Peer reviewed article

Transcription regulation and human diseases

Jean Villard

Immunology and Transplant Unit, Division of Immunology and Allergology, Geneva University Hospital, Geneva, Switzerland

Summary

Regulation of gene expression is a crucial step in the maintenance of cellular homeostasis. The control of gene expression can occur in multiple steps. The overwhelming majority of regulatory events occur at the level of transcription. To initiate transcription eukaryotic polymerase RNA II requires the close collaboration of a battery of proteins collectively termed transcription factors. Transcription factors are generally divided into two groups: (1) the basal transcription factors which are ubiquitous and recruit the RNA polymerase II multi-protein complex to the minimal promoter; (2) gene-specific transcription factors that activate or repress basal transcription. These proteins bind to regulatory sequences organized in a series of regulatory modules along the DNA. Thus, the molecular basis for transcriptional regulation of gene expression is the binding of *trans*-acting proteins (transcription factors) to *cis*-acting sequences (binding sites). A growing list of human diseases are due to genetic defects in transcription factors. In most cases, mutations in transcription factors lead to pleiotropic effects. Clinical observations can be explained at the molecular level by the fact that these *trans*-acting factors control the expression of many genes, usually in combination with one or more further activators. In addition, many events that lead to the process of tumourigenesis in leukaemias and in solid tumours implicate overexpression or mutations of transcription factors. This review describes human diseases attributable to mutations in the genes coding for transcription factors or mutations in their cognate binding sites.

Key words: transcription factors; genetic diseases; promoter; chromatin structure; cancer; MHC class II deficiency

Introduction

The phenotypic differences that distinguish the various kinds of cells in higher eukaryotes are largely due to differences in the expression of genes that code for proteins. Cellular proliferation and differentiation processes underlying human development are controlled by programs of regulated gene expression within the embryo. As is the case in development after birth, humans must continue to respond to a variety of endogenous and exogenous stimuli to maintain the correct functioning of the adult organism. These stimuli induce specific adaptive responses that involve quantitative or qualitative changes in gene expression. Regulation of gene expression is the crucial step that allows the maintenance of cellular and organism homeostasis. Gene expression can be regulated at any of at least seven potential control steps [1, 2].

(1) chromatin structure
(2) initiation of transcription
(3) processing of the transcript
(4) transport to the cytoplasm
(5) translation of mRNA
(6) mRNA stability
(7) protein activity stability.

No financial support declared.

The first step is implied by the fact that the genes may exist in either of two structural conditions. The genomes of eukaryotic organisms are packaged into chromatin. Some forms of chromatin are so highly compacted that the packaged genes are transcriptionally silent. In some cases, this silencing can be reversed and the packaged genes activated. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed. Histone acetylation and the state of gene methylation are two mechanisms that are mainly implicated in the change of structure which can render the gene "transcribable" [3, 4]. Although the control of gene expression can occur at multiple stages, it is well demonstrated that the overwhelming majority of regulatory events occur at the initiation of gene transcription [5].

Although prokaryotic RNA polymerase can initiate transcription on its own, eukaryotic polymerases require the prior assembly of general transcription factors at the promoter. This assembly provides several steps at which the initiation of transcription can be regulated, and many eukary-

otic regulatory proteins are thought to work by facilitating or hindering this process. Some of these proteins are transcriptional activators, while others are transcriptional repressors. These proteins bind to regulatory sequences organized into a series of regulatory modules along the DNA. Thus, the molecular basis for transcriptional regulation of gene expression is the binding of *trans*-acting proteins (transcription factors) to *cis*-acting sequences (binding sites) [5]. This review focuses on these two elements in relation to human diseases attributable to mutations in the genes coding for transcription factors or mutations in their cognate binding sites.

Cis-acting DNA elements

Transcription is initiated by assembly of a transcription initiation complex composed of RNA polymerase II and other associated basal factors (general transcription factors). This multi-protein complex binds to a short DNA sequence called the minimal promoter, which often contains a conserved DNA sequence motif called the TATA box situated 20–30 nucleotides upstream of the transcription initiation site. The general transcription factors are characterized by their ability to control the activity of RNA polymerase II on the minimal promoter (Fig. 1). This step is in principle sufficient to initiate transcription but its regulation is mediated by others factors that bind the different regulatory elements briefly described below [6, 7].

