NF-κB

Since many years, our laboratory work has been focused on the transcription factor NF-κB, and more particularly, its role in cancer (Viatour *et al*., 2003; Viatour *et al*., 2004; Bonif *et al*., 2006; Benoît *et al*., 2006). NF-κB was first identified more than 20 years ago in B-cells, as a protein complex capable of binding to a 10 base pairs sequence within the transcriptional enhancer of the immunoglobulin κ light chain gene and was therefore named: Nuclear Factor for κB (Sen and Baltimore, 1986). If NF-κB was initially thought to be specific of B-lymphocytes, it was rapidly found to be an important transcriptional factor, present in a wide variety of cell types and implicated in a large number of cellular processes. NF-κB mediates the cellular responses to multiple extracellular stimuli, through the regulation of target genes expression. To date, more than 100 genes (listed at www.nf-kb.org) have been shown to contain one or more κB sites (5'-GGGRNNYYCC-3', R=purine, Y=pyrimidine, N=purine or pyrimidine) that confer NF-κB-responsive transcriptional regulation. They reflect the critical regulator role of NF-κB in many physiological processes, such as innate and adaptative immune responses, inflammation, cell adhesion, proliferation, control of apoptosis, cellular-stress response, … It is therefore not surprising to see that deregulation of NF-κB can underlie diverse pathologies like immune and inflammatory diseases, but also induces an imbalance between cell proliferation and apoptosis, leading to oncogenesis.

The NF-κB and IκB families

❖ *The Rel/NF-κB family*

In mammals, the NF-κB family consists of five related transcription factors: **RelA** (p65), **RelB, c-Rel**, **p105/p50** (NF-κB1) and **p100/p52** (NF-κB2), which exert their function as homo-or heterodimers. The heterodimer p50/p65 is the most common form, but all combinations are possible, with the exception of RelB, which interacts only with either p50 or p52. These different NF-κB dimers contribute to differentially control gene expression and appear to mediate distinct biological functions, although some redundancy also exists among some of them.

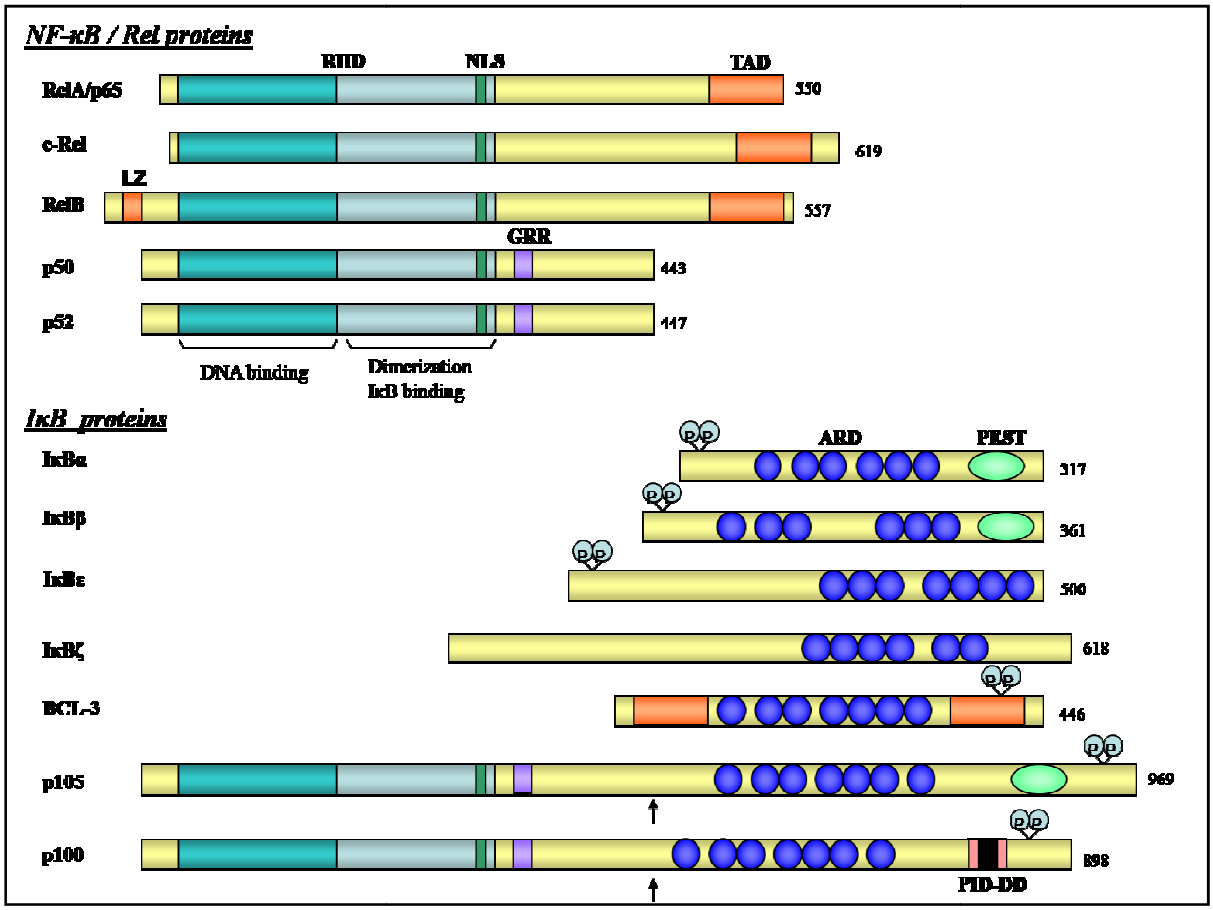


Figure 1: Schematic representation of NF-κB/Rel and IκB proteins. The mammalian **NF-κB** family consists of five members: RelA (p65), c-Rel, RelB, p50 (NF-κB1) and p52 (NF-κB2). These proteins all share an N-terminal Rel homology domain (RHD), which mediates DNA binding, dimerization and interaction with inhibitors and which contains the nuclear localization signal (NLS). RelA, c-Rel and RelB are synthesized as mature proteins and possess a transactivation domain. p50 and p52 are synthesized as longer precursors, p105 and p100 respectively. The **IκB** (Inhibitor of NF-κB) family consists of the "classical" IκBs (IκBα, IκBβ and IκBε), the NF-κB precursors (p105 and p100) and the "unusual" IκBs (BCL-3 and IκBζ). They contain an ankyrin repeats domain (ARD) composed of five to seven ankyrin repeat sequences and responsible for the interaction with the NF-κB proteins. Unlike the other IκBs, BCL-3 is nuclear and possesses two transactivation domains. The number of amino acids of each protein is indicated on the right. The arrows indicate the presumed site for cleavage for p100 and p105. **RHD**: Rel homology domain, **NLS**: nuclear localization signal, **TAD**: transactivation domain, **LZ**: leucine zipper, **GRR**: glycine-rich region, **ARD**: ankyrin-repeats domain, **PEST**: polypeptide sequence enriched in proline, **PID**: processing inhibitory domain, **DD**: death domain.

These proteins all share a highly conserved N-terminal domain of approximately 300 amino acids, called the **Rel homology domain** (**RHD**) (**Figure 1**). This domain mediates DNA-binding, dimerization, interaction with the inhibitory factors IκBs and nuclear localization via a **nuclear localization signal** (**NLS**).

The five NF-κB members can be further classified into two groups. The first one, called the *Rel family*, contains RelA, RelB and c-Rel. These proteins are synthesized as mature proteins and harbor a C-terminal **transactivation domain** (**TAD**), which allows them to activate target genes expression. In contrast, proteins of the *NF-κB subfamily*, p50 and p52, are generated through the C-terminal processing of their large precursor, p105 and p100, respectively. These proteins do not contain a transactivation domain and need to associate with RelA, c-Rel or RelB, to form transcriptional activating dimers. Homodimers p50 and p52 act as transcriptional repressors, unless bound to the coactivator BCL-3, which contains two transactivation domains. They can therefore regulate κB-mediated transcription either positively or negatively, depending on the relative concentration of their partners in the nucleus.

❖ ***The IκB family***

NF-κB activity is regulated by translocation of the dimers from the cytoplasm to the nucleus in response to cell stimulation. In most unstimulated and untransformed cells, NF-κB dimers are sequestered in the cytoplasm as inactive complexes by interaction with specific NF-κB inhibitors, known as the IκB (Inhibitor of κB) proteins (**Figure 1**). This family includes the structurally related proteins **IκBα**, **IκBβ** and **IκBε**. These inhibitors contain an N-terminal regulatory domain, with lysine and serine residues, and a central **ankyrin repeats domain (ARD)** which interacts with the RHD of the NF-κB proteins, masking the NLS and preventing dimers to move into the nucleus. However, in the complex IκBα/p65/p50, the NLS of p50 remains accessible, but the simultaneous presence of the nuclear export sequence (NES) of IκBα results in a constant shuttling of the complex between the nucleus and the cytoplasm (reviewed in Hayden and Ghosh, 2004 and Perkins, 2007).

Interestingly, the extended IκB family also includes **p100** and **p105**, the precursors of p52 and p50, respectively. Indeed, these proteins contain in their C-terminal part, ankyrin repeat motifs, similar to those found in IκB proteins which form an auto-inhibitory IκB-like domain (Blank *et al*., 1991; Henkel *et al*., 1992; Mercurio *et al*., 1992; Naumann *et al*., 1993a, 1993b; Rice *et al*.,

1992; Scheinman *et al*., 1993). The processing of these precursors into the active forms p52 and p50 involves the proteasome-mediated degradation of this C-terminal inhibitory domain (Fan and Maniatis, 1991; Sienbenlist *et al*., 1994) and requires a **glycine-rich region** (**GRR**) (Lin and Gosh, 1996). Whereas the processing of p105 is constitutive (Harhaj *et al*., 1996; Belich *et al*., 1999; Heissmeyer *et al*., 1999, Orian *et al*., 2000), p100 processing is tightly controlled due to the presence of a **processing inhibitory domain** (**PID**) in its C-terminal (Xiao *et al*., 2001). As RelB has a low affinity for small IκBs, p100 is its main inhibitor and p100 processing appears to be the major regulatory pathway for the nuclear translocation of RelB-containing complexes.

Two other proteins, BCL-3 (B-cell lymphoma 3) and IκBζ are defined as "unusual" IκBs because they also contain ankyrin repeats in their central part, but unlike the other IκB members, they are predominantly localized in the nucleus and they do not bind transactivating dimers. Instead, due to its two transactivation domains, BCL-3 can activate gene transcription when bound to p50 or p52 homodimers.

NF-κB activating signaling pathways

NF-κB is the prototype of a latent cytoplasmic factor, whose activation does not require *de novo* protein synthesis, but relies on the degradation of the IκB inhibitory proteins and the subsequent migration of the free NF-κB dimers into the nucleus to modulate target genes expression (Traenckner *et al*., 1994). The two main NF-κB activating signaling pathways are referred as the "classical" or "canonical" pathway and the "alternative" or "non-canonical" pathway (**Figure 2**). These two NF-κB pathways can be activated by overlapping, but distinct sets of stimuli, and also target activation/repression of overlapping, but distinct sets of target genes. The canonical pathway is primarily essential for innate immunity, while the alternative pathway is mostly involved in lymphoid organ development and adaptative immunity.

❖ ***The classical pathway***

The classical pathway is the most frequently observed and the best characterized NF-κB activating signaling cascade. It concerns mostly the dimers containing RelA or c-Rel and can be induced by various stimuli, including pro-inflammatory cytokines such as tumor-necrosis factor (TNFα) or interleukine-1 (IL-1), byproducts of bacterial and viral infections, such as lipopolysaccharide

(LPS), mitogens and the ligand for the T-cell receptor (TCR). In response to these signals, IκBs proteins become phosphorylated on two conserved serine residues (S32 and S36 for IκBα, S19 and S21 for IκBβ) by the activated IKK (IκB kinase) complex. The classical IKK complex is a 700-900 kDa multiprotein complex which contains two catalytically active kinases, IKKα (IKK1) and IKKβ (IKK2) and a regulatory subunit NEMO (NF-κB essential modulator, also known as IKKγ), whose integrity is required for the activation of this pathway. While IKKβ is both necessary and sufficient for phosphorylation of IκBs, IKKα is relatively dispensable. ELKS is another essential regulatory subunit of the IKK complex and seems to recruit downstream IκB to the IKK complex for phosphorylation (Ducut Sigala *et al*., 2004).

Phosphorylation of IκBs triggers their ubiquitination by the $SCF^{\beta TrCP}$ (S-phase kinase-associated protein1-Cullin1-F-box; β-transducin repeat-containing protein) E3-ubiquitin ligase complex and their subsequent degradation by the 26S proteasome. The NF-κB dimers, mainly p50/p65, are then free to migrate into the nucleus, where they may be further regulated by modifications that affect their transactivation potency, such as phosphorylation and/or acetylation of their subunits. Phosphorylation of p65 involves different kinases including protein kinase A (PKA) or phosphatitdyl inositol 3 kinase (PI3K)/Akt (Zhong *et al*., 1998), while coactivators such p300/CBP would be implicated in acetylation of p65. These modifications facilitate recruitment of coactivators and binding to the κB sites in the promoter or enhancer region of their target genes. Activated NF-κB rapidly induces transcription of genes involved in inflammation, cell survival, innate immunity and angiogenesis and also increases expression of its own inhibitors, the members of the IκB family (except IκBβ). Deacetylation of p65 enhances its binding to the newly synthesized inhibitors, which, in turn, remove the NF-κB dimers from DNA and export them in the cytoplasm, providing thus a negative feedback control, which assures a transient nuclear expression of NF-κB (Karin and Ben-Neriah, 2000).

