3K4me3

As for acetylation, first studies into the occurrence and function of histone tail methylations were carried out in yeast. Methylation of lysine 4 on histone 3 (Figure 1-3) was one of the first modifications that awakened general interest. H3K4me1, -me2, and -me3 occur at distinct positions relative to TSSs[25]. For H3K4me3 basically the same picture as for the hyperacetylations emerged and its occurrence was associated with active transcription[53,60]. However, H3K4me2 was distributed equally throughout the genic regions and H3K4me1 was most enriched at the end of genes. For these modification states no correlation with expression was observed.

In higher eukaryotes a first study mapping the occurrence of four modifications on human chromosomes 21 and 22 reported punctuate modification sites with H3K4me3 correlating with transcription start sites. However, as for several other modifications, this correlation does not seem to be as straightforward as in yeast: In a study covering 60.3 Mb of the mouse genome only 63% of H3K4me3 corresponded to TSS, whereas 80% of TSS were covered by H3K4me3 sites nearly all of which belonged to genes coding for transcription factors (93%)[61]. Similarly, a recent study mapping active promoters in the human genome[62] has found, that although nearly all promoters of actively transcribed genes are associated with hyperacetylated histone 3 (99%) and H3K4me2 (97%) quite a large amount of untranscribed genes, showing neither expression nor association with RNA Polymerase II still carry these epigenetic marks (20% and 31%, respectively). The function of these sites remains to be elucidated, single-gene studies point to a possible role in the regulation of development. In embryonic stem (ES) cells it was reported that developementally important loci are marked by overlapping domains of (activating) H3K4me3 and (repressive) H3K27me3[61]. This observation was confirmed by a study investigating resting and induced human T-cells[59].

In addition to a possible function in controlling developmentally inducible transcriptional changes it has been suggested that H3K4me2 and H3K4me3 may act as markers for recently transcribed genes thereby providing a mechanism of cellular memory[39,63]. This would be in accordance with the observation that during mitosis most proteins – but not histones – are displaced from the chromosomes[64]. This would point to histone methylation as a carrier of transcriptional memory in cell division[65]. In this case, the marks must be placed as a necessary consequence of transcription. However, other reports arrive at opposing conclusions.

A study investigating the di- and tri-methylation pattern of H3K4 at the β-globin locus in chicken embryo erythrocytes at two developmental stages shows a strong increase or decrease of methylation of genes becoming active or inactive, respectively, from day 5 to day 15. Significant levels of H3K4me are also present on inactive genes compared to heterochromatin. Therefore, the authors suggest that this modification implicates a 'poised' chromatin state[66]. In this model modifications are present prior to transcription suggesting a signaling function in development.

Recently, a new aspect has moved into the focus of attention regarding histone tail methylation. H3K4me2 and -me3 can recruit chromatin remodeling factors[67] and histone acetyl transferases[68]. Proteins containing chromodomains have long been known to indiscriminately bind to methylated lysine residues[69-71]. Recently, transcription factors containing PHD domains were reported to specifically bind to H3K4me3[72-76], indicating specific roles as signaling marks. In fact, the opposite scenario has now also been observed. The PHD finger protein BHC80 can only bind to the histone 3 tail in the absence of any methylation of lysine 4[77].

1.1.1.3 The Histone Code Hypothesis vs. Transcriptional Memory

The wide variety of histone modifications as well as the observation of specific effects associated with their occurrence led to the formulation of the *histone code hypothesis* by

Strahl and Allis in 2000[78]: "distinct histone modifications on one or more tails act sequentially or in combination to form a *histone code* that is read by other proteins to bring about distinct downstream events". According to this theory, the specific pattern of histone posttranslational modifications in a locus extends the information stored in the genetic code. The modification marks would be recognized by proteins containing specific domains such as the bromo- and chromo-domains, thereby specific processes such as transcriptional activation and repression, chromosome condensation or repair could be initiated. The overall mixture of histone modifications may be the same over large chromatin regions, but the distinct occurrence at specific nucleosomes could create local structures leading to functional subdomains. In summary the histone code hypothesis suggests that the modifications contain information which leads to functional read-outs and have major functions as signaling marks. This implies they are placed prior to transcription.

In contrast, the theory suggesting histone modifications to be carriers of transcriptional memory[39] indicates that the modification marks are placed during or after transcription. Liu *et al*[79] investigated twelve different histone modifications in yeast. In this study no evidence for a deterministic code with discrete states was found. The authors suggest that the continuous pattern of modifications they observe distinguishes only nucleosome positions (e.g. promoter vs. 3'-end of genes) and reflects transcription.