Upstream of the minimal promoter *cis*-acting sequences, in an orientation-dependent fashion, called proximal promoter elements are found; these sequences are bound by specific transcription factors whose presence increases or decreases transcriptional activity of the genes (Fig. 1). In addition to promoter regions, other *cis*-acting elements can be arranged within several hundred or thousand base pairs of DNA situated 5' or 3' of the initiation site (in an intron for example). These elements are also binding sites for sequence-specific transcription factors. In contrast to promoters, the position and orientation of these elements are variable with regard to the genes (Fig. 1). These elements are called enhancers if the specific factors that bind to the elements activate transcription, and silencers if they repress transcription [8]. The importance of a particular *cis*-acting element can vary greatly in different cell types and in response to different physiological stimuli because the transcription factors that bind to these elements may vary in abundance or in ability to function in different tissues and under different circumstances [5, 9].

Overlapping or superimposed binding sites for multiple factors can result in different positive and negative factors competing for sites. In some cases, synergistic effects that are dependent on strict spacing between adjacent *cis*-acting elements have been observed. Various types of silencer elements can block activity of *cis*-linked enhancers. Although the interplay between the various factors that may bind to these regions is not always reported, it is the combination of *cis*-acting elements arranged in a unique configuration that seems to confer an individualized spatial and temporal transcription program on each gene [5, 9].

Some genes respond to additional regulatory elements as well as to promoters and enhancers. These different *cis*-acting elements exert their effects over very large distances that may encompass entire clusters of genes. Such elements are called locus control regions (LCR). They appear to establish chromatin domains in which transcription factors are able to gain access to their binding sites and activate transcription [10, 11].

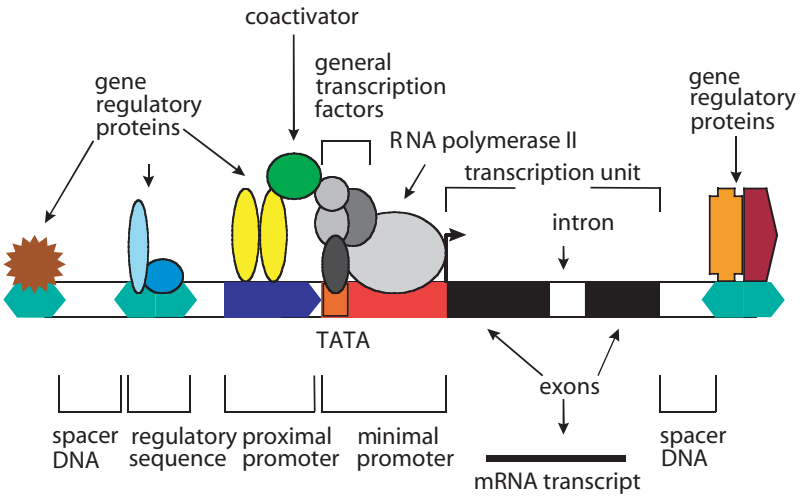


Figure 1

Prototypical promoter with the transcription factors. The minimal promoter is the DNA sequence at which the general transcription factors and RNA polymerase II assemble. The regulatory DNA sequence motifs serve as binding sites for regulatory proteins. These sequences can be located adjacent to the minimal promoter (proximal promoter, unidirectional), far upstream of it (bi-directional, up to several kilobases), or even downstream of the gene. Enhancers are cis-acting sequences that serve as specific binding site for transcription factors to activate transcription. Silencers are cis-acting sequences that serve as specific binding site for transcription factors to inhibit transcription. DNA looping is thought to allow gene regulatory proteins to bind at any of these positions to interact with the proteins that assemble at the promoter. The combination of regulatory proteins and their binding sites relative to the promoter are different for each gene.