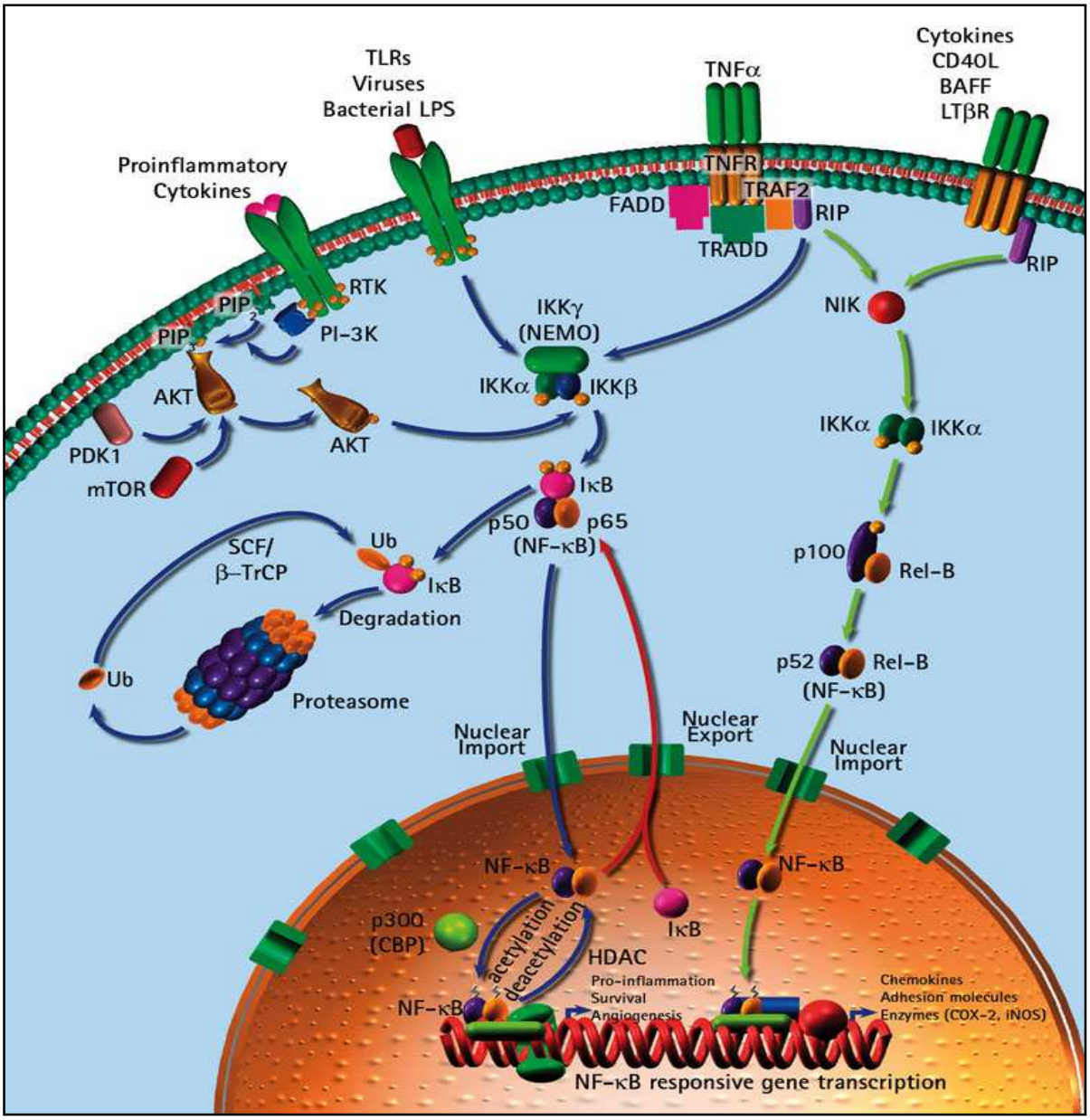


Figure 2: The NF-κB signaling pathways. The ***classical*** pathway is activated by a large number of stimuli, including byproducts of bacterial and viral infections (LPS,…) or proinflammatory cytokines such as TNFα. They trigger signaling cascades involving sequential phosphorylation of critical proteins which ultimately lead to the activation of the IKK complex, containing the catalytic subunits IKKα and IKKβ and the regulatory subunit IKKγ (NEMO). IKKβ, in turn, phoshorylates inhibitory proteins, such as IκBα, on specific residues, which results in its rapid ubiquitination and subsequent proteolytic degradation by the proteasome. The NF-κB dimer, mostly p50/p65, is therefore free to translocate into the nucleus where it undergoes further posttranslational modifications such as acetylation, which increases its DNA affinity. Activated NF-κB rapidly induces transcription of genes involved in inflammation, cell survival, innate immunity and angiogenesis and also increases expression of its own inhibitor, IκBα. Deactylation of p65 enhances its binding to IκBα, which, in turn, exports the NF-κB complexes back to the cytoplasm, thereby restoring the resting state.

In the ***alternative*** pathway, a restricted number of stimuli, such CD40L, BAFF or LTβR stabilizes the NF-κB inducing kinase (NIK), which activates an IKKα homodimer. Phosphorylation of p100 by IKKα induces its proteasomal-dependent processing, followed by the migration of the p52/RelB dimers into the nucleus. This pathway is involved in lymphoid organ development, maturation of B-cells and adaptative humoral immunity. **TLR**: Toll like receptor; **PIP**: phosphatidyl inositol phosphate; **PI3K**: phosphatidyl inositol 3 kinase; **RTK**: receptor tyrosine kinase; **m-TOR**: mammalian target of rapamycin, **PDK**: phosphoinositide-dependent kinase; **LPS**: lipopolysaccharide; **TNF**: tumor necrosis factor; **TNFR**: TNF receptor; **FADD**: Fas-associated protein with death domain; **TRADD**: TNF receptor associated death domain; **TRAF2**: TNF receptor associated factor 2; **RIP**: receptor interacting protein; **BAFF**: B-cell activating factor; **LTβR**: lymphotoxine β receptor; **Ub**: ubiquitin; **HDAC**: histone deacetylase; **CBP**: CREB binding protein; **COX2**: cyclooxygenase 2; **iNOS**: inducible nitric oxide synthase. *From* www.merckbiosciences.com.

❖ *The alternative pathway*

A limited number of stimuli including members of the TNF superfamily, such as LTβ (lymphotoxine β), BAFF (B-cell activating factor), CD40L (CD40 ligand), RANK ligand (receptor activator of NF-κB), TWEAK (TNF-like weak inducer of apoptosis),… have been shown to activate an alternative NF-κB signaling pathway, leading to the C-terminal degradation of p100. As p100 is found in the cytoplasm mostly dimerized with RelB, activation of this alternative pathway preferentially releases active p52-RelB dimers, which can then accumulate into the nucleus and regulate target genes expression. This pathway requires the NF-κB inducing kinase (NIK) and an alternative IKK complex, composed of two IKKα subunits, which functions independently of the larger classical IKK complex. Upon stimulation, NIK is stabilized (Qing *et al*., 2005b), activates and recruits the IKKα homodimer into the p100 complex via two C-terminal serines of p100 (S866 and S870) (Xiao *et al*., 2004). In turn, IKKα phosphorylates p100 on both N- and C-terminal serines (S99, S108, S115, S123, S872), triggering it for βTrCP-mediated polyubiquitination and subsequent processing by the 26S proteasome. Activation of the alternative pathway is delayed relative to that of the classical pathway and the heterodimers p52/RelB released have a higher affinity for distinct κB elements and might therefore regulate a distinct subset of NF-κB target genes.

NF-κB and cancer

Deregulated function of NF-κB contributes to the development of a variety of human diseases, particularly, immune-related diseases and cancer. Constitutive NF-κB activity has been found in many types of human tumors from either hematological or solid origin, such as melanomas, breast, prostate, ovarian, pancreatic, colon and thyroid carcinomas (Baldwin, 2001; Karin *et al*., 2002; Pacifico and Leonardi, 2006). The first indication suggesting the involvement of NF-κB in oncogenesis comes from the studies with v-Rel, the retroviral homolog of c-Rel, which was found to be highly oncogenic and to cause aggressive tumors in chickens (Theilen *et al*., 1966; Franklin *et al*., 1974; Gilmore, 1999). In humans, the *c-Rel* gene is frequently amplified or rearranged in different lymphomas, particularly in diffuse large B-cell lymphomas (DLBCL) and follicular lymphomas (FL) (Rayet and Gelinas, 1999; Gilmore *et al*., 2004). Genes that encode NF-κB2 (p52/p100) and BCL-3 are also located within regions of the genome involved in rearrangements or amplifications. The *bcl3* gene was identified in a t(14,19) chromosomal translocation in a subset of B-cell chronic lymphocytic leukemias (Ohno *et al*., 1990; Mc Keithan, 1997). The translocation t(10,14) involving the *nfκb2* gene, originally described in a case of B-cell non-Hodgkin's lymphoma, occurs in a number of lymphoid neoplasms, particularly in cutaneous lymphomas (Fracchiolla *et al*., 1993). Loss of function mutations in the *IκBα* gene, leading to a constitutive NF-κB activity have also been observed in Hodgkin's lymphomas.

Beside these chromosomal translocation events affecting the NF-κB/IκB family members, NF-κB has been linked to oncogenesis by some oncoproteins which signal through NF-κB activation (Baldwin, 2001; Kim *et al*., 2006). For example, in murine fibroblasts, p65/RelA and c-Rel are required for efficient cellular transformation induced by oncogenic Ras (Hanson *et al*., 2004; Finco *et al*., 1997; Norris and Baldwin, 1999). The EGF receptor family member Her-2/Neu (ErbB2), which is amplified and active in a significant proportion of breast cancer, activates NF-κB through a PI3K/Akt signaling pathway (Zhou *et al*., 2000; Pianetti *et al*., 2001). The Bcr-Abl fusion oncoprotein associated with chronic myelogenous leukemia was also shown to activate NF-κB and inhibition of NF-κB with super-repressor IκBα expression blocked Bcr-Abl-induced tumor growth (Reuther *et al*., 1998). NF-κB is also activated by a number of viral transforming proteins, such as the HTLV-I tax protein (human T-cell leukemia virus-I) (Chu *et al*., 1998), the latent membrane protein 1 (LMP1) of Epstein-

Barr virus (EBV),… and, in some cases, is required for virus-induced transformation (Yamaoka, 1996; Mosialos, 1997).

Chronic inflammation contributes also to the genesis of many solid tumors, such as gastric, colon or hepatic carcinomas. Recently, NF-κB was shown to activate a network between epithelial and inflammatory cells, which is thought to be crucial for the genesis and the establishment of carcinomas. Karin and co-workers utilized a colitis-associated cancer (CAC) animal model and showed that deletion of IKKβ in enterocytes reduced tumor incidence but not tumor size, while deletion of IKKβ in myeloid cells decreased tumor size (Greten, 2004). Similarly, Ben-Neriah and colleagues showed that inactivation of NF-κB during late phases of tumor promotion blocked progression to hepatocarcinoma in an *mdr2*$^{-/-}$ mouse model of cholangitis (bile-duct inflammation) (Pikarsky, 2004). Therefore, NF-κB activation by the classical IKKβ-dependent pathway seems to be a key component in inflammation-induced tumor growth and progression, providing a link between inflammation and cancer.

NF-κB participates in many aspects of oncogenesis. Cancer cells share common features that contribute to drive malignant growth (Hanahan and Weinberg, 2000): self-sufficiency in growth, strong resistance to growth inhibitory signals, evasion to apoptosis, limitless replicative potential, enhanced angiogenic potential and ability to invade local tissue and to metastasize to distant sites. Many of the genes able to mediate such effects are under transcriptional control of NF-κB. NF-κB can promote cell growth by upregulating transcription of genes involved in the cell cycle such as Cyclin D1, CDK2 kinase, c-Myc,… and which are altered in several types of human tumors. By inducing expression of genes whose products inhibit apoptosis, including inhibitors of apoptosis (IAPs: XIAP, cIAP-1, c-IAP-2), TNF receptor associated factors (TRAF1, TRAF2), members of the Bcl2 family (A1/Bfl-1, Bcl-xL), c-FLIP,… NF-κB prevents the death of tumor cells with DNA damage, either associated with carcinogenic initiation or cancer therapy. NF-κB also plays a role in later-stages of oncogenesis, and tumor progression by regulating expression of genes involved in angiogenesis, tumor invasion and metastasis, such as vascular endothelial growth factor (VEGF), matrix metalloproteinases (MMP2, MMP9), adhesion molecules (ICAM1, VCAM1, ELAM1), tissue plasminogen activator (tPA), chemokine receptors (CXCR4),…

❖ *NF-κB in lymphoid neoplasms*

The World Health Organization (WHO) classifies the lymphoid neoplasms into three main categories: the B-cell neoplasms, the T-cell neoplasms and the Hodgkin's lymphomas (**Table 1**). These groups can be further divided, depending on the following parameters: the postulated developmental stage of the cell of origin of the disease, the various sites of involvement, the morphologic, immunophenotypic, genetic and cytogenetic features of the cells, as well as the prognostic and predictive clinical factors (Harris *et al*., 1999). **Figure 3** illustrates the different stages of B and T cell differentiation and the relationship to the major B and T cell lymphoproliferative disorders.

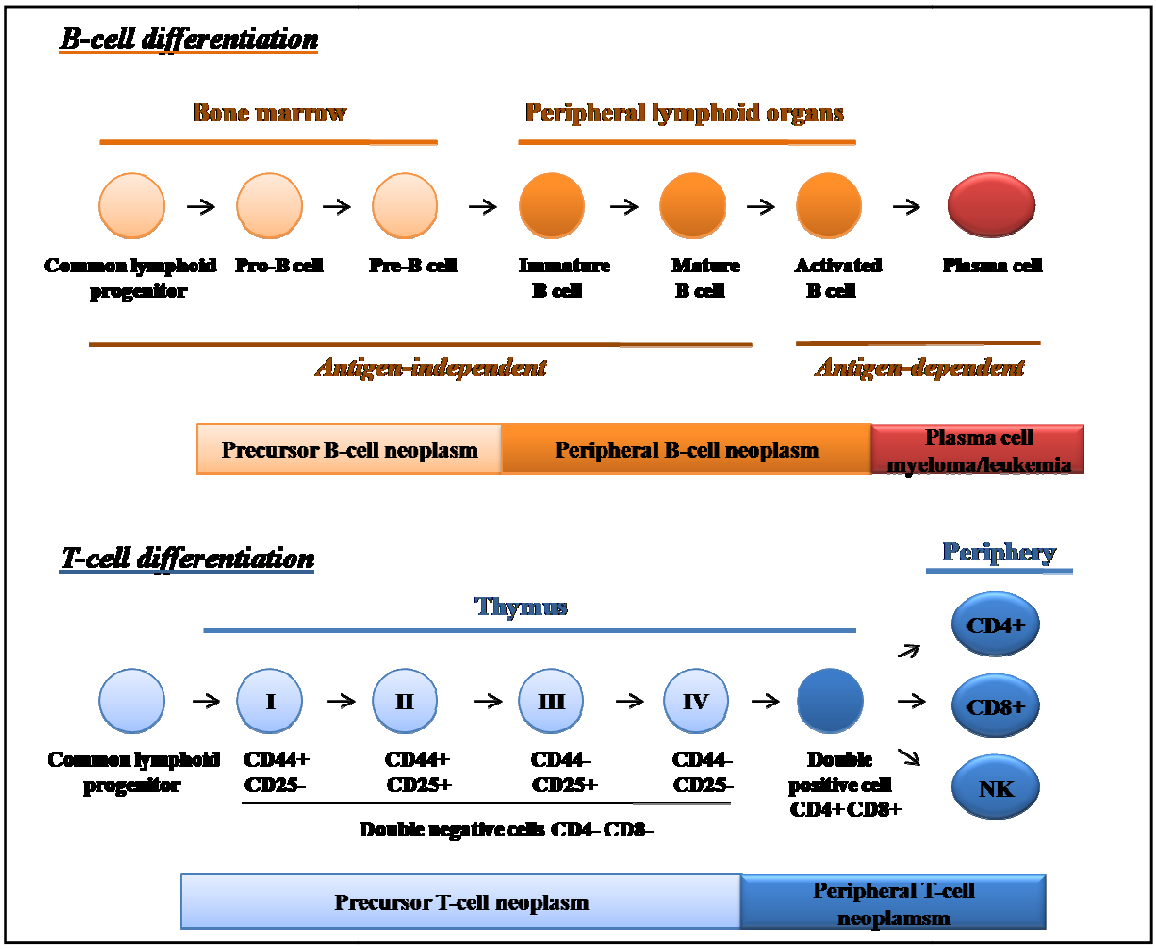


Figure 3: Schemes of B and T cell differentiation and their relationship to the major B and T cell lymphoproliferative disorders.