1.1.1.4 Histone-Modifying Enzymes & Disease

Epigenetic marks have long been implicated in disease, but unil now only DNA methylation was in the center of attention. Recently, malfunction of histone-modifying enzymes and aberrant histone modifications have been shown to play a major role in a wide variety of diseases including Lupus erythematosus[80], cardiovascular disease, imprinting and pediatric syndromes and reproductive disorders[81]. Aberrant transcription due to altered expression or mutation of genes that encode histone acetyltransferase (HAT) or histone deacetylase (HDAC) enzymes or their binding partners has been clearly linked to carcinogenesis. Consequently, histone deacetylase inhibitors are a new promising class of anticancer agents[82] that inhibit the proliferation of tumor cells in culture and *in vivo* by inducing cell-cycle arrest, terminal differentiation, and/or apoptosis. The role of specific HATs and HDACs in tumor formation[83] as well as the efforts being made to develop inhibitors of these enzymes have recently been extensively reviewed[84,85].

1.1.1.5 Histone Modifications and Cardiovascular Disease

Several histone-modifying enzymes are expressed early in development and show clearly restricted expression patterns to certain cells types, marking them as key factors in development and disease. Recent studies point to the importance of enzymes that control histone acetylation as stress-responsive regulators of gene expression in the heart[86]. Illi *et al*[87]. report that the exertion of shear stress on embryonic stem cells results in cardiovascular lineage commitment which is accompanied by an increase in the histone modifications H3K14me, H3S10phos, and H3K79me as well as an early induction of cardiovascular markers.

The class II histone acetyl transferases (HDACs 4, 5, 7 and 9) are expressed at the highest levels in heart, brain and skeletal muscle and contain a C-terminal catalytic domain and an N-terminal extension that mediates interactions with other transcription factors. Consistent with the notion, that class II HDACs suppress pathological growth, mice lacking HDAC5 or HDAC9[88] display hypertrophy. Mice lacking both enzymes are prone to embryonic and early postnatal death from a spectrum of cardiac abnormalities including ventricular septal defects and thin-walled myocardium[89].

Recently, it has been shown that also members of the ubiquitously expressed class I HDACs are essential for heart development[90]. Mice with global deletion of either HDAC1 or HDAC2 are embryonic lethal with a variety of phenotypes. Cardiac-specific deletion of one of these enzymes does not evoke any phenotype whereas cardiac-specific deletion of both HDAC1 and HDAC2 results in neonatal lethality, accompanied by cardiac arrhythmias and dilated cardiomyopathy. These results demonstrate that HDAC1 and HDAC2 have redundant roles in regulating cardiac morphogenesis, growth and contractility.

1.1.2 Transcription Factors in Congenital Heart Diseases

Congenital heart diseases (CHD) are the most common birth defects in humans. They arise during development of the embryo and affect 1 in every 100 live births and an even higher number in miscarriages[91,92]. Despite advances in surgical therapy, morbidity and mortality remain high. Some malformations are not amenable to surgery, and even treated defects may result in a reduced life span due to residual heart disease. It is therefore highly important to understand the causes of these common birth defects.

To gain insight into the formation of cardiac anomalies molecular genetic studies of human patient populations have been carried out. Linkage analysis and candidate-gene approaches have led to the identification of a variety of CHD-causing gene mutations. At the core of the molecular and developmental pathways are transcription factors. These proteins play fundamental roles in regulating the pattern and timing of expression of genes responsible for the cardiac lineage determination, valvulogenesis, conduction-system development and heart chamber formation[93]. Understanding the function of these TFs is the next step towards comprehending the causes of CHDs. Current studies indicate that transcription factor mutations demonstrate variable expression, i.e., varied phenotypes even from the same mutation. This suggests that the regulatory context of the TF plays an important role in the disease manifestations. Therefore, the function of TFs must be viewed in the context of transcriptional networks which also include the interplay between different TFs and co-regulatory elements as well as epigenetic effects.

Four TFs were selected which are evolutionarily conserved, necessary for heart development in mouse models and associated with CHD in human patients. Previous results suggested that these TFs might form a subnetwork, as physical interactions and mutual regulation had been described at least on a pair-wise basis. The overexpression of each of these factors initiates cardiomyogenesis in P19 cells and can also activate the expression of each other (Gata4 → *Nkx2.5*[94,95]), (Srf → *Gata4*[96]), (Srf → *Nkx2.5*[97]), (Srf → *Srf*[98]). In a cell culture model where function of all Mef2 proteins was abolished, the expression of *Gata4* and *Nkx2.5* was reduced[99]. It remains to be shown whether they are directly regulated and by which of the Mef2 family members. These transcription factors (Gata4, Mef2a, Nkx2.5, and Srf) will be introduced in the following paragraphs.