B-cell neoplasms	
Precursor B-cell neoplasm	Precursor B lymphoblastic leukemia/ lymphoma
Peripheral (mature) B-cell neoplasm	B-cell chronic lymphocytic leukemia/small lymphocytic lymphoma
	B-cell prolymphocytic leukemia
	Lymphoplasmacytic lymphoma/immunocytoma
	Mantle cell lymphoma
	Follicular lymphoma
	Extranodal marginal zone B-cell lymphoma of mucosa-associated lymphoid tissue (MALT type)
	Nodal marginal zone B-cell lymphoma
	Splenic marginal zone lymphoma
	Hairy cell leukemia
	Plasma cell myeloma/plasmacytoma
	Diffuse large B-cell lymphoma
	Burkitt's lymphoma/Burkitt's cell leukemia
T-cell and NK-cell neoplams	
Precursor T-cell neoplasm	Precursor T-lymphoblastic leukemia/ lymphoma
Peripheral (mature) T-cell neoplasm	T-cell prolymphocytic leukemia
	T-cell granular lymphocytic leukemia
	Aggressive NK-cell leukemia
	Adult T-cell lymphoma/leukemia (HTLV+)
	Extranodal NK/T-cell lymphoma
	Enteropathy-type T-cell lymphoma
	Hepatosplenic T-cell lymphoma
	Subcutaneous panniculitis-like T-cell lymphoma
	Mycosis fungoides/Sezary's syndrome
	Peripheral T-cell lymphoma, unspecified
	Angioimmunoblastic T-cell lymphoma
	Anaplastic large cell lymphoma
Hodgkin's lymphoma	
Nodular lymphocyte predominance Hodgkin's lymphoma	
Classical Hodgkin's lymphoma	

Table 1: WHO classification of lymphoid neoplasms.

Both the classical and the alternative NF-κB pathways play a crucial role for the survival of T and B lymphocytes from early lymphopoiesis to later stages of development and maturation. It is therefore not surprising to see that aberrant, persistent activation of NF-κB in humans has been reported in various autoimmune diseases and in a number of lymphoid malignancies (reviewed in

Keutgens *et al*., 2006) including acute myelogenous leukemia (AML) (Guzman *et al*., 2001; Baumgartner, 2002), acute lymphoblastic leukemia (ALL) (Kordes *et al*., 2000), Hodgkin Reed Sternberg cells (Thomas *et al*., 2004), B-cell lymphomas of mucosa-associated lymphoid tissue (MALT) (Zhou *et al*., 2004; Zhou *et al*., 2005), multiple myeloma (Mitsiades *et al*., 2002),… as well as in the bone marrow of patients with myelodysplastic syndrome (Braun *et al*., 2006).

In this work, we are interested in a product of an *nfκb2* gene rearrangement, Hut78/p80HT (also referred as to p85), identified the HUT78 cutaneous T-cell lymphoma cell line. *nfκb2* rearrangements can be found at low frequency (1 to 2% of cases) in B-cell non Hodgkin's lymphoma, chronic lymphocytic leukemia (B-CLL), multiple myeloma and more commonly (14%) in cutaneous T-cell lymphomas (CTCL), such as mycosis fungoides (MF) and Sezary's syndrome (SS). These latter are low-grade, indolent lymphomas, characterized by infiltration of the skin by malignant mature T-lymphocytes, which present usually a mature CD4+ phenotype. After a variable period of time, mycosis fungoides typically evolve from the early patch/plaque stage to the eventual development of cutaneous tumors, as well as nodal and visceral involvement. Alternatively, evolution to diffuse erythroderma, which is almost invariably accompanied by the presence of neoplastic T-lymphocytes in the peripheral blood, is known as Sezary's syndrome (SS). They typically occur in middle-aged adults and a wide variety of therapies can produce responses in MF/SS, such as topical nitrogen mustard, PUVA or total skin electron beam, but are generally not curative (Diamandidou *et al*., 1996).

nfκb2 rearrangements

The NF-κB2 genomic locus is localized on the human chromosome 10q24 and the *nfκb2* gene was originally cloned from a chromosomal translocation t(10;14) (q24;q32) resulting in its juxtaposition to the immunoglobulin heavy chain locus (IgHCα), in a case of B-cell non Hodgkin's lymphoma (Neri *et al*., 1991). Contrarily to the chromosomal translocation t(14;19) (q32;q13.1) which drives the proto-oncogene *bcl3* under the dependence of a strong promoter, resulting in an intact, but overexpressed BCL-3 protein, chromosomal rearrangements involving the *nfκb2* gene generate oncogenic chimeric proteins. The breakpoints all cluster within the 3'*ankyrin*-encoding portion of the *nfκb2* gene and the sequence 3' to the breakpoint is substituted by sequences from different chromosomes. Depending on this sequence, the reading frame may be

interrupted by the presence of a stop codon at (LB40) or in the close vicinity (Hut78) to the breakpoint, or remains open, resulting in the production of a fusion protein where the N-terminal domain of p100 is fused to a heterologous domain (Lyt-10 Cα, EB308). **Figure 4** provides a schematic representation of the proteins encoded by these rearrangements.

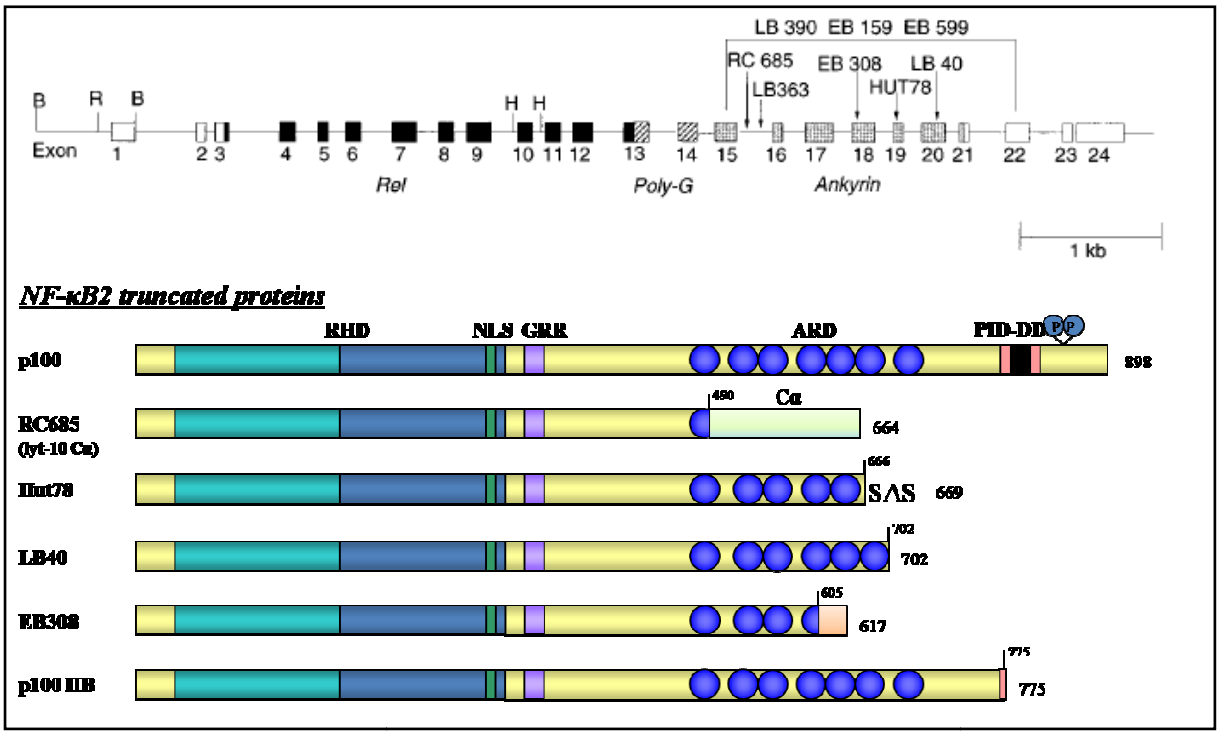


Figure 4: Position of chromosomal breakpoints within the *nfκb2* gene (*from* Chang *et al*., *Molecular and cellular biology*, 1995, 5180-5187) **and schematic representation of the corresponding truncated NF-κB2 proteins**: rearrangements cluster within the 3'coding portion of the gene corresponding to the ankyrin domain and lead to the production of C-terminally truncated proteins which lack variable portions of the ankyrin domain. The number of amino acids of each protein is indicated on the right. The number of amino acids derived from the normal NF-κB2 sequence is also shown for each protein.

♦ ***lyt-10Cα (RC685)*** was identified in the B-cell lymphoma associated translocation t(10;14) (q24;q32) which juxtaposes the immunoglobulin Cα locus to the NF-κB2 locus, resulting in the production of a chimeric protein in which the first ankyrin repeat of p100 is fused to a 174 amino acid tail derived from an off-frame immunoglobulin heavy chain Cα coding region (Neri *et al*., 1991). This mutant can transform mouse immortalized fibroblasts (Ciana *et al*., 1997). However, it failed to transform human lymphoblastoid cell lines, probably due to an apparent cytotoxic effect.

♦ ***Hut78***, which was identified in a cutaneous T-cell lymphoma (CTCL), is composed of the first 666 amino acids of p100 fused to three extra amino acids (serine-alanine-serine) (Thakur *et al*., 1994; Zhang *et al*., 1994).

♦ ***LB40*** harbors a stop codon in the sixth ankyrin repeat and has been identified in a case of B-cell chronic lymphocytic leukemia (B-CLL) (Migliazza *et al*., 1994).

♦ ***EB308*** results from the fusion of three and half ankyrin repeats from p100 to 12 amino acids from an unknown locus and was observed in a cutaneous T-cell lymphoma (CTCL) (Migliazza *et al*., 1994).

♦ ***p100HB*** is a 775 amino acids protein detected in some human tumor cell lines, originating from a point mutation that generates a premature stop codon at position 2576 of the p100 sequence (Derudder *et al*., 2003).

These NF-κB2 mutants escape the physiological regulatory mechanisms of NF-κB2 proteins: in contrast to p100, mutant proteins are primarily localized in the nucleus, present a defective IκB-like function and undergo constitutive processing, resulting in an enhanced production of p52 (except for p100HB). The common feature of these *nfκb2* rearrangements is the production of C-terminally truncated proteins lacking the PID and variable portions of the ankyrin domain, with or without fusion to heterologous tails, suggesting that C-terminal deletion may be the general mechanism underlying the constitutive activation of NF-κB2 *in vivo*.

The recurrent association between *nfκb*2 rearrangements and lymphoid neoplasms suggest that these alterations play a role in lymphomagenesis. However, the mechanisms by which these *nfκb2* gene rearrangements contribute to the development of lymphomas remain unclear. In this work, we try to gain insights into the molecular mechanisms underlying the oncogenic potential of the truncated NF-κB2 mutants, through the identification of the target genes of one of them, the Hut78 protein.

❖ ***Hut78***

The HUT78 cell line is derived from a patient with a type of cutaneous T-cell lymphoma (CTCL) called Sezary's syndrome (Gazdar *et al*., 1980; Mann *et al*., 1989). The karyotype of HUT78 cells shows a single copy of an abnormal chromosome 10 (Saglio *et al*., 1986), indicating that one allele is rearranged while the other one is lost. The rearrangement leads to the production of a large

amount of a truncated NF-κB2 protein, Hut78 (p80HT, p85), as well as the processed p52, while no wild type p100 could be detected in these cells (Thakur *et al*., 1994; Zhang *et al*., 1994). Such rearrangement was also observed in two CTCL patients, but contrarily to the HUT78 cell line, the normal allele was retained in addition to the rearranged allele.

The sequence 3' to the breakpoint of the NF-κB2 locus is substituted by a sequence from an undetermined origin, which contains multiple Alu repetitive sequences. The resulting Hut78 protein corresponds to the 666 first amino acids of the wild-type p100 linked to three extra amino acids (SAS) prior to a premature stop codon. This protein retains therefore the first five ankyrin repeats, but lacks a part of the sixth, the entire seventh and the processing inhibitory domain (PID). The functional properties of this abnormal protein may account for its role in tumorigenesis:

♦ ***nuclear localization***: conversely to p100 which is cytoplasmic, Hut78 is located in the nucleus like p52 (Thakur *et al*., 1994; Zhang *et al*., 1994; Chang *et al*., 1995). Loss of the C-terminal domain allows thus the protein to escape the cytoplasmic retention and to translocate into the nucleus without further proteolytic processing. The consequence of this nuclear localization may be a constitutive activation of NF-κB-mediated transcription, which may play a role in transformation (Zhang *et al*., 1994).

♦ ***altered transcriptional activity***: p52 can activate transcription when associated with p65 or BCL-3, and repress transcription by forming inactive DNA-bound homodimers. Hut78 was shown to bind κB sites in an unprocessed form and to maintain the ability to mediate transcriptional activation via heterodimerizing with p65. However, conversely to p52, Hut78 appears to have acquired intrinsic transactivation properties on some promoters and lost the transrepression function associated with homodimer formation (Zhang *et al*., 1994; Chang *et al*., 1995).

♦ ***loss of the IκB-like, inhibitory activity***: loss of the potent inhibitory IκB-like activity of p100 does appear to be a common feature of the NF-κB2 C-terminal truncations associated with tumors (Kim *et al*., 2000). These proteins no longer exhibit the strong inhibition of NF-κB-mediated transcriptional activation associated with p100 wild type, which possibly leads to enhanced constitutive NF-κB activation in tumor cells. As transcription of NF-κB2 is autoregulated (Liptay *et al*., 1994; Lombardi *et al*., 1995), the NFκB2 promoter could itself represent a target promoter whose expression is altered by the loss of the IκB-

like properties. This pathway would explain the very high levels of NFκB2 mRNA and protein observed in the HUT78 cell line.

♦ ***constitutive processing***: deletion of the processing inhibitory domain, whose core region corresponds to a death domain (DD) is sufficient to trigger constitutive processing of p100 (Xiao *et al*., 2001). This latter, contrarily to the p100 inducible processing, seems to be independent of the $SCF^{\beta\text{-}TrCP}$ ubiquitin ligase (Fong and Sun, 2002) and regulated by nuclear translocation of the protein (Liao and Sun, 2003). Xiao and colleagues also demonstrated that binding of p100ΔCs to promoter DNA-containing κB sites is essential for their constitutive proteasomal-mediated processing and that the processed p52 plays a role in oncogenic transformation through deregulating expression of tumor-associated NF-κB target genes (Qing *et al*., 2007). Moreover, genetically manipulated mice lacking p100, but still containing a functional p52 protein, develop gastric and lymphoid hyperplasia (Ishikawa *et al*., 1997), which supports the idea that deregulated processing of p100 and generation of p52 may contribute to the development of lymphoid malignancies.

Recently, the generation of transgenic mice expressing Hut78 in lymphocytes, provided direct evidence of the oncogenic activity of Hut78 *in vivo*, which constitutes the first demonstration of a causal role for *nfκb2* mutation in lymphomagenesis (Zhang *et al*., 2007). The transgenic mice display a marked expansion of peripheral B-cell population and develop small B-cell lymphomas with extensive multi-organ metastases. This study shows that Hut78 promotes lymphomagenesis by inducing TRAF1 (TNF receptor associated factor1), which in turn suppresses specific apoptotic responses critical for the maintenance of lymphocyte homeostasis. However, the tumorigenic activity of Hut78 does not seem to be totally dependent on p52 production, as mice with targeted overexpression of p52 in their lymphocytes do not develop lymphomas, but are predisposed to inflammatory auto-immune diseases and show only transient upregulation of TRAF1 (Wang *et al*, 2008). The lymphoma development seen in Hut78 mice is characterized by a prolonged latent period, which suggests that additional mutations may be required for malignant transformation of Hut78 lymphocytes. This hypothesis is supported by the observation that the HUT78 cell line also harbors other alterations, for example in *c-myc* (Finger *et al*., 1988), *c-cbl* (Blake and Langdon, 1992) and p53 (Tolomeo *et al*., 1998). Therefore NF-κB2 mutations are likely to be only one of several mutations involved in the generation of human T or B cell lymphomas.

Matrix metalloproteinases: MMPs

Matrix metalloproteinases are a family of zinc-dependent proteolytic enzymes originally described to degrade the components of the extracellular matrix (ECM) including collagen, laminin, fibronectin and proteoglycans (Nagase and Woessner, 1999). In addition to this role in ECM homeostasis, the identification of new MMP substrates such as cytokines, latent growth factors, cell-matrix adhesion molecules and cell-surface receptors, and the development of genetically modified animal models with gain or loss of MMP function, have demonstrated the relevance of these proteases in multiple physiological processes (Chakraborti *et al*., 2003; McCawley and Matrisian, 2001; Sternlicht and Werb, 2001; Egeblad and Werb, 2002; Vu and Werb 2000). MMPs are well-known mediators of cell migration, invasion and proliferation, which, in normal physiology, are involved in embryonic development, wound healing, tissue remodeling, angiogenesis, ovulation, macrophage and neutrophil function,… (Van den Steen, 2002; Yoon, 2003). Deregulation of MMPs activity plays a role in pathologies, often characterized by the excessive degradation of ECM, such as rheumatoid arthritis, osteoarthritis, periodontal diseases, auto-immune diseases, gastric ulcer, arteriosclerosis, cardiovascular diseases, tumor invasion and metastasis (Westermarck and Kahari, 1999; Vu and Werb, 2000).

Structure and classification

In humans, the MMP family currently consists of 25 secreted or transmembrane enzymes, synthesized as pro-zymogens which acquire their active form after proteolytic cleavage, thus allowing for local degradation of their substrates. Matrix metalloproteinases belong to the metzincin superfamily, a family of zinc endopeptidases, characterized by a highly conserved motif containing three histidines that bind to zinc at the catalytic site, followed by an invariant methionine involved in a structural feature called the “Met-turn” (Stocker and Bode, 1995). The MMPs share a common basic structure, made up of the following homologous domains (reviewed in Bode *et al*., 1999; Sternlicht and Werb, 2001):

♦ ***signal peptide*** or ***predomain***: contains a leader sequence, which directs MMPs to the secretory or plasma membrane insertion pathway.

♦ ***propeptide*** or ***prodomain***: contains a well conserved cysteine residue which ligates the active site zinc ion, making the catalytic enzyme inaccessible to the

substrate and conferring latency to the proMMP (Springman *et al.*, 1990; Van Wart and Birkedal-Hansen 1990).

♦ ***catalytic domain***: contains a zinc-binding sequence characteristic of the metzincin superfamily.

♦ ***hemopexin-like C-terminal domain***: mediates interactions with endogenous inhibitors and substrates, conferring specificity to the enzyme.

♦ ***hinge region***: rich in proline residues, links the catalytic and the hemopexin domains (except for MMP 7, 23 and 26).

Depending on their substrate specificity, MMPs were traditionally subdivided into collagenases, stromelysins and gelatinases, but recently, new classifications, based on structure and cellular location, have been proposed (Egeblad and Werb, 2002; Brinckerhoff and Matrisian, 2002) (**Figure 5**):

❖ ***Secreted MMPs***

♦ ***minimal domain MMPs*** (MMP 7 and 26): they lack the hemopexin-like domain (Muller *et al*., 1988; de Coignac *et al*., 2000), but yet display substrate degradation and are known as matrilysins.

♦ ***simple hemopexin domain-containing MMPs*** **(**MMP 1, 3, 8, 10, 12, 13, 19, 20 and 27): this subgroup of archetypal MMPs is composed of three collagenases (MMP 1, 8 and 13), two stromelysins (MMP 3 and 10) and four additional MMPs with unique structural characteristics (MMP 12, 19, 20 and 27).

♦ ***gelatin-binding MMPs*** (MMP 2 and 9): the gelatinases contain in their catalytic domain, a fibronectin-like insert organized into three internal repeats, homologous to the type II motif of the collagen binding domain of fibronectin and required for binding and cleaving collagen and elastin (Collier *et al*., 1988; 1992; Murphy *et al*., 1994; Shipley *et al*., 1996). In addition, MMP9 possesses at one end of the catalytic domain, a unique 54 amino acid long proline rich domain, homologous to the α2 chain of type V collagen (Wilhelm *et al*., 1989).

♦ ***furin-activated secreted MMPs*** (MMP 11 and 28): they contain a recognition motif furin-like serine proteinases between the propeptide and the catalytic domain, allowing intracellular activation prior to the secretion (Thomas 2002; Zucker *et al*., 2003).

♦ ***vitronectin-like insert MMP*** (MMP 21): MMP 21 also contains a furin-cleavage site and a domain related to vitronectin, not found in other MMPs.

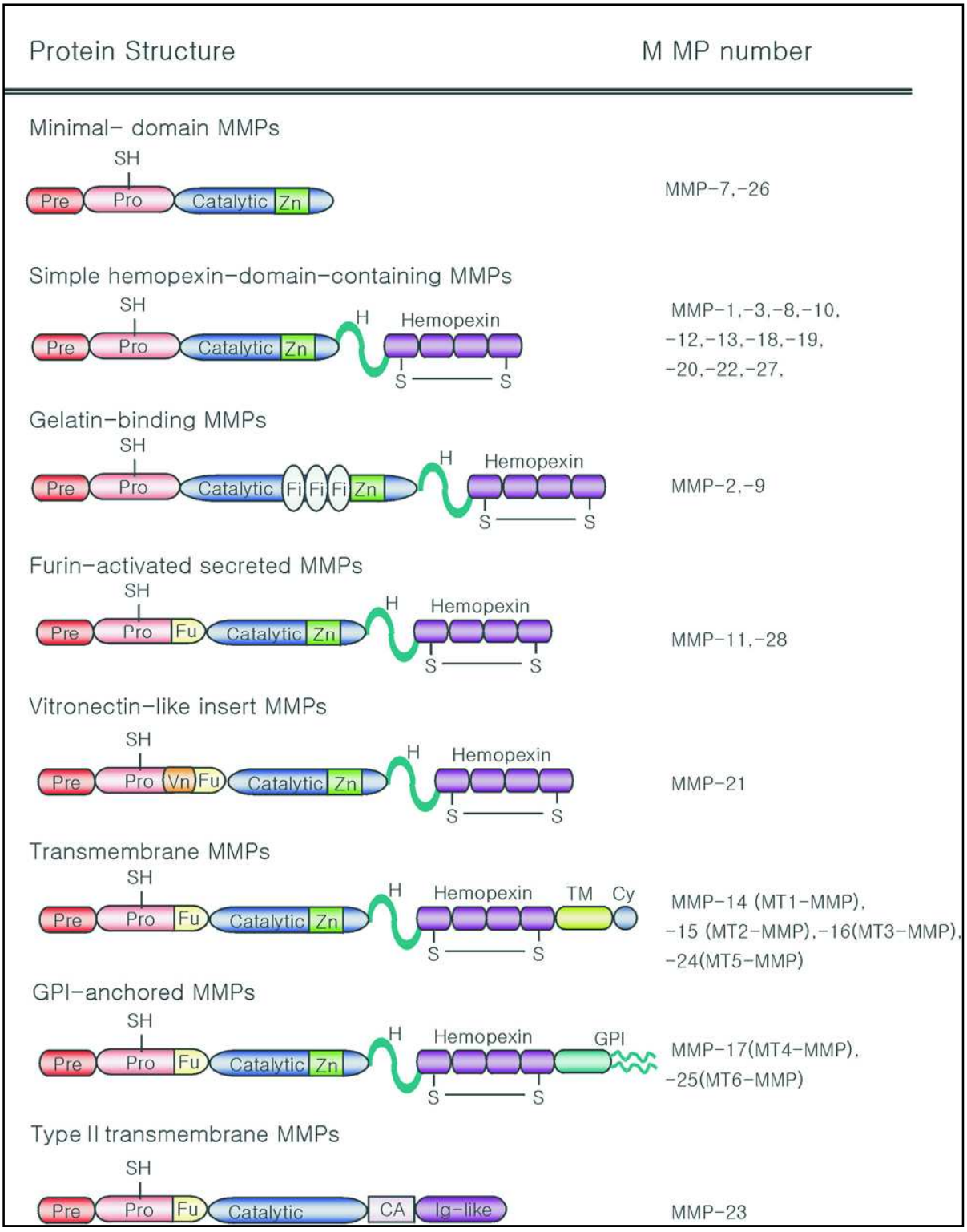


Figure 5: Protein structures of MMPs. Pre: predomain or signal peptide, **Pro**: propeptide, **H**: hinge region, **Hemopexin**: hemopexin-like domain, **Fu**: furin susceptible site, **TM**: transmembrane domain, **Cy**: cytoplasmic domain, **GPI**: glycophosphatidyl inositol-anchoring domain, **SA**: signal anchor domain, **CA**: cysteine array domain, **Vn**: vitronectin-like domain, **Ig-like**: immunoglobulin-like domain. *From* Yoon *et al.*, *Journal of Biochemistry and Molecular Biology*, Vol. 36, No. 1, January 2003, pp. 128-137.

❖ ***Membrane-bound MMPs***

♦ **(*type I) transmembrane MMPs*** (MMP 14, 15, 16 and 24): they posses a transmembrane domain and a small cytoplasmic tail at the C-terminus.

♦ ***glycosyl-phosphatidyl inositol (GPI)-linked MMPs*** (MMP 17 and 25): these MMPs are anchored to the membrane by a C-terminal hydrophobic region with a glycosyl-phosphatidyl inositol (GPI) domain (Itoh *et al*., 1999; Kojima *et al*., 2000).

♦ ***type II transmembrane MMPs*** (MMP 23): instead of the conserved hemopexin-like domain, MMP23 possesses a cysteine-array (CA), an immunoglobulin-like-domain, as well as an amino-terminal signal anchor targeting it to the cell membrane (Pei *et al*., 2000).

All these transmembrane MMPs also contain a furin-cleavage site for intracellular activation.

MMPs in tumor invasion and metastasis

Like embryonic growth and tissue morphogenesis, tumor invasion and metastasis require the disruption of ECM barriers to allow cell migration and matrix microenvironment remodeling. Because of their ability to degrade structural components of ECM and basement membranes, MMPs were initially thought to be directly involved in these processes. This idea was supported by a correlation between high levels of MMP activity, in particular MMP2 (gelatinase A) and MMP9 (gelatinase B), and increased tumor metastasis (Johnsen, 1998; Ohashi *et al*., 2000; Kurahara *et al*., 1999; Bando *et al*., 1998; Kanayama *et al*., 1998). Tumor metastasis is a multistep process, in which a subset or individual cancer cells escape form a primary tumor and disseminate to distant, secondary organs or tissues. This includes disconnection of intercellular adhesions and separation of single cells from solid tumor tissue, proteolysis of the extracellular matrix, migration and invasion of the surrounding tissues and lymph- and blood vessels, immunologic escape in the circulation, adhesion to endothelial cells, extravasation from lymph- and blood vessels, proliferation and induction of angiogenesis (Bohle and Kalthoff, 1999; Nash *et al*., 2002; Yoon *et al*., 2003; Deryugina and Quigley, 2006). MMPs expressed by tumors and stromal cells have been shown to be involved in many steps in tumor metastasis, such as tumor invasion, migration, host immune escape, extravasation, angiogenesis and tumor growth. However, the role of MMPs in cancer

progression is much more complex than that derived from their direct degradative action on ECM components (Egeblad and Werb 2002; Freije *et al*., 2003; Hojilla *et al*., 2003; Folgueras *et al*., 2004). Besides degrading ECM, MMPs target other substrates like growth-factor receptors, cell adhesion molecules, chemokines, cytokines, apoptotic ligands and angiogenic factors, which can exhibit pro-metastatic, as well as anti-metastatic activities. The opposite effects of bioactive molecules processed by MMPs are also involved in earlier stages of tumor evolution and are illustrated in **Figure 6**.

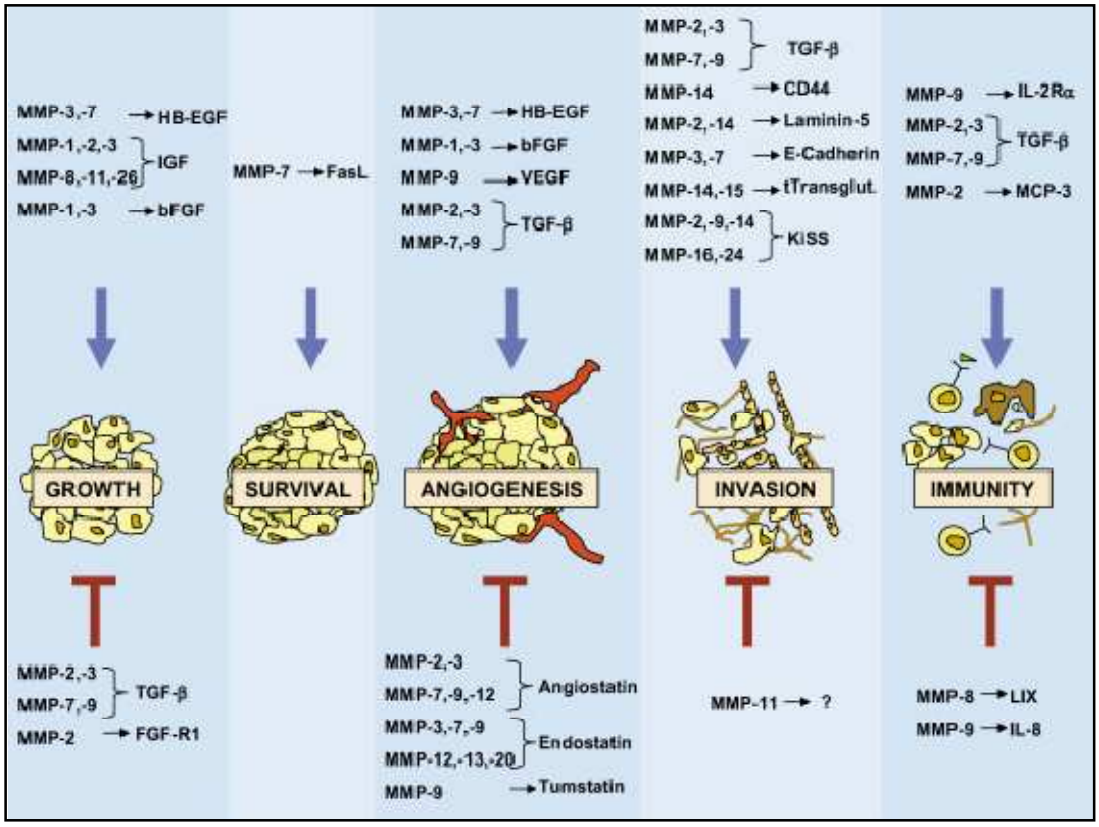


Figure 6: Dual functions of MMPs in tumour progression. MMPs activate many bioactive molecules which were shown to have opposite effects on cancer development. *From* Folgueras *et al*., **Int. J. Dev. Biol. 48**: 411-424 (2004).

The diversity and complexity of MMP functions in cancer rely on the substrate specificity and the regulation of each individual MMP, certain MMPs having dual effects on tumor development (Andarawewa *et al*., 2003). The generation of genetically modified animal models of gain or loss of MMP function allowed a better understanding of the biology of each MMP. We will next focus on the roles of MMP9 in cancer as we observed upregulation of this protease in murine fibroblasts overexpressing p52 and/or Hut78.

MMP9

The type IV collagenase MMP9, also called gelatinase B, is synthesized as a 78,4 kDa prepropeptide and secreted as a glycosylated 92 kDa proenzyme (Wilhelm *et al*., 1989), which yields an active MMP9 enzyme of 82 kDa after proteolytic activation (Ogata *et al*., 1992). Macrophages and neutrophils constitute the major source of MMP9, but it is also produced by stromal cells, endothelial cells and osteoclasts.