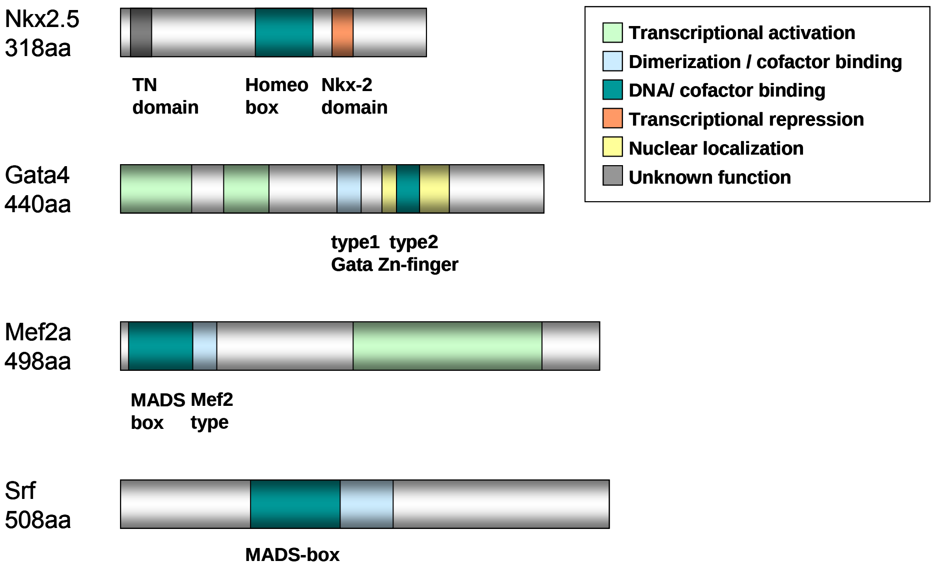

Figure 1-4. Schematic structures of investigated transcription factors.

1.1.2.1 GATA Binding Protein 4 (Gata4)

The GATA family consists of six members numbered 1-6. The name is derived from their binding motif GATA. Factors *GATA1* to *-3* are important regulators of hematopoietic stem cells and their derivatives, while *GATA4* to *-6* are expressed in various mesoderm and endoderm-derived tissues[100,101]. Like other proteins of the family, GATA4 is approximately 50 kDa in size and contains two zinc finger domains of the form C-X-N-C-(X17)-C-N-X-C[102](Figure 1-4). Most of the protein–protein interactions of GATA factors are mediated by their C-terminal zinc finger, while the N-terminal zinc finger interacts with 'Friend of GATA' (FOG) transcription factors: Fog-2[103] and probably also Fog-1[104]. Further cardiac interaction partners described either in mouse or human are Mef2[105], Nkx2.5[106-108], SRF[109,110], Hand2[111], GATA6[112], NF-AT3[113], p300[114], RXRA[115], KLF13[116], SP1[117], YY1[118], and Tbx5[119]. In skeletal muscle Gata4 is known to interact with MyoD and Myf5[120]. The Gata4 protein is subject to post-translational modifications which affect its DNA-binding activity, transactivation and/or localization within cardiac myocytes. Phosphorylation of serine-105 through kinases leads to an increased binding of GATA4 to DNA[121].

During murine development *Gata4* is expressed in heart, gut, gonads, liver, and endoderm[122]. *Gata4* is one of the earliest transcription factors expressed in developing murine cardiac cells[123] as early as 7.0–7.5 days postcoitum (dpc) in the precardiac mesoderm and both the transcript and protein are found during formation and bending of the heart tube (8.0 dpc) in endocardium, myocardium and precardiac mesoderm[123]. Expression continues in cardiac myocytes throughout the life of the mouse[122,123]. In adult animals, *Gata4* transcripts are additionally found in gonads, lung, liver and small intestine.

GATA4 binds to the *Nkx2.5* promoter and activates its expression[124]. The transcription factors GATA4 and MEF2 physically interact to synergistically activate cardiac gene expression[125]. Similarly combinatorial action is known between Gata4 and Srf. Other co-activators of human GATA4 include HAND2[111], STAT-1[126], KLF13[116], and YY1[118]. Retinoid X receptor alpha (RXRA) represses GATA4-mediated transcription via a retinoid-dependent interaction with the cardiac-enriched repressor FOG-2[115].

Cell culture experiments in P19 cells have shown that overexpression of *GATA4* increases differentiation of cardiomyocytes[127], while *GATA4* knockdown leads to apoptosis and blocks differentiation[128]. ES cells lacking *GATA4* have reduced ability to differentiate to cardiac myocytes[129]. Mice lacking *Gata4* die between 8.0 and 9.0 dpc because of failure of ventral morphogenesis and heart tube formation[130,131]. However, in these knockouts some direct target genes of Gata4 are still expressed. This points to a possible partial compensation by Gata6[130,131]. Gata4 has been implicated to play a role in hypertrophy, as its DNA-binding activity increases during initiation of hypertrophic response of cardiac myocytes *in vitro* and *in vivo*[132].

In human patients with 8p23.1 monosomy *GATA4* haploinsufficiency is associated with congenital heart defects[133]. Furthermore, mutations leading to amino-acid exchange in GATA4 at position 296 (glycine to serine) leads to cardiac septal defects[119], possibly due to disruption of GATA4 and TBX5 interactions. *GATA4* and *Nkx2.5* mutations have also been reported as disease genes of familiar atrial septal defect[134,135] and implicated in causing *Tetralogy of Fallot* (TOF)[136] (Figure 1-5).

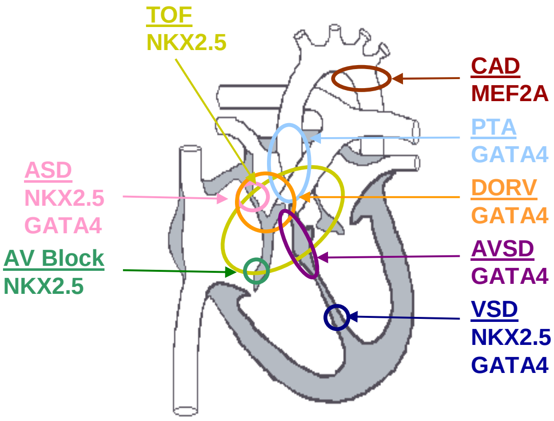

Figure 1-5. Sites of structural anomalies associated with mutations of the investigated TFs. ASD: atrial septal defect; AV: atrioventricular; AVSD: atrioventricular septal defect; CAD: coronary arterie disease, DORV: double-outlet right ventricle; PTA: persistent truncus arteriosus; TOF: tetralogy of Fallot; VSD: ventricular septal defect.

1.1.2.2 The Transcription Factor Nkx2.5

The *tinman* gene in the fruit fly *Drosophila melanogaster* encodes a homeobox-containing transcription factor which is required for the formation of the primitive heart[137]. The identification of Nkx2.5 (also called cardiac-specific homeobox (Csx)) in mammals[138] attracted much attention to the key regulatory roles of cardiac transcription factors in heart development. Nkx2.5 belongs to the NK2 homeobox gene family which commonly have highly conserved structure composed of the N-terminal TN domain, the homeodomain, and the NK2-specific domain, located just C-terminal to the homeodomain[139] (Figure 1-4). The name 'Nkx2.5' originated from a taxonomical standpoint that it is the fifth vertebrate gene identified from the NK2 homeobox gene family.[140] The homeodomain of Nkx2.5 has a helix-turn-helix motif that binds to the specific consensus DNA sequence 5' T(C/T)AAGTG 3'[141].

Nkx2.5 is highly expressed in the early heart progenitor cells in both primary and secondary heartfields during murine embryogenesis and continues to be expressed at high levels in the heart throughout adulthood[138,140]. A transient elevation of *Nkx2.5* levels has been observed in specialized myocardial conduction cells during the period of conduction system formation, suggesting a significant role for Nkx2.5 in the development of the conduction system[142].

Cell culture experiments have shown that Nkx2.5 function is essential for commitment of mesoderm into the cardiac muscle lineage, and the N-terminal region, together with the homeodomain, is sufficient for cardiomyogenesis in P19 cells[143]. Gain-of-function studies in *Xenopus* embryos of *XNkx2.5*[144] and injection of *Nkx2.5* into zebrafish embryos[145] leads to increased myocardial cell numbers (hyperplasia). Nkx2.5 is also critical for the regulation of the secondary heart field proliferation and the development of the outflow tract morphology in mice[146]. Targeted interruption of *Nkx2.5* in mouse embryos resulted in abnormal heart morphogenesis, growth retardation and embryonic lethality at approximately 9-10 dpc. Heart tube formation occurred normally, but looping morphogenesis was not initiated at the linear heart tube stage (8.25-8.5 dpc)[147].

In humans, mutation in the *NKX2.5* gene provided the first evidence that the genetic factors are crucial in non-syndromic congenital heart disease (CHD)[148]. More than ten disease-related mutations in *NKX2.5* have been documented in patients with a spectrum of congenital heart diseases (reviewed by Akazawa *et al.*[139]). The most common phenotypes are secundum ASD and AV conduction disturbance, but other cardiac abnormalities have been reported, such as VSD, TOF, double-outlet right ventricle (DORV), tricuspid valve abnormalities including *Ebstein's anomality* and *hypoplastic left heart syndrome* (Figure 1-5).

1.1.2.3 Myocyte Enhancer Factor 2a (Mef2a)

Myocyte enhancer factors 2 (Mef2 proteins) belong to the MADS-box family of transcription factors. Members of this family contain a conserved MADS-box domain and a MEF2 domain at the N-terminus of the protein which are involved in DNA-binding and dimerization, and a C-terminal transactivation domain involved in the regulation of transcriptional activity (Figure 1-4). The four Mef2 proteins in vertebrates (Mef2A, Mef2B, Mef2C and Mef2D[149,150]) bind to the A-T-rich Mef2 sites CTA(A/T)$_4$TAG as homo- and heterodimers[149]. Mef2 proteins can act both as activators and as repressors by interaction with histone acetyl transferases (HATs) and histone deacetylases (HDACs).

Mef2 was originally identified in myotubes, but is now known to be expressed in most tissues. The four Mef2 proteins are expressed in precursors of the three muscle lineages as well as in neurons[149]. *Mef2c* is the first member of the family to be expressed in the mouse with transcripts appearing in the precardiac mesoderm at embryonic day 7.5 (E7.5)[151]. Expression of all four *MEF2* transcripts (*MEF2A, MEF2B, MEF2C*, and *MEF2D*) can be detected in all developmental stages of the human heart[152]. Promoter analysis has shown that MEF2 sites are essential for the expression of muscle-specific genes in cardiac muscle and skeletal muscle[149,153]. Mice lacking *Nkx2.5*, and *Drosophila* lacking the *Nkx2.5* homologue *tinman* or the *GATA4* homologue *pannier*, show a downregulation of *Mef2*[154-156].