The absence of distinct phenotypes in most genetically-engineered mice with knockdown of specific MMPs finding suggests functional redundancy among the different MMPs (Ny *et al*., 2002; Solberg *et al*., 2003). However, studies with MMP9-deficient mice have demonstrated the *in vivo* role of this protease in a number of developmental processes. MMP9 is required for the release of vascular endothelial growth factor (VEGF), a chemoattractant for osteoblast recruitment (Engsig *et al*., 2000), which explains the defect in long bone development observed in MMP9-null mice (Vu *et al*., 1998). MMP9 also contributes to adipogenesis by promoting adipocyte differentiation (Bouloumie *et al*., 2001). MMP9-deficient mice exhibit lower levels of apoptosis induced by TNF-α (Wielockx *et al*., 2001), resulting in a delayed apoptosis of hypertrophic chondrocytes at the skeletal growth plates and vascularization (Vu *et al*., 1998).

In many cancers, high serum levels of MMP9 are positively correlated with rapid tumor progression, metastasis and poor prognosis (Nakajima *et al*., 1993; Johnsen *et al*., 1998; Nikkola *et al*., 2005). However, whether elevated levels of MMPs actually results in more functional activity of the enzymes during cancer progression or simply reflects their deregulated expression remain to be determined as soluble MMPs are found mainly in the proenzyme form (Deryugina and Quigley, 2006). MMP9 deficiency produces a substantial inhibitory impact in different metastasis models (Itoh *et al*., 1999; Bergers *et al*., 2000; Coussens *et al*., 2000; Huang *et al*., 2002; Acuff *et al*., 2006) and the majority of effects observed have been related to reduced or altered angiogenesis. Therefore, the pro-angiogenic roles of MMP9 appear to be the most critical aspect in the functional contribution of host MMP9 to tumor progression.

- ***MMP9 and angiogenesis***

The formation of capillaries irrigating tumor cells is necessary to sustain solid tumor growth, by allowing gas exchange and nutrient supply (Carmeliet and Jain, 2000). Several studies involving MMP9-deficient mouse models demonstrated that MMP9 is mainly supplied by the tumor-associated inflammatory cells and plays a key role in tumor-induced angiogenesis, without affecting the development of normal vasculature (Bergers *et al*., 2000; Bergers and Benjamin, 2003; Huang *et al*., 2002; Hiratsuka *et al*., 2002; Jodele *et al*., 2005; Chantrain *et al*., 2004; Riedel *et al*., 2000; Coussens *et al*., 2000; Zijlstra *et al*., 2004; Zijlstra *et al*., 2005). The MMP9 pro-angiogenic effect can be mediated by activation of the vascular endothelial growth factor (VEGF) and/or the transforming growth factor TGFβ, triggering the switch from vascular quiescence to angiogenesis (Bergers *et al*., 2000; Giraudo *et al*., 2004; Mira *et al*., 2004; Aalinkeel *et al*., 2004). However, MMP9 can also inhibit angiogenesis by cleaving the ECM components into anti-angiogenic factors, like angiostatin, generated from plasminogen, endostatin derived from the proteolysis of type XVIII collagen and tumstatin, the NC1 domain of α3 chain of collagen IV (Ferreras *et al*., 2000; Hamano *et al*., 2003; Heljasvaara *et al*., 2005; Pozzi *et al*., 2000). A recent study has correlated the decreased levels of tumstatin in MMP9 knock-out mice with the increased angiogenesis and accelerated tumour growth (Hamano *et al*., 2003). Therefore, the dual role of MMP9 in angiogenesis contributes to regulate the "angiogenic balance" in tumor development.

- ***MMP9 and the immunologic escape in the circulation***

Once in the circulation, the intravasated tumor cells have to acquire certain mechanisms to avoid their recognition and elimination by immunocompetent cells. By disrupting the IL-2Rα signaling, which is involved in the development and propagation of T-cells, MMP9 suppresses the proliferation of the encountered T-lymphocytes and thus contributes to this immunologic escape of the tumor cells (Sheu *et al*., 2001). MMP9 also activates TGFβ (Yu and Stamenkovic, 2000), an important inhibitor of the T-lymphocytes response against tumors (Gorelik and Flavell, 2001). Moreover, the MMP9-dependent ICAM1 release mediates the resistance of breast cancer cells to natural killer (NK) cell-mediated cytotoxicity.

Regulation of MMP9

The regulation of MMPs occurs at multiple levels:

- ***Transcriptional regulation***

MMP transcripts are generally expressed at low levels, but these levels rise rapidly in response to extracellular stimuli, such as cytokines, growth factors or oncogenes (Westermarck and Kahari, 1999). The MMP expression is cell-specific and MMP induction mechanisms appear to be different depending on the characteristics of the diverse cells with ability to produce these enzymes. Analysis of the different MMP promoters reveals the presence of a variety of functional *cis*-acting elements allowing for the regulation of *mmp* genes expression by a subset of *trans*-activators, like AP-1 (activator protein 1), PEA3 (polyomavirus enhancer A-binding protein 3), Sp1, β-catenin/Tcf4 or NF-κB. Based on their basic conformation, the MMP promoters are divided into three groups (Yan and Boyd, 2007):

- TATA box at around -30bp and AP-1 site at approximately -70bp (relative to the transcription start site): MMP 1, 3, 7, 9, 10, 12, 13, 19, and 26. Most of these promoters harbor also an upstream PEA3 binding site (Benbow and Brinckerhoff, 1997).
- TATA box, but no proximal AP-1 site: MMP 8, 11 and 21.
- No TATA box, nor proximal AP-1 site: MMP 2, 14 and 28.

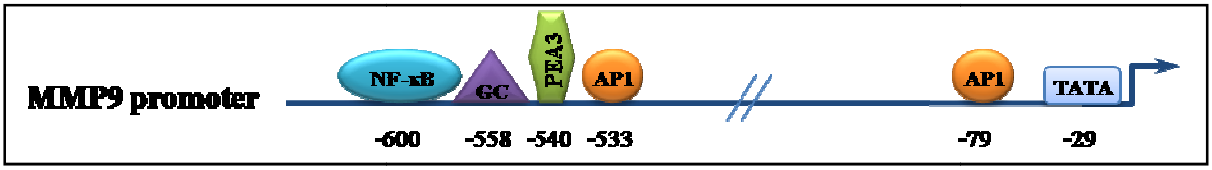


Figure 7: Schematic representation of the *cis*-elements of the human MMP9 promoter (Data from Yan and Boyd, J. Cell. Physiol. 211: 19–26, 2007). **AP-1**: Activator protein 1, **GC**: GC-box, **PEA3**: polyoma enhancer A binding protein-3.

The MMP9 promoter (**Figure 7**) belongs to the first group, which renders the *mmp9* gene responsive to the dimeric AP-1 complex, composed of Jun and Fos proteins. Activation of AP-1 by extracellular stimuli is mediated by three classes of mitogen-activated protein kinases (MAPKs): the mitogen-activated extracellular signal-regulated kinases 1, 2 (ERK 1, 2), the stress-activated Jun N-terminal kinases and the p38 MAPK (Westermarck and Kahari, 1999; Cho and Choi, 2002). The PEA3 site is another *cis*-element present in the MMP9

promoter, which binds the Ets family of oncoproteins (Crawford *et al*., 2001) and can cooperate with AP-1 to enhance transcription (Benbow and Brinckerhoff, 1997; Westermarck and Kahari, 1999; Chakraborti *et al*., 2003). AP-1 and/or PEA3 represent a convergence point for regulation of MMP expression by a variety of cytokines, growth factors or ECM proteins (EGF, VEGF, TNFα, TGFβ,…) (Yan and Boyd, 2007). The GC box binds Sp1 and Sp3, and potentially other GC-binding proteins. The NF-κB signaling pathway also mediates the synergistic induction of MMP9 expression by inflammatory cytokines and growth factors (Bond *et al*., 1998), through the presence of a NF-κB binding site located at -600 bp in the human MMP9 promoter (He, 1996). Therefore, the inhibition of NF-κB was shown to down-regulate MMP9 and reduce tumor invasion (Yoon *et al*., 2002). MMP9 is also regulated by Runx2, which fits with its role in endochondral ossification (Ortega *et al*., 2004).

Epigenetic modifications such as DNA methylation and chromatin remodeling, also contribute to the regulation of MMP genes expression. Hypermethylation of cytosines within the CpG dinucleotides in a promoter region is usually associated with inhibition of gene expression (Fuks, 2005). In lymphoma cell lines, a correlation was shown between silencing of MMP9 promoter and hypermethylation at its promoter (Chicoine *et al*., 2002).

The roles of the different histone post-translational modifications in remodeling the structure of chromatin in order to modulate access of a promoter to the basal transcriptional machinery will be further discussed in the next chapter. However, we can mention here some studies suggesting that epigenetic mechanisms contribute to the control of *mmp9* gene expression. For example, MTA1, a component of the NuRD (nucleosomal remodeling and histone deacetylation) repressor complex was shown to recruit HDAC2 and the chromatin-remodeling enzyme Mi-2 to the MMP9 promoter, repressing MMP9 expression (Yan et *al*., 2003). Additionally, PMA-induced MMP9 expression involves the recruitment of AP-1, NF-κB, Sp1 and Brg1, the ATPase subunit of the Swi/Snf complex on the MMP9 promoter, removing HDAC1- and HDAC3-containing complexes and leading to modifications in histone acetylation, phosphorylation and methylation (Ma *et al*., 2004). This PMA-induced MMP9 expression is inhibited by MTA, a specific protein methyltransferase inhibitor, and enhanced by HDAC inhibitors (Ma *et al*., 2004; Martens *et al*., 2003; Clark *et al*., 2008).

- ***Proenzyme activation***

MMPs are synthesized as inactive zymogens and therefore, activation of proMMPs represents another step in the regulation of MMP activity. *In vivo*, MMPs are proteolytically activated in the pericellular space by tissue or plasma proteinases, bacterial proteinases or other MMPs.

- ***Endogenous inhibitors***

The activity of MMPs may be also controlled by a series of endogenous inhibitors. Four specific tissue inhibitors of metalloproteinases (TIMPs) have been identified in vertebrates (Brew *et al*., 2000): TIMP-1, TIMP-2 and TIMP-4 are secreted proteins whereas TIMP-3 is anchored in the ECM. They all reversibly inhibit MMPs in a stoechiometric manner and the balance between MMPs and TIMPs determines the proteolytic potential of tumors (Folgueras *et al*., 2007).

In addition to TIMPs, MMP functions may be regulated by other protease inhibitors, like the α1-proteinase inhibitor and α2-macroglobulin, the procollagen C terminal protease enhancer (Mott *et al*., 2000), the NC1 domain of type IV collagen (Petitclerc *et al*., 2000) and the tissue factor pathway inhibitor-2 (Herman *et al*., 2001). RECK (reversion-inducing cysteine-rich protein with kazal motifs) is a membrane-anchored protein that inhibits the secretion of pro-MMP9 as well as the MMP2, MMP9, and MT1-MMP catalytic activity (Takahashi *et al*., 1998; Oh *et al*., 2001).

The Histone Code

In eukaryotes, DNA is compacted into the nucleus as chromatin. The fundamental repeating unit of the eukaryotic chromatin is called the nucleosome and consists of 146bp of DNA wrapped around an octamer core of pairs of histone proteins (H4, H3, H2A an H2B) (Kornberg, 1974). The histone H1 binds to the linker DNA region between the histone beads and facilitates the formation of higher-order structures of chromatin (**Figure 8A**).

The structure of the chromatin determines the accessibility of DNA to various enzymes implicated in transcription, replication, recombination and repair process. For example, an open, diffuse chromatin structure (euchromatin) is mostly associated with actively transcribed genes, whereas a highly condensed chromatin (heterochromatin) prevents the transcription machinery to gain access to the DNA template (Weintraub *et al*., 1976) (**Figure 8B**). Chromatin structure can be modulated by different mechanisms, such as methylation of cytosine residues in the CpG islands of the genome and histones modifications, which we will discuss here. These mechanisms often operate in a coordinated way on a given locus and constitute the "*epigenetic*" regulation which refers to an inherited state of regulation independent of the genetic information encoded within DNA itself (Shilatifard, 2006).

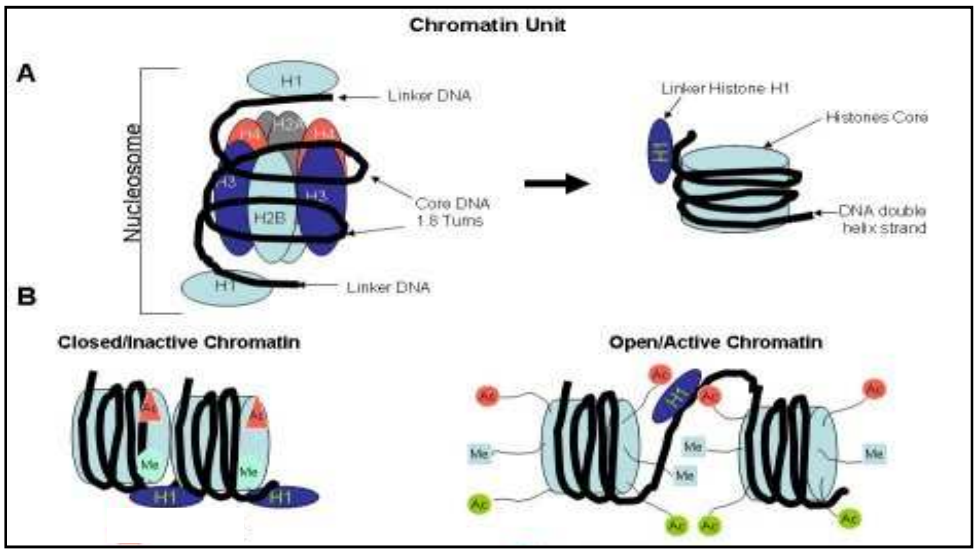


Figure 8: A. **The nucleosome structure**: the DNA is wrapped around the histone octamer, which consists of two copies of H2A, H2B, H3 and H4. H1 is the linker histone **B**. **Histone modifications have been associated with either 'active' or 'inactive' chromatin states. Ac**: acetyl- ; **Me**: methyl-. *From* http://www.abcam.com.

The "histone code"

Histone proteins have a globular C-terminal domain critical to nucleosome formation and a flexible N-terminal tail that extends away from the nucleosome core, providing an attractive signaling platform (Santos-Rosa and Caldas, 2005). These amino-terminal tails of histone proteins are subject to post-translational covalent modifications such as *acetylation* (lysine), *methylation* (lysine, arginine), *phosphorylation* (serine, threonine), *ubiquitination* (lysine) and *SUMOylation* (**Figure 9**). These modifications occur at selected residues and can be recognized by enzymes that affect the status of other residues, including those on neighboring histones. The precise pattern of the histone marks in the chromatin surrounding a gene constitutes the "*histone code*" and controls chromatin structure and regulation of gene expression through recruitment of downstream effector proteins containing a specific interacting domain (chromodomain, bromodomain) (Strahl and Allis, 2000; Santos-Rosa and Caldas, 2005).