The strongest indication of an essential role for MEF2 in muscle development comes from studies in *Drosophila*. Loss-of-function mutations of the single *Mef2* gene resulted in a block in the development of all muscle cell types in the embryo. Although the cardiac muscle contractile proteins were not expressed in these mutants, a dorsal vessel formed normally and expressed tinman, a homeobox transcription factor. Together with the finding that *Mef2* contains essential tinman-binding sites[157,158] these data suggest that Mef2 controls the conversion of cardiomyoblasts into cardiomyocytes.

Most mice lacking Mef2a died suddenly within the first week of life and exhibited pronounced dilation of the right ventricle, myofibrillar fragmentation, mitochondrial disorganization and activation of a fetal cardiac gene program. The few *Mef2a*$^{(-/-)}$ mice that survived to adulthood also showed a deficiency of cardiac mitochondria and susceptibility to sudden death[159]. Experiments using dominant-negative MEF2A protein have shown that MEF2A is essential for skeletal myogenesis in myoblast cell lines[160].

In humans, mutations in *MEF2A* have been proposed as a cause of coronary artery disease and myocardial infarction[160-164] and was termed the first non-lipid-related gene causing these diseases (Figure 1-5). However, there are also studies where a large number of patients with these diseases were screened and no mutations within *MEF2A* could be detected[165,166]. Therefore, the extent to which coronary artery disease and myocardial infarction are caused by mutations in the *MEF2A* gene is currently still controversial[167,168].

1.1.2.4 Serum Response Factor (Srf)

Serum response factor (Srf) is a further member of the MADS family of TFs. (Figure 1-4). Srf binds to the CArG-box motif CC(A/T)$_6$GG. Target genes with only one CArG-box (typically *c-Fos*[169]) are known, however, the majority of SRF targets have duplicate CArG-boxes[170] that function cooperatively. The expression of the human SRF protein has been described to be autoregulatory[171].

Srf is known to associate both with positively and negatively acting cofactors. Positively acting SRF cofactors include the cardiac members of the GATA family[110] (GATA-4, -5 and -6) and the NKX2.5 family of homeodomain proteins[172], which can form complexes both with SRF and their own adjacent recognition site[173]. Myocardin (Myocd) is probably the best studied cofactor of Srf[174,175]. Myocardin binds Srf and potently stimulates Srf-dependent transcription, leading to the activation of smooth muscle genes in the process of differentiation. It has been proposed that interaction of Myocardin and Srf may function as a molecular switch between the proliferation and the differentiation programs of smooth muscle cells[176]. A recent investigation suggests that Myocardin may even function as a switch between skeletal and smooth muscle gene expression[177].

Negatively regulating SRF cofactors include the heart-enriched homeodomain-only cofactor HOP. HOP does not bind DNA by itself[178] and acts downstream of NKX2.5[179]. Although interaction with HOP inhibits the DNA-binding activity of SRF[180,181], HOP has also been implicated in recruitment of histone deacetylase activity to SRF target genes[182]. Protein interaction studies demonstrated physical association of HDAC4 with Srf in living cells leading to repression of Srf activity[183]. Mef2 and Srf can compete for transcription factor binding sites (TFBSs) for example at the *MyoD* promoter[184] in skeletal muscle cells.

Cell culture experiments have shown that SRF is essential for serum-dependent cell growth and skeletal muscle differentiation[185-187] as well as differentiation of chicken coronary smooth muscle cells[188]. Consistent with this function, the SRF protein is most abundant in embryonic heart, skeletal, and smooth muscle tissues, but barely detected in liver, lung, and spleen tissues. During early mouse development, *in situ* hybridization analysis revealed enrichment of *Srf* transcripts in the myotomal portion of somites, the myocardium of the heart, and the smooth muscle media of vessels of mouse embryos[171].

ES cells lacking *SRF* display defects in spreading, adhesion and migration. These defects correlate with defective formation of cytoskeletal structures, namely actin stress fibers and focal adhesion plaques[189]. In *SRF-null* neonatal cardiomyocytes, severe defects in the contractile apparatus are observed[190]. Homozygous *Srf-null* mutation in mice result in lethality at gastrulation[191]. In Cre mice lacking skeletal muscle *Srf* expression died during the perinatal period from severe skeletal muscle hypoplasia[192]. Cre mice with knockout of *Srf* in 80% of the cardiomyocytes display severe heart defects and die at dpc 11.5[193].