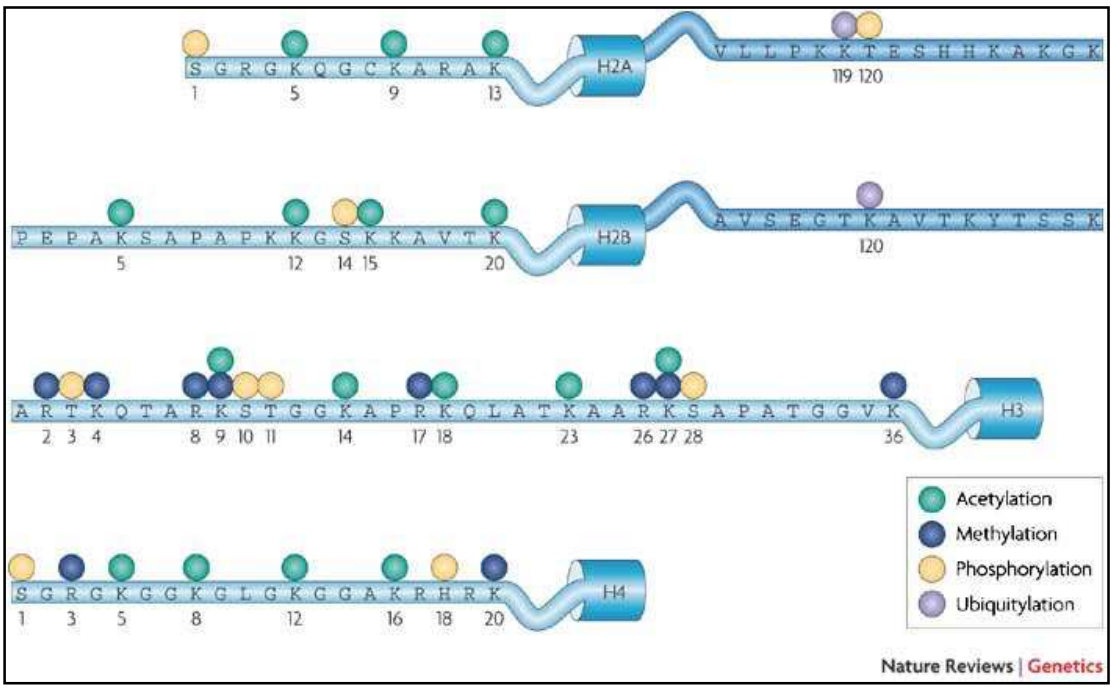


Figure 9: Histone post-translational modifications (methylation, acetylation, phosphorylation, and ubiquitination): they occur on specific residues, which are generally located on the aminoterminal tail of histones. *From* Spivakov and Fisher, *Nature Reviews Genetics* (2007) 8, 263-271.

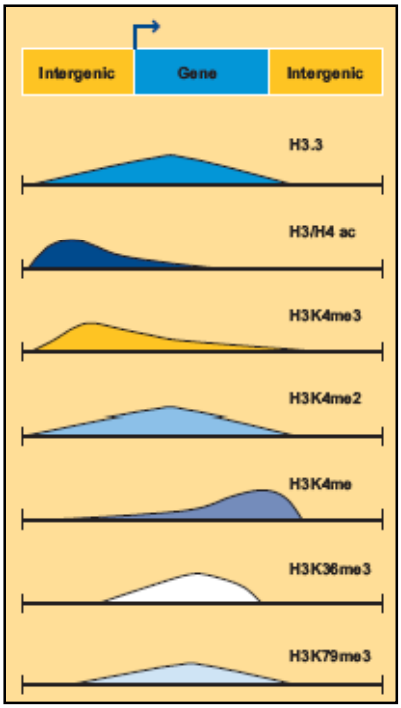


Figure 10: **Gradients of histone H3 methylation and acetylation in active genes**. A schematic representation of active gene is shown. The extent and level of modification are depicted by the height of the colored shape. *From* www.abcam.com.

Methylation marks at H3K4 (lysine 4 of histone H3) in the promoter region of a gene seem to be necessary to initiate transcription whereas H3K36 methylation along the coding region is linked to efficient elongation and H3K79 methylation marks have a broad distribution across promoter and open-reading frame regions (Krivtsov and Armstrong, 2007). Hyperacetylation at histone H3 (H3K9ac, H3K14ac) and histone H4 has also been extensively associated with active regions of the genome, while methylation at H3K9, H3K27 and H4K20 participates in the establishment and maintenance of silent domains (Navarro *et al*., 2006) (**Figure 10**).

Adding to the complexity of epigenetic regulation, the methylation of lysine residues can occur at three states (mono-, di- or trimethylated) and the degree of methylation has also considerable influence on transcriptional activation or repression (Xu *et al*., 2008). For example, in *Saccahromyces cerevisiae,* di-methylated H3K4 is associated with both transcriptionally active and inactive genes, peaking toward the middle of the coding region whereas the tri-methylated H3K4 is present exclusively at transcriptionally active genes, in the 5’ and promoter regions, and monomethylation is most abundant at the 3’ end of genes (Santos-Rosa *et al*., 2002; Ruthenburg *et al*., 2007). Because the loss of trimethylation is a very slow process, histone methylation is considered as a much more stable mark in comparison to other histone modifications. It can indicate that the transcription of a given gene has occurred in the recent past but is not necessarily happening in the present time. It has been proposed that the pattern of histone H3 methylation can serve as a mark of transcriptional “short-term memory”, informing the cell of the transcription status of a given gene (Shilatifard, 2006).

Many enzymatic complexes are involved in writing this covalent language either by introducing (*writer*) or removing (*eraser*) a posttranslational modification. In this work, we will briefly introduce the different classes of writers and erasers, and more focus on enzymes writing the H3K4 methyl mark.

Histone acetylation

Acetylation of the ε-amino group of lysine residues neutralizes the positive charge of the histone tails and decreases their affinity for DNA which, in turn, makes the chromatin more accessible to transcriptional regulatory proteins. On the contrary, deacetylation of the histones leads to a more compacted chromatin, diminishing its accessibility for transcription factors (Santos-Rosa and Caldas, 2005). Histone acetylation is now recognized as a fundamental process that strongly affects the regulation of gene expression, but it appears also to be involved in other processes, such as DNA replication, recombination and repair, chromosome dynamics, cell proliferation, differentiation and apoptosis (Struhl, 1998, Roth *et al*., 2001).

❖ ***HATs (Histone Acetyltransferases)***

Given the link between histone acetylation and transcriptional activation, it is not surprising that many histone acetyltransferases (HATs) were first identified as transcriptional co-activators. These enzymes, capable of transferring acetyl groups from acetyl coenzyme A (acetyl-CoA) onto histone acceptors, all share a highly conserved acetyl-coA binding site. They are usually components of multisubunit complexes and can be organized into families, based on sequence similarity (Santos-Rosa and Caldas, 2005) (**Figure 11**):

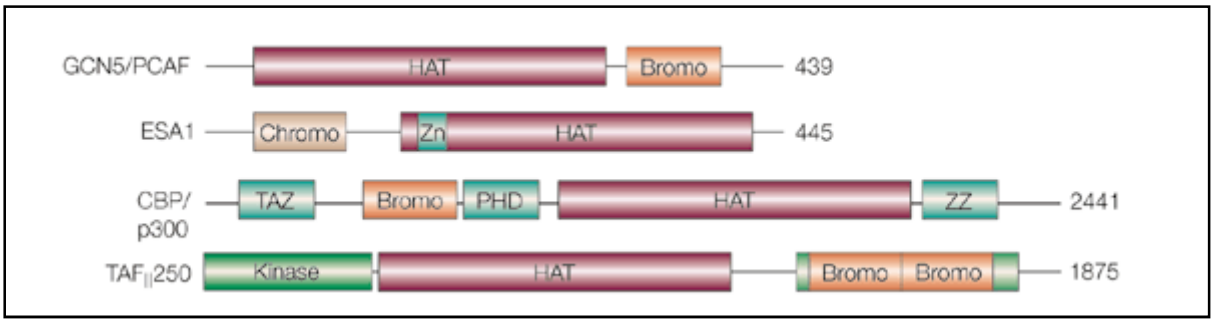


Figure 11: **Representative members of the different histone acetyltransferase (HAT) families: GCN5/PCAF (GNAT), yeast ESA1 (MYST), CBP/p300 (CBP/p300) and TAFII250**. **Bromo**: bromodomain; **chromo**: chromodomain; **HAT**: histone acetyltransferase domain; **Kinase**: kinase domain; **TAZ**, **PHD** and **ZZ**: conserved cysteine-histidine-rich regions; **Zn**: zinc-binding domain. From Marmorstein, *Nature Reviews Molecular Cell Biology* **2**; 422-432 (2001).

♦ ***The GNATs*** (Gcn5-related *N*-acetyltransferase) superfamily: members of this HAT group display several conserved acetylation-related structural motifs, including a HAT domain and a C-terminal bromodomain able to recognize and bind acetyl-lysine residues. They function as co-activators for a subset of transcriptional activators and include proteins linked to transcriptional initiation (Gcn5 and PCAF), elongation (Elp3) and telomeric silencing (Hat1) (Grant, 2001).

♦ ***The p300/CBP family***: these nuclear HATs contain a HAT domain, a bromodomain and three zinc fingers regions (cys, TAZ, and ZZ) that are believed to mediate protein-protein interactions. These proteins serve coactivator functions for transcription: they do not bind directly to DNA but are recruited to particular promoters through interaction with multiple DNA-bound transcription factors (Roth *et al*., 2001).

♦ ***The MYST family*** (named for its founding members: MOZ, Ybf2/Sas3, Sas2, and Tip60): beside the highly conserved MYST domain common to all MYST family members, many of them contain a cysteine-rich zinc-binding domain as well as N-terminal chromodomains. They are involved in a wide range of regulatory functions, including transcriptional activation, transcriptional silencing and cell cycle progression.

Other proteins such as basal transcription factors (TAFII250, ATF2) or nuclear hormone-receptor cofactors (SRC-1, ACTR) also harbor a histone acetyltransferase activity.

❖ ***HDACs (Histone Deacetylases)***

HDACs (histone deacetylases), the enzymes responsible for deacetylation of histones, do not bind directly to DNA, but are recruited to DNA by multiprotein chromatin complexes involved in transcriptional repression. In mammals, the HDACs have been ordered in three classes, based on homology to their yeast counterparts (Marks *et al*., 2003) (**Figure 12**):

♦ ***Class I (Rpd3-like)***: this class regroups HDACs 1, 2, 3 and 8. These proteins, expressed in a wide variety of tissues, present almost exclusively a nuclear localization and share homology in their catalytic site containing a zinc molecule.

♦ ***Class II (Hda1-like)***: HDACs 4, 5, 6, 7, 9 and 10 are components of this group. These proteins also present a zinc-active catalytic site, but a narrower tissue distribution than Class I HDACs and they shuttle between the nucleus and

the cytoplasm upon certain cellular signals. HDACs 6 and 10 are particular because they harbor two catalytic domains homologous with the class II catalytic site and HDAC 11 shares homology to both class I and II HDACs.

♦ ***Class III (Sir2-like)***: the conserved nicotinamide adenine dinucleotide-dependent Sir2 family of deacetylases differs from the two first classes of HDACs in their structure and their catalytic mechanism, which depends on the cofactor NAD^+. This class of proteins is also known to deacetylate non-histone proteins, such as alpha tubulin, p53,...

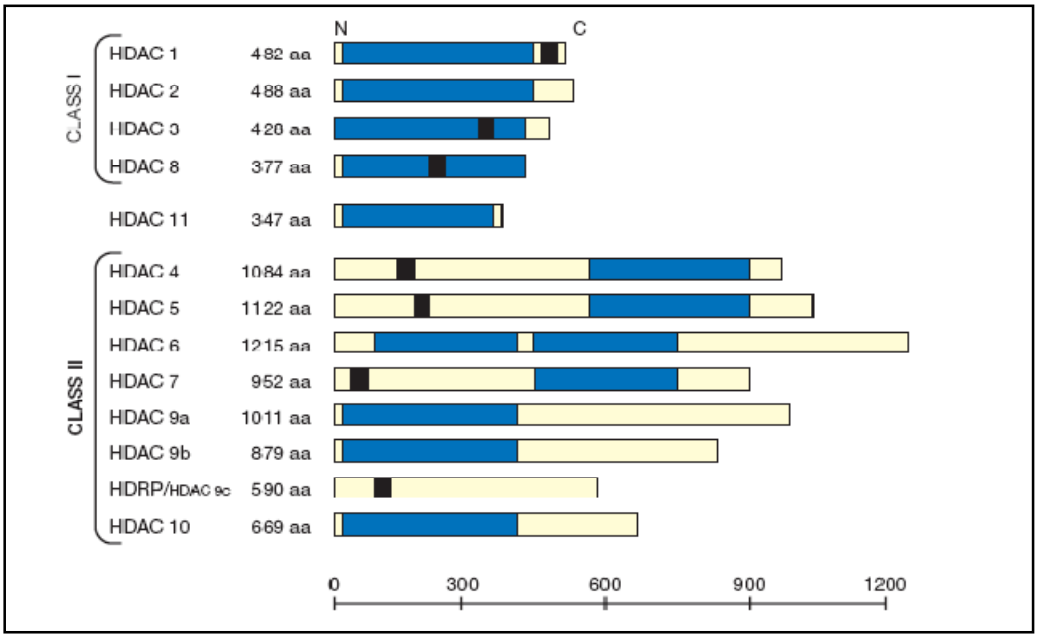


Figure 12: Schematic representation of the different histone deacetylases (HDACs). The catalytic domain is shown in blue. Black depicts a NLS. **N**: N-terminus; **C**: C-terminus. *From* de Ruijter *et al.*, Biochem. J. (2003) *370*, 737±749.

By many different mechanisms, HDACs may be involved in tumor development and they represent therefore an attractive target for anti-cancer therapy. Several molecules, such as Trichostatin A (TSA), phenylbutyrate or valproïc acid inhibit Class I and Class II HDACs by binding to their zinc-containing catalytic domain. HDACs inhibitors are usually well tolerated and have been shown to induce cancer cell growth arrest, differentiation and/or apoptosis *in vitro* and *in vivo*. Many drugs are now under clinical trials as monotherapy or in combination with other therapies in the treatment of cancer and recently (2006), the vorinostat (SAHA, suberoylanilide hydroxamic acid) has been approved by the U.S. FDA for the treatment of refractory cutaneous T-cell lymphoma (Zolinza®).

Histone methylation

The transfer of methyl groups from the S-adenosyl-L-methionine (SAM) to histone proteins occurs predominantly on specific lysine or arginine residues on histones H3 and H4. The ε-amino group of lysine residues can accept up to 3 methyl groups, resulting in mono-, di- or trimethylated forms, while arginine residues can only be mono- or dimethylated. Contrarily to histone acetylation, which is almost exclusively linked with transcriptional activation, histone methylation correlates with either transcriptional activation or repression, depending on the modified residue, the degree of methylation and the pattern of modifications on adjacent residues (Santos-Rosa and Caldas, 2005, Shilatifard, 2006).

❖ ***Histone Arginine MethylTransferases***

Protein arginine methyltransferases (PRMTs) catalyze the methylation of arginine residues of many cytosolic and nuclear proteins. Three distinct forms of methylation can occur on arginine residues on histone tails: monomethylation and symmetric or asymmetric dimethylation, depending on whether both guanido nitrogens are methylated or only one guanido nitrogen receives two methyl groups. The type I histone methyltransferases catalyze asymmetric dimethylation of arginine and include PRMT1, CARM1/PRMT4, PRMT2 and PRMT3, while the type II PRMTs catalyze the formation of symmetric dimethylarginine and include PRMT5/JBP1 and PRMT7. All these PRMTs share a conserved catalytic core, but present little similarity outside this core domain (Santos-Rosa and Caldas, 2005, Shilatifard, 2006).