1.2 Experimental Methods

The first step towards understanding the functions of DNA-binding proteins is to obtain information on where this binding occurs. Here, both detailed information on single binding sites and global maps are highly relevant. Traditional *in vitro* methods such as gel-shift assays and oligonucleotide-based selection techniques are used to determine consensus binding sequences of transcription factors but show little predictive power for the regulation of single genes *in vivo*[194]. One reason for this is the small length of DNA-binding motifs of typically 5 to 10 bp which may occur thousands of times in the genome while only a fraction is actually functional. In addition several alternative motifs may be possible and the binding may also be cell-type specific. This fact has also made the development of reliable computationally based predictions difficult.

1.2.1.1 Chromatin Immunoprecipitation

A powerful technique to map binding sites of proteins *in vivo* is the technique of chromatin immunoprecipitation (ChIP). Combining ChIP with microarray analysis (ChIP on chip) has made it possible to screen a huge number of DNA regions for binding sites of a protein of interest or even the whole genome. As starting material cells from cell culture[195], tissue[196], primary cells and embryos[197] can be used.

Cell lines have the advantage of higher homogeneity and clearly defined cell states, can be easily dispersed and grown in high numbers. Cells are cultured under the desired experimental conditions and proteins are cross-linked to DNA using formaldehyde. Formaldehyde can also crosslink proteins to each other by reacting with the ε-amino groups of lysine residues and an adjacent peptide bond. DNA-protein crosslinks are formed with the –CO-NH moiety at position 1 (N-1) of a guanine or the exocyclic amino groups of an adenine, guanine, or cytosine. Although other crosslinking reagents have been employed[198] formaldehyde remains the most widely used as the reaction can be reversed by heat.

After cross-linking the chromatin, the cells are either directly lysed or the extract is first enriched for nuclei. The chromatin is sheared to fragments of the desired size by sonication or through micrococcal nuclease digest to a size of usually 0.5-1 kb. Part of the fragmented chromatin is directly purified as total genomic reference DNA (input). The fragments bound to the protein of interest are usually enriched by immunoprecipitation with an antibody against the respective protein. If such an antibody is not available it is also possible to overexpress the protein with a tag (e.g. biotin tag[199], FLAG tag[200] or TAP tag[201]) and then to enrich via the tag; these approaches, however, have the disadvantage that the necessary overexpression leads to a disturbance of the normal physiological protein concentration and therefore the results obtained using tagged proteins may not reflect the normal binding behavior accurately. On the other hand, using protein-specific antibodies makes optimizing IP conditions of each individual antibody necessary and often these may show unwanted cross-reactivity. In parallel to the actual ChIP experiment one sample is processed with the addition of pre-immune serum from the host organism of the specific antibody. This serves as control to identify fragments unspecifically enriched e.g. by adhesion to the samples tubes.

The formaldehyde cross-links are then heat-reversed and the precipitated DNA fragments are purified. Yields from ChIP are usually low but sufficient for subsequent PCR or qPCR analysis. However, for microarray-based detection (ChIP-chip) amplification of the DNA is generally necessary, although ideally the ChIP reactions are scaled up and amplification avoided. Three amplification methods have so far been widely used: Randomly-primed[202] and ligation-mediated PCR[203] as well as amplification on the basis of T7 DNA polymerase[204]. Although input samples usually give enough material for microarray hybridization they should also be amplified so that if an amplification bias occurs this should ideally be the same in the ChIP as in the input sample.

The enriched and the reference DNA are then fluorescently labeled. Although one color platforms where both samples have the same label e.g. biotin are hybridized on separate arrays the use of two color platforms is often preferred, as this minimizes the influence of microarray batch effects on the experimental results. In this case the ChIP DNA is labeled with a fluorescence dye such as Cy5 or Alexa 647 and the input with Cy3 or Alexa 555, the samples are combined and hybridized to a single DNA microarray. The relative intensities of the two dyes allow identification of the fragments enriched in the IP thereby enabling the finding of protein-DNA interaction sites.

For a comprehensive and unbiased analysis microarrays used in ChIP-chip applications ideally represent the entire genome of the organism in the form of overlapping fragments (so called *tiling arrays*). For larger genomes such as for higher eukaryotes these are, however, not available or only at very high monetary cost. Therefore arrays are often custom designed for specific applications. The resolution of the identified binding sites depends on the size of the sheared DNA and the size and spacing of the probes on the arrays. With the sinking cost of high-throughput sequencing recently protocols coupling ChIP with whole genome sequencing (ChIP-Seq) have been developed[205].