❖ ***Histone Lysine Methyltransferases***

The enzymes catalyzing addition of methyl groups on the ε-nitrogen of histone lysine residues can be divided into two groups: the *SET domain-containing lysine histone methyltransferases*, which represent the majority of histone lysine methyltransferases and are involved in methylation of lysines 4, 9, 27 and 36 of histone H3 and lysine 20 of histone H4; and the *non-SET domain-containing lysine histone methyltransferases*, methylating lysine 79 of histone H3 (Shilatifard, 2006). The catalytic SET domain, which is a 130- to 140-amino acids motif found in a variety of chromatin-associated proteins, was first recognized as a conserved sequence in three *Drosophila melanogaster* proteins

where it takes its name from: Su(var)3-9, Enhancer of zeste [E(z)], and Trithorax (Dillon *et al*., 2005). These latter belong to two superfamilies, the trithorax (TRX) group and the polycomb group (PcG), also identified in vertebrates and known to activate or repress transcription, respectively. Equilibrium between their opposing roles is essential to maintain the expression state of homeotic (Hox) genes during development, hematopoiesis, X-chromosome inactivation and control of cell proliferation (Santos-Rosa and Caldas, 2005, Schuettengruber *et al*., 2007). Some TRXG and PcG components possess methyltransferase activities directed towards specific lysines, whereas other TRXG and PcG proteins interpret these histone marks (Schuettengruber *et al*., 2007).

- *H3K4*

In yeast *S. cerevisiae*, Set1 is the sole enzyme responsible for mono-, di- and trimethylation of histone H3K4 (Briggs *et al*., 2001; Roguev *et al*., 2001). Nevertheless, for its full H3K4 methyltransferase activity, this *Drosophila* Trithorax-related protein needs to associate to other polypeptides to form the multimeric complex COMPASS (standing for complex proteins associated with Set1). The recruitment of the COMPASS complex to chromatin by Paf1 (Polymerase II-associated factor), a complex associated with the elongating form of RNApol II, establishes a link between transcriptional elongation and histone methylation (Krogan *et al*., 2003).

In mammals, at least ten known or predicted H3K4 methyltransferases have been isolated. They catalyze mono-, di- or trimethylation, depending on the presence of specific residues in their catalytic domain and the proteins they associate with. Some of them contain SET domains related to the SET domain of yeast Set1 and *Drosophila* Trx and are referred as the MLL (Myeloid/Lymphoid or Mixed-Lineage Leukemia) or Set1-like family. This latter includes: Set1A (hSet1), Set1B, MLL1 (HRX, ALL1), MLL2 (TRX2), MLL3 and MLL4 (ALR). Other proteins, like ASH1, SET7/9 and SMYD3 display also a H3K4 methyltransferase activity, but their SET domains are not related to the one of Set1 and Trx (Ruthenburg *et al*., 2007) (**Figure 13**).

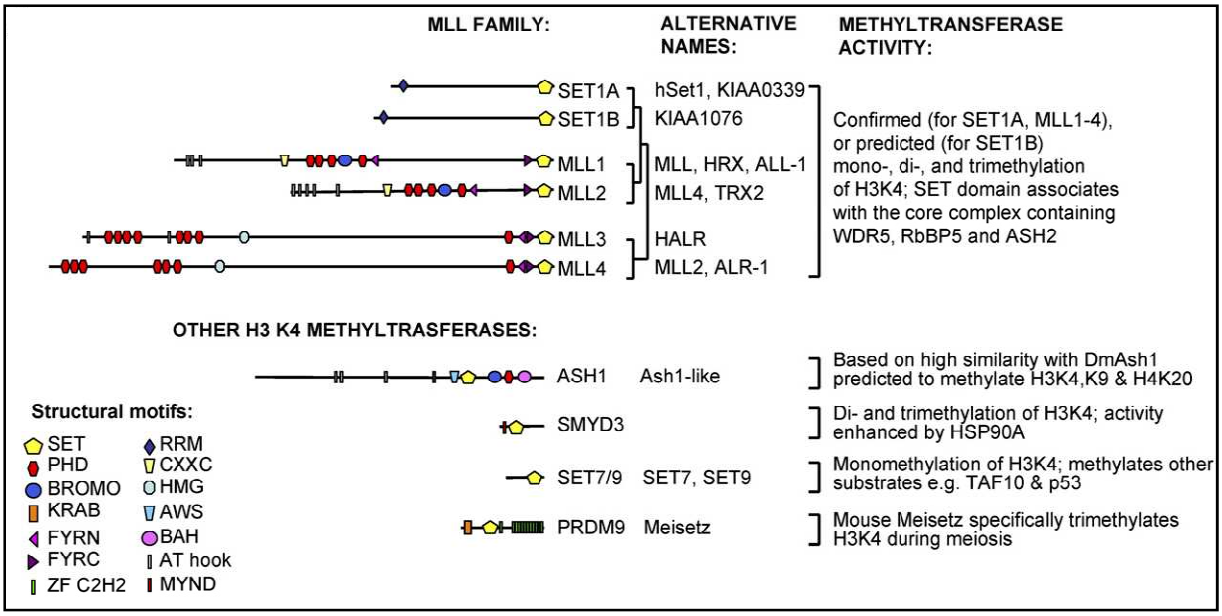


Figure 13: **Schematic representation of the domain architecture of known H3K4 histone methyltransferases (HMTs).** The H3K4 HMTs are represented by two groups: proteins containing SET domains related to the SET domain of yeast Set1 and Drosophila Trx (MLL family), or those unrelated yet able to methylate H3 at K4 (other H3K4 methyltransferases). *From* Ruthenburg *et al*, Molecular Cell *25*, January 12, 2007.

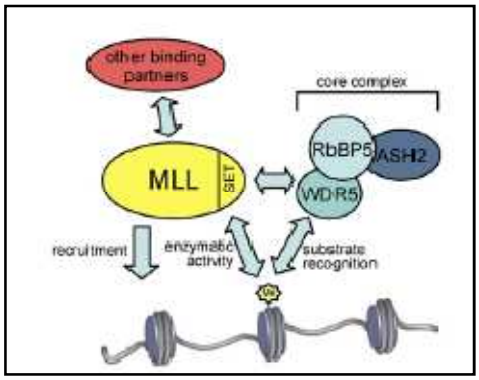


Figure 14: Schematic representation of the interactions between MLL complex components. The core complex containing RbBP5, WDR5 and ASH2 cooperates with the catalytic SET-domain of the MLL-family HMTs to methylate H3 at K4, whereas other regions of the MLL proteins are involved in association with other protein partners and in recruitment to target genes. *From* Ruthenburg *et al*, Molecular Cell *25*, January 12, 2007.

Like most histone-modifying complexes, the MLL-methyltransferases exist in multiprotein complexes. They all share three common subunits: WDR5, RbBP5 and ASH2 (ASH2L), which are homologs of yeast COMPASS/Set1 complex subunits Swd3, Swd1 and Bre2, respectively. These three proteins form a subcomplex that can exist independently of the catalytic subunit and can associate with the SET domain-containing COOH-terminal fragment of the different MLL-family members (Ruthenburg *et al*., 2007, Cho *et al*. 2007) (**Figure 14**). All these components are required for the H3K4 methyltransferase activity of the complex: WDR5 is thought to act as a presenter which recognizes the H3 tail with a preference for the dimethylated H3K4 form and presents it for

further methylation within the MLL complex, while ASH2L seems to be required for H3K4 trimethylation (Dou *et al*., 2006; Wysocka *et al*., 2005). Besides these three shared proteins and the SET-domain enzymatic subunit, MLL complexes contain distinct but overlapping subunits, including the cell proliferation regulator HCF-1, the tumor suppressor MENIN (Multiple Endocrine Neoplasia type I gene), hDPY3, ... The composition of the different MLL complexes is resumed in **Table 2**.

Subunit compositions of human Set1-like histone H3 K4 methyltransferase complexes

Yeast COMPASS/Set1 complex	Human Set1-like histone H3 K4 methyltransferase complexes			
	hSet1 complex	MLL complex	MLL2 complex	MLL3/MLL4 complex
Set1	hSet1	MLL	MLL2	MLL3/MLL4
Bre2	ASH2L	ASH2L	ASH2L	ASH2L
Swd1	RBBP5	RBBP5	RBBP5	RBBP5
Swd3	WDR5	WDR5	WDR5	WDR5
Sdc1	hDPY-30	hDPY-30	hDPY-30	hDPY-30
Swd2	hSwd2			
Spp1	CXXC1			
	HCF1	Menin	Menin	PTIP
		HCF1/HCF2		PA1
				NCOA6
				UTX

Table 2: Subunit compositions of human Set1-like histone H3K4 methyltransferase complexes. *From* Cho *et al*., The Journal of Biological Chemistry Vol. 282, N°28, pp. 20395–20406, July 13, 2007.

The greater number of H3K4 methyl writers in higher organisms compared to yeast, likely results from their functional specialization, in order to deal with the regulatory complexity of vertebrate development. Genetic disruption of the *MLL1* or *MLL2* genes in mice leads to embryonic lethality, revealing that each is important in early development, albeit in different ways (Glaser *et al*., 2006; Lubitz *et al*., 2007). The distinguishable phenotypes and/or the lethality observed when deletions or truncations of the different family members, strongly suggest that the multiple mammalian H3K4 methyltransferases provide non-redundant functions. Moreover, Milne shown in 2005 that loss of MLL1 has negligible effects on global H3K4 methylation but strongly influences the methylation status and expression of specific genes (Milne *et al*., 2005). All these examples support the notion of function specialization for each H3K4 methyltransferase, likely through differential expression patterns and/or target genes specificity. However, the nature of these

target genes and the specific mechanisms of recruitment remain elusive (Ruthenburg *et al*., 2007; Lee *et al*., 2007).

Based on the literature, different models have been proposed to explain the targeting of MLL-complexes to gene loci (reviewed in Ruthenburg *et al*., 2007). First, MLL-like proteins have been shown to associate with sequence-specific DNA binding factors, including estrogen receptor α (Mo *et al*., 2006), E2F6 (Dou *et al*., 2005),… (**Figure 15a**). This suggests a gene-specific recruitment by interaction with specific transcription factors. Other studies shown, like for the yeast COMPASS complex, an interaction between the MLL-complexes and the phosphorylated form of RNA polymerase II, as well as an overlapping between the H3K4me3 marks and the early elongating Pol II (Milne *et al*., 2005, Ng *et al*., 2003) (**Figure 15b**). The recognition of histone tail modifications by the MLL-complex itself (Wysocka *et al*., 2005) (**Figure 15c**) and some noncoding sequence-specific RNAs (Sanchez-Elsner *et al*., 2006) (**Figure 15d**) may also play a role in the recruitment of H3K4 methyltransferases to their target genes. It is likely that, *in vivo*, all these mechanisms act in combination to recruit and/or stabilize on chromatin the H3K4 methyltransferases.

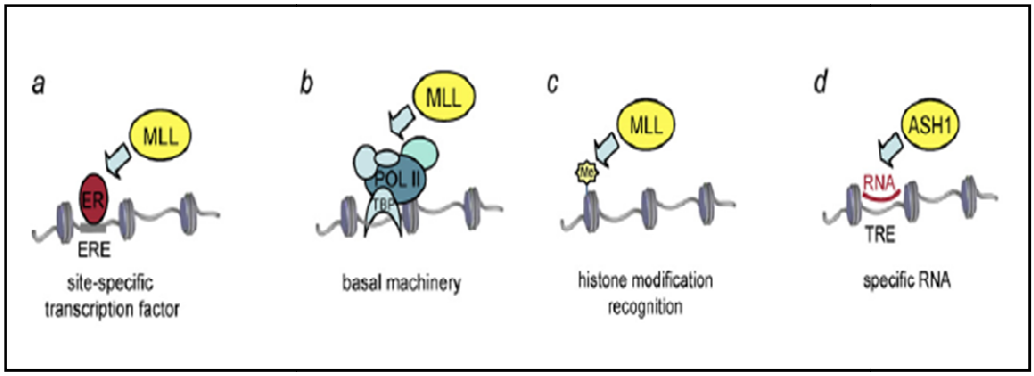


Figure 15: Mechanisms of H3K4 methyltransferase recruitment to the target genes: (a) association with site-specific transcription factors, **(b)** association with basal machinery, **(c)** histone modification recognition and **(d)** specific RNAs. *From* Ruthenburg *et al*, Molecular Cell *25*, January 12, 2007.

Deregulation of H3K4 methyltransferase activity might result in aberrant regulation of gene expression and cellular transformation. The human *MLL* gene was initially cloned on the basis of its transcription properties associated with the pathogenesis of several different forms of hematological malignancies, including acute myeloid leukemia (AML) (Ziemin-Van der poels *et al*., 1991). Reciprocal chromosomal rearrangements involving the *MLL* gene are indeed detected in approximately 80% of infants diagnosed with acute forms of both

myeloid and lymphocytic leukemias. They also occur in approximately 5% of adult patients with AML, and up to 10% with ALL (acute lymphoblastic leukemia). MLL is a 3968 amino acids protein which is specifically cleaved shortly after translation into two peptides that non covalently associate with each other. The amino-terminal portion contains an AT-hook DNA-binding domain, as well as a region that binds CpG-rich DNA and recruits transcriptional repressors (HDACs,…). The carboxy-terminal portion contains a transactivation domain and the catalytic SET domain (Xia *et al*., 2005). In most of the chromosomal translocations involving the *MLL* gene, the amino terminal part is fused to about 60 different partners, the most recurrent being AF4, AF9, ENL, AF10 and ELL. The gain of function of the resulting fusion protein could partly explain its role in leukemia and Wiederschain (Wiederschain *et al*., 2005) showed that inactivation of p53, a critical tumor suppressor proteins implicated in cell cycle regulation and pro-apoptotic responses, is a common characteristic of multiple MLL fusions. However, the precise mechanisms underlying the oncogenic function of MLL fusions are still poorly understood. Other H3K4 methyltransferases have also been linked to tumorigenesis: *MLL2* has been found to be amplified in epithelial cancers such as glioblastoma and pancreatic cancer (Huntsman *et al*., 1999), while SMYD3 is overexpressed in most colorectal and hepatocellular carcinomas (Hamamoto *et al*., 2004). Moreover, loss of interaction between MLL1 or MLL2 and the protein MENIN, observed in some tumor-derived cells, could abolish the tumor suppressor activity of this latter (Hughes *et al*., 2004; Milne *et al*., 2005; Yokoyama *et al*., 2004).

- *H3K9*

In contrast to H3K9 acetylation, H3K9 methylation is linked to gene silencing in many organisms, except in *S. cerevisiae* where it is not methylated. H3K9 methylation participates in the establishment and maintenance of heterochromatin in higher eukaryotic organisms through recruitment of the heterochromatin formation protein, HP1. This protein possesses a chromodomain which recognizes the H3K9 marks (mainly trimethylation) and it recruits DNA methyltransferases. Therefore, DNA methylation and H3K9 methylation appear to be tightly connected (He and Lehming, 2003).

The “SuV39 family” which belongs to the Polycomb Group, regroups the enzymes capable of methylating histone H3 on lysine 9, including the proteins Suv39h1, Suv39h2, G9a, ESET/SETDB1 and EuHMTaseI. These enzymes seem to display distinct patterns of localization on chromatin as well as different catalytic activity: Suv39h1/Suv39h2 mediates the H3K9 trimethylation in

pericentric heterochromatin, whereas, G9a is the major H3K9 dimethylase in euchromatin (Shilatifard, 2006; Santos-Rosa and Caldas, 2005).

- *H3K27*

Methylation of K27 on histone H3 is a signal for transcription repression and maintenance of stable epigenetic silencing via recruitment of the Polycomb Repressive Complex (PRC1) (Cao *et al*., 2002; Santos-Rosa and Caldas, 2005). It also seems to play a role in the initial stage of chromosome X inactivation (Plath *et al*., 2003; Silva *et al*., 2003).

Methylation of K27 of histone H3 is catalyzed by EZH2, a component of the Polycomb Repressive Complex 2 (PRC2) or EED-EZH2 complex, which contains a SET domain (Muller *et al*., 2002; Cao *et al*. 2002). Moreover, G9a, the H3K9 methyltransferase is also capable of methylating K27 on histone H3 (Tachibana, 2001).