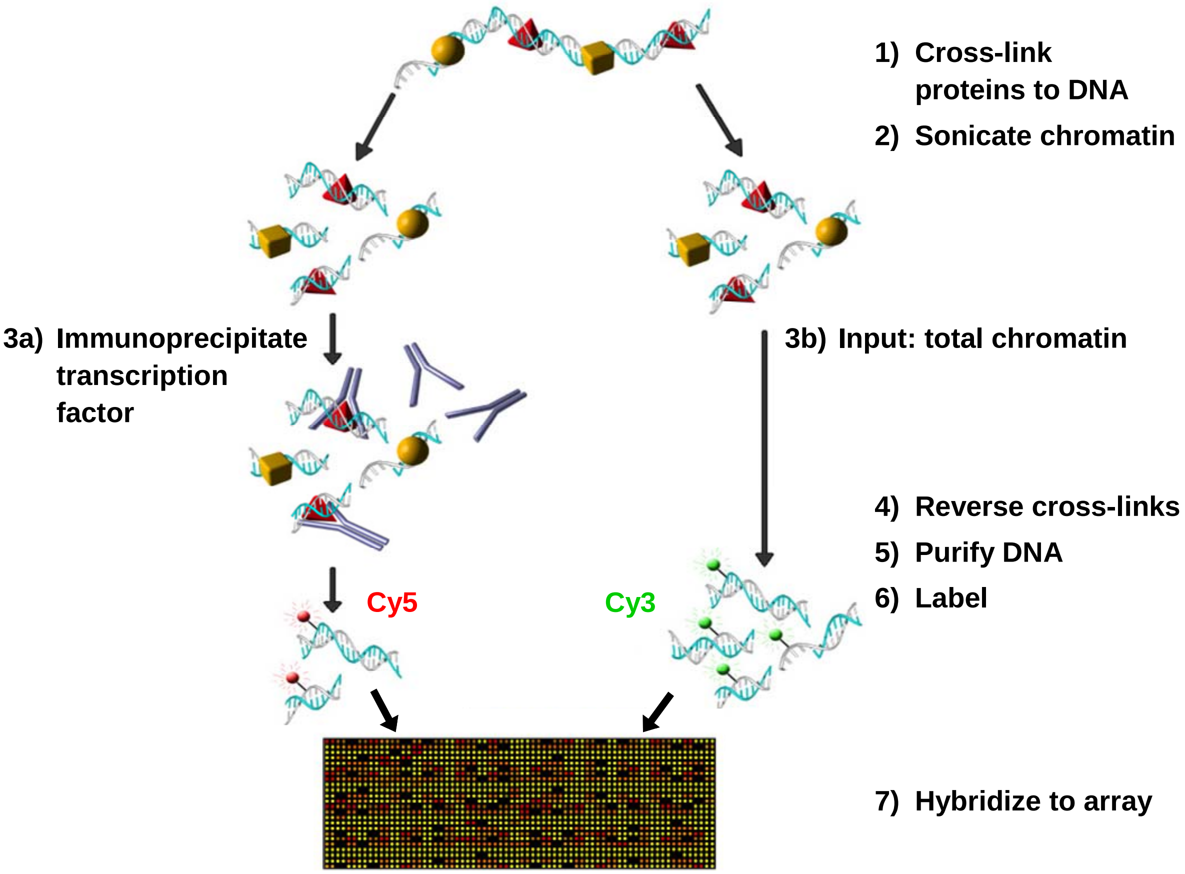
1) Cross-link proteins to DNA
2) Sonicate chromatin
3a) Immunoprecipitate transcription factor
3b) Input: total chromatin
4) Reverse cross-links
5) Purify DNA
6) Label
Cy5
Cy3
7) Hybridize to array

Figure 1-6. Schematic representation of ChIP-chip.

1.2.1.2 RNA interference

ChIP-chip analysis is a powerful tool to investigate the location of histone modifications or transcription factor binding *in vivo*. However, the effect on transcription may be activating, repressive or non-functional and this cannot be determined by ChIP-chip. Consequently, to gain functional insights it is necessary to couple ChIP with the analysis of the transcriptome of the cells. The expression status of the genes in the vicinity of the epigenetic marks or binding sites can be analyzed. This can give information on the overall association between the investigated factor and transcription. However, in case of transcription factors the effect on each target gene may be different.

The method of choice is to compare the expression status of cells depleted of the transcription factor with that of normal cells. Thereby information can be gained on all genes influenced by the presence of the TF. In combination with ChIP-chip, direct targets can be separated from downstream pathways and the influence on each target gene can be determined. Common methods to achieve such a depletion are: knockdown by RNA interference (RNAi) or antisense methods (e.g. phosphorothioate-linked DNA[206], morpholinos[206,207]) or genetic knockouts[208].

RNA interference (RNAi) is an intrinsic cellular mechanism which is conserved in most eukaryotic species. It plays a role in the regulation of gene expression, differentiation and defense against viral infections. RNA interference plays an important role in determining cell fate and survival. The relevance of the field has recently been acknowledged by the Nobel prize in Physiology or Medicine 2006 to Andrew Z. Fire[209] and Craig C. Mellow[210] for the first description of the phenomenon[211,212].

The natural mechanism has been utilized to artificially silence particular genes and thereby to gain insight into their functions[213]. Since the first discovery of gene-specific silencing by RNA molecules, efforts have been made to understand the mechanisms involved in the natural pathways and in artificially induced responses. Both have been extensively reviewed[214-220], therefore here only those aspects relevant to the study will be briefly introduced.