- *H3K36*

H3K36 methylation marks within the coding region of transcriptionally active genes are necessary for efficient elongation. In yeast, Set2, a SET-domain containing protein, is the sole enzyme identified to display an HMTase activity specific for lysine 36 of histone H3 (Strahl *et al*., 2002). This latter requires the association of Set2 with the elongating form of the RNApol II (Hampsey *et al*., 2003). In human, several Set2 homologues have been identified (NSD1, NSD2/MMSET, WHSC1, NSD3, ASH1, HIF1), but their HMTase activity still needs to be defined.

- *H3K79*

H3K79 methylation marks also correlate with transcriptional activation and present broad distribution across promoter and open-reading frame regions of actively transcribed genes (Krivstov and Armstrong, 2007). Contrarily to other lysine residues, lysine K79 is not located on the N-terminal tail, but lies in the globular domain of histone H3. Dot1 (standing for disruptor of telomeric silencing 1), the enzyme responsible for the methylation of H3K79 is also different from the other HMTases in the way that it lacks the characteristic catalytic SET domain (Feng *et al*., 2002; van Leeuwen , 2002).

- *H4K20*

In human, methylation of lysine 20 on histone H4 is involved in chromatin condensation and is catalyzed by SET8 (Pr-Set7) (Rice *et al*., 2002; Santos-Rosa and Caldas, 2005).

❖ ***Histone Demethylases***

Loss of histone methylation is a very slow process: some observations indicate that the half-life of histones and methyl-residues within them are the same (Byvoet *et al*., 1972; Thomas *et al*., 1972). Because of this high thermodynamic stability of methyl marks and the lack of identification of histone demethylases, histone methylation, in particular trimethylation, has been considered for a long time as an irreversible mechanism. However, the recent discovery of histone demethylases has challenged this dogma.

Recently, a human arginine deiminase (PAD4/PADI4) was shown to convert either non- or monomethylated arginine to citrulline, therefore effectively erasing arginine from histone tails and preventing further methylation of this residue (Cuthbert *et al*., 2004; Wang *et al*., 2004b). However, this deimination reaction does not affect the dimethylated arginines and it is thus not considered as a proper "arginine demethylation event".

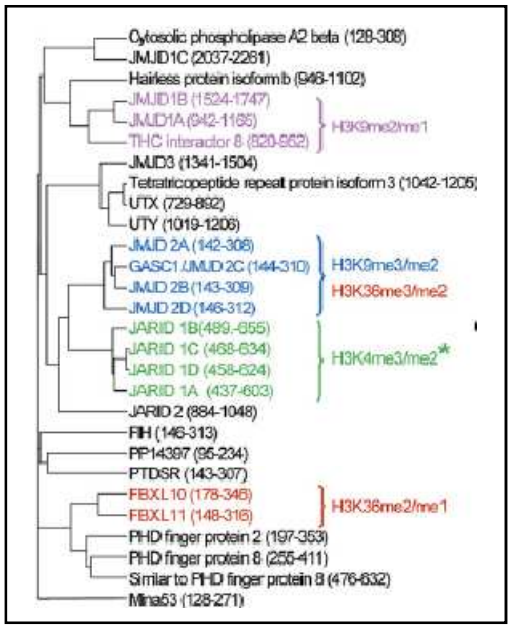
The first histone demethylase identified was the FAD-dependent amine oxidase LSD1 (lysine-specific demethylase, KIAA0601). LSD1 can specifically remove mono- and dimethylation from lysine 4 of histone H4. However, it is unable to demethylate trimethylated H3K4, most likely due to the absence of a protonated nitrogen required for oxidation (Shi *et al*., 2004). Therefore, it is likely that other oxidative mechanisms exist to mediate demethylation of trimethylated lysines. Different dioxygenases were tested for a histone demethylase activity and proteins harboring a Jumonji C domain (JmjC) were identified as enzymes that catalyze demethylation of methylated lysine residues by using a hydroxylation reaction which requires iron and α-ketoglutarate as cofactors (Trewick *et al*., 2005; Cloos *et al*., 2006; Tsukada *et al*., 2006; Whetstine, 2006). The JmjC domain has been shown to be the catalytic domain for histone demethylation (Tsukada *et al*., 2006) and based on this demethylase signature motif, several proteins possessing this domain, were identified as histone demethylases. So far, five families of JmjC histone demethylases have been described, all submembers within each family sharing the same enzymatic specificity for methylated histone lysine residues (**Figure 16**):

♦ ***The FBXL10/11 family***: its members, JHDM1a (JmjC domain histone demethylase 1, FBXL-11) and JHDM1b (FBXL 10), mediate demethylation of mono- and dimethylated H3K36 (Tsukada *et al*., 2006).

♦ ***The JMJD1 family***: JMJD1A/JHDM2a has been shown to demethylate dimethylated H3K9 (Yamane *et al*., 2006).

♦ ***The JMJD2 family***: these enzymes were the first enzymes shown to reverse lysine trimethylation. This family regroups four members, JMJD2 A-D, which exhibits dual-substrate specificity for di- and trimethylated H3K9 and H3K36 (Cloos *et al*., 2006, Fodor *et al*., 2006, Klose *et al*., 2006b; Whetstine *et al*., 2006).

♦ ***The JARID1 family***: this family, including RBP2 (JARID1 A), PLU1 (JARID1 B) and SMCX (JARID1 C), was identified as a family of histone demethylases targeting di- and trimethylated H3K4 (Christensen *et al*., 2007).



♦ ***The UTX-UTY family***: the member Jmjd3 was shown to be an efficient H3K27me3 demethylase (De Santa *et al*., 2007).

Figure 16: Phylogenetic analysis of the human JmjC family. The reported specificities of the different JmjC families are indicated. From Christensen *et al*., Cell 128, 1063–1076, March 23, 2007.

The field of demethylases is moving forward at a very fast pace and new histone demethylases will probably be discovered soon. About 30 proteins in mouse and human have been shown to possess a JmjC domain, although some of them are unlikely to be involved in histone demethylation due either to their cytoplasmic localization or to aminoacidic substitutions in the iron- and/or cofactor-binding site (De Santa *et al*., 2007). However, new demethylases may also use alternative enzymatic domains and chemical reaction mechanisms as opposed to the LSD1 and JmjC proteins (Shi and Whetstine, 2007).

Histone phosphorylation

Phosphorylation of histones occurs mainly on serines 10 and 28 and on threonine 11 of histone H3. During cell division, this modification is involved in chromosome condensation and is dynamically regulated by the opposing effects of the members of the aurora AIR2-Ipl1 kinase family, and the type 1 phosphatases (PP1). Moreover, phosphorylation of H3S10 has an important role in the transcriptional activation of eukaryotic genes by promoting acetylation of H3K14.

Histone ubiquitination

In yeast, monoubiquitination of histone H2B on lysine 123 (corresponding to H2B K120 in human) is required for methylation of histone H3 at K4 and K79, by COMPASS and the non SET domain enzyme Dot1, respectively (Dover *et al*., 2002; Sun and Allis, 2002). Ubiquitination is a reversible, dynamic process which is catalyzed by the ubiquitin ligase complex BRE1/Rad6 (Wood *et al*., 2003) and by the deubiquitinating enzymes Ubp8 and Ubp10 (Henry *et al*., 2003; Emre *et al*., 2005). The Paf1 complex, which is involved in the recruitment of COMPASS to the early elongating RNApol II, has been shown to be also required for the functional activation of Rad6/Bre1 in histone monoubiquitination (Wood *et al*., 2003).

Ubiquitination of histone H2A on lysine 119 by PcG proteins belonging to PRC1 (PRC1-L, PRC1-like) has been linked to polycomb silencing and chromosome X inactivation (Wang *et al*., 2004a; de Napoles, 2004).

Cross-talk

The large number of possible combinations of histone modifications constitutes the “histone code”. This latter can be read by proteins containing bromo- or chromodomains, which have been shown to recognize and bind acetyl- and methyl-lysine residues respectively. The wide distribution of such domains among enzymes that methylate, acetylate or remodel chromatin suggests an important crosstalk between all these modifications. Moreover, the code writers are often components of multiprotein complexes that contain or recruit other writers with which they can work synergistically. Histone modifications are thus interconnected as one mark can either promote or prevent another one. Some histone modifications, such as H3K4 methylation or H3K14 acetylation, have been linked to transcriptional activation and are generally incompatible with repressive marks like H3K9 methylation. The crosstalk between the different histone modifications determines also a certain chronology in the establishment of a specific pattern. For example, phosphorylation of H3 S10 facilitates acetylation of K14 and methylation of K4, which in turn, impairs the ability of Su(var)3-9 to methylate K9 (Wang *et al*., 2001). Moreover, this crosstalk may also occur between modifications on different histones. This is the case for the ubiquitination of H2B K120, which is required for histone H3K4 methylation and subsequent activation of transcription.

Crosstalk between histone modifications and DNA methylation is also required for proper epigenetic control of gene expression. For example, in fungi, H3K9 methylation is found to be upstream DNA-methylation, likely marking regions for cytosine methylation (Tamaru *et al*., 2001; Tamaru *et al*., 2003). However, the temporal order of events is not yet clearly defined and may differ from organism to organism and from gene to gene (Santos-Rosa and Caldas, 2005).

New nomenclature for chromatin-modifying enzymes

Recently, a new nomenclature was proposed in order to unify and rationalize the field of chromatin-modifying enzymes (Cell *131*, November 16, 2007). The enzymes have been given a more generic name that reflects the type of enzymatic activity they perform and the type of residue they modify. The lysine demethylases will therefore become KDMs (**Table 3**), acetyltranferases will become KATs (**Table 4**) and lysine methyltransferases will become KMTs (**Table 5**).

New Name	Human	*D. melanogaster*	*S. cerevisiae*	*S. pombe*	Substrate Specificity	Function
KDM1	LSD1/BHC110	Su(var)3-3		SpLsd1/Swm1/ Saf110	H3K4me1/2, H3K9me1/2	Transcription activation and repression, hetero-chromatin formation
KDM2			Jhd1		H3K36me1/2	Transcription elongation
KDM2A	JHDM1a/FBXL11				H3K36me1/2	
KDM2B	JHDM1b/FBXL10				H3K36me1/2	
KDM3A	JHDM2a				H3K9me1/2	Androgen receptor gene activation, spermatogenesis
KDM3B	JHDM2b				H3K9me	
KDM4			Rph1		H3K9/ K36me2/3	Transcription elongation
KDM4A	JMJD2A/JHDM3A				H3K9/ K36me2/3	Transcription repression, genome integrity
KDM4B	JMJD2B				H3K9/ H3K36me2/3	Heterochromatin formation
KDM4C	JMJD2C/GASC1				H3K9/ K36me2/3	Putative oncogene
KDM4D	JMJD2D				H3K9me2/3	
KDM5		Lid	Jhd2	Jmj2	H3K4me2/3	
KDM5A	JARID1A/RBP2				H3K4me2/3	Retinoblastoma-interacting protein
KDM5B	JARID1B/PLU-1				H3K4me1/2/3	Transcription repression
KDM5C	JARID1C/SMCX				H3K4me2/3	X-linked mental retardation
KDM5D	JARID1D/SMCY				H3K4me2/3	Male-specific antigen
KDM6A	UTX				H3K27me2/3	Transcription activation
KDM6B	JMJD3				H3K27me2/3	Transcription activation

Table 3: Lysine (K) demethylases: KDMs.

New Name	Human	*D. melanogaster*	*S. cerevisiae*	*S. pombe*	Substrate Specificity	Function
KAT1	HAT1	CG2051	Hat1	Hat1/ Hag603	H4 (5, 12)	Histone deposition, DNA repair
KAT2		dGCN5/PCAF	Gcn5	Gcn5	H3 (9, 14, 18, 23, 36)/ H2B; yHtzl (14)	Transcription activation, DNA repair
KAT2A	hGCN5				H3 (9, 14, 18)/H2B	Transcription activation
KAT2B	PCAF				H3 (9, 14, 18)/H2B	Transcription activation
KAT3		dCBP/NEJ			H4 (5, 8); H3 (14, 18)	Transcription activation, DNA repair
KAT3A	CBP				H2A (5); H2B (12, 15)	Transcription activation
KAT3B	P300				H2A (5); H2B (12, 15)	Transcription activation
KAT4	TAF1	dTAF1	Taf1	Taf1	H3 > H4	Transcription activation
KAT5	TIP60/PLIP	dTIP60	Esa1	Mst1	H4 (5, 8, 12, 16); H2A (yeast 4, 7; chicken 5, 9, 13, 15); dH2Av/yHtzl (14)	Transcription activation, DNA repair
KAT6		(CG1894)	Sas3	(Mst2)	H3 (14, 23)	Transcription activation and elongation, DNA replication
KAT6A	MOZ/MYST3	ENOK			H3 (14)	Transcription activation
KAT6B	MORF/MYST4				H3 (14)	Transcription activation
KAT7	HBO1/MYST2	CHM		(Mst2)	H4 (5, 8, 12) > H3	Transcription, DNA replication
KAT8	HMOF/MYST1	dMOF (CG1894)	Sas2	(Mst2)	H4 (16)	Chromatin boundaries, dosage compensation, DNA repair
KAT9	ELP3	dELP3/ CG15433	Elp3	Elp3	H3	
KAT10			Hap2		H3 (14); H4	
KAT11			Rtt109		H3 (56)	Genome stability, transcription elongation
KAT12	TFIIIC90				H3 (9, 14, 18)	Pol III transcription
KAT13A	SRC1				H3/H4	Transcription activation
KAT13B	ACTR				H3/H4	Transcription activation
KAT13C	P160				H3/H4	Transcription activation
KAT13D	CLOCK				H3/H4	Transcription activation

Table 4: Lysine (K) acetyltransferases: KATs.

New Name	Human	*D. melanogaster*	*S. cerevisiae*	*S. pombe*	Substrate Specificity	Function
KMT1		Su(Var)3-9		Clr4	H3K9	Heterochromatin formation/silencing
KMT1A	SUV39H1				H3K9	Heterochromatin formation/silencing
KMT1B	SUV39H2				H3K9	Heterochromatin formation/silencing
KMT1C	G9a				H3K9	Heterochromatin formation/silencing
KMT1D	EuHMTase/GLP				H3K9	Heterochromatin formation/silencing
KMT1E	ESET/SETDB1				H3K9	Transcription repression
KMT1F	CLL8					
KMT2			Set1	Set1	H3K4	Transcription activation
KMT2A	MLL1	Trx			H3K4	Transcription activation
KMT2B	MLL2	Trx			H3K4	Transcription activation
KMT2C	MLL3	Trr			H3K4	Transcription activation
KMT2D	MLL4	Trr			H3K4	Transcription activation
KMT2E	MLL5				H3K4	Transcription activation
KMT2F	hSET1A				H3K4	Transcription activation
KMT2G	hSET1B				H3K4	Transcription activation
KMT2H	ASH1	Ash1			H3K4	Transcription activation
KMT3			Set2	Set2	H3K36	Transcription activation
KMT3A	SET2				H3K36	Transcription activation
KMT3B	NSD1				H3K36	
KMT3C	SYMD2				H3K36 (p53)	Transcription activation
KMT4	DOT1L		Dot1		H3K79	Transcription activation
KMT5				Set9	H4K20	DNA-damage response
KMT5A	Pr-SET7/8	PR-set7			H4K20	Transcription repression
KMT5B	SUV4-20H1	Suv4-20			H4K20	DNA damage response
KMT5C	SUV4-20H2					
KMT6	EZH2	E(Z)			H3K27	Polycomb silencing
KMT7	SET7/9				H3K4 (p53 and TAF10)	
KMT8	RIZ1				H3K9	Transcription repression

Table 5: Lysine (K) methyltransferases: KMTs.