The RNAi signaling pathway is part of a larger network in which small RNA molecules are used as regulators for cellular signals and dsRNA is used as a switch for sequence specific gene silencing. In mammalian cultured cells short synthetic siRNAs of 19-23 nucleotide length are commonly used (Figure 1-7), because introduction of dsRNA longer than $\approx 30\,\text{bp}$ induces an antiviral interferon response[221]. A common mode of introducing the siRNAs into the cytoplasm is by lipofection. In the cytoplasm the siRNA is incorporated into the RNA-inducing silencing complex (RISC) which consists of an Argonaute protein (Ago) as one of its main components[222]. The dsRNA is unwound in an ATP-dependent process. Ago cleaves and discards the passenger strand (sense) of the siRNA duplex leading to an activation of RISC. The strand whose 5'-end is at the less stable end of the duplex siRNA is the final guide strand. The remaining guide (antisense) strand of the siRNA directs RISC to its homologous mRNA, resulting in the endonucleolytic cleavage of the target mRNA. In the RNAi pathway the mature RISC complex containing the siRNA operates as a multiple turnover enzyme: once the mRNA is cleaved the complex dissociates from the fragments to repeat the cleavage multiple times.

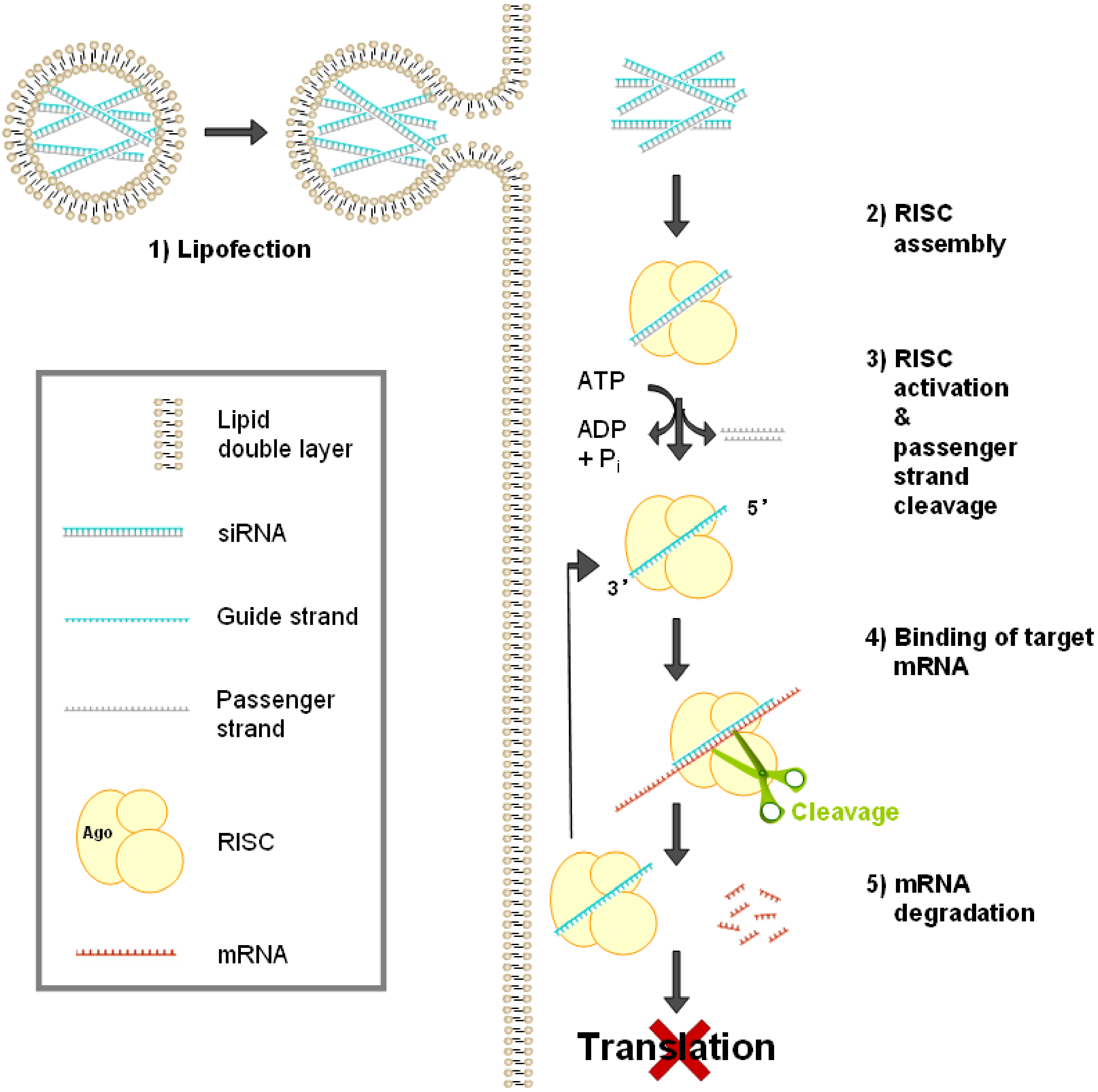

Figure 1-7. Schematic outline of the RNA interference pathway. Here only lipofection of siRNA as used in the study is shown.