Trans-acting factors

General transcription factors

RNA polymerases (I, II and III) cannot initiate transcription in eukaryotic genes on their own. They require additional proteins, the general transcription factors (Fig. 1). These proteins have to assemble with RNA polymerase into a complex on the promoter DNA. To date, at least 40 proteins grouped into subcomplexes have been identified. Careful order of addition experiments with purified components of the general transcription machinery have suggested a stepwise assembly of initiation complexes in vitro [9, 12]. The exact role of all these proteins is not clear. Some are implicated in modifications that occur before the beginning of transcription. These steps include conformational changes, stimulation of enzymatic activities, for example, via phosphorylation, or chromatin remodelling activity. In the context of this review, one of these general transcription factor multiprotein complexes is of special interest, TFIIH. TFIIH consists of 9 subunits with helicase, ATPase and kinase activities. Some of these subunits are also essential for the mechanisms of DNA repair coupled with transcription [7, 12, 13]. Genetic defects in two TFIIH subunits are implicated in a rare human disease which is discussed below.

Transcription activators

Regulatory transcription factors are proteins that positively or negatively affect the rate of transcription (activators and repressors, respectively) by specific interaction with regulatory DNA sequences and by interactions with other proteins (Fig. 1). Most of these transcription factors are activators and have several modules in common.

(1) A DNA binding domain that positions the protein on specific sequences. When the DNA binding domains are compared, many transcription factors fall into groups defined by related motifs that define a protein structure capable of binding DNA. The four well-described motifs are the basic helix-loop-helix (HLH) motif, the basic leucine zipper (LZ) motif, the zinc finger (ZF) motif, and helix-turn-helix (HTH) domains.

(2) Transcription factors generally contain activation domains which are often characterized by motifs rich in glutamine, rich in proline or rich in acidic amino acids. Activation domains could work by recruiting or accelerating the assembly of the general transcription factors on the promoter, but their mode of action remains unclear.

(3) The majority of transcription factors bind DNA as homo- or heterodimers, and they thus have a dimerization domain.

(4) Some transcription factors also have ligand binding domains, such as hormone binding domains, which are essential for controlling this activity.

The exact way regulatory transcription factors affect the level of transcription has not been clearly demonstrated but they probably work by the recruitment and/or stabilization of general transcription factors, the induction of a modification, the induction of a conformational change or the stimulation of an enzymatic activity of the basal transcription machinery [14, 15]. Moreover, some trans-acting factors may be implicated in chromatin remodelling to permit enhanced accessibility to general transcription factors or specific activators [4]. These different roles can be promoted directly via protein–protein interactions with the basal transcription machinery, or via interactions with other trans-acting factors. In addition, some regulatory transcriptions factors do not contact the basal machinery directly, but instead bind co-activators that in turn contact the basal apparatus. Co-activators may be regarded as transcription factors whose specificity is conferred by the ability to bind to DNA-binding transcription factors instead of directly to DNA (Fig. 1) [16, 17]. The formation and dissociation of trans-acting factor complexes are an integral part of the regulation of many cellular processes. The combinatorial action of trans-activators confers specificity and synergistic activation of gene expression.

Cis-acting DNA elements and disease

Haemophilia B Leyden

Mutations in the X-linked factor IX gene result in haemophilia B. The majority of cases are due to mutations in the coding protein sequence. However, in a small number of cases the disease is due to mutations in the regulatory region of the factor IX gene. Patients with haemophilia B Leyden present with severe bleeding symptoms and <1% of the normal amounts of plasma factor IX in childhood. After puberty, the clinical symptoms improve gradually and plasma factor IX concentrations rise to 60% of normal. All patients studied have mutations within 20 bp of the factor IX gene transcription initiation site. These mutations disrupt the binding of transcription factors that is essential for factor IX gene expression. For example, mutation at –20 disrupts binding of hepatocyte nuclear factor 4 (HFN4). In addition, the promoter region of factor IX contains a binding site for the androgen receptor at –22 to –38 which overlaps the HFN4 binding site. At puberty, binding of the androgen receptor to this site can com-

pensate for the lack of HFN4 or other transcription factors to activate the factor IX gene. Certain mutations of the –22 to –38 region, named the Brandbourg variant, prevent this compensation, so that there is no improvement at puberty [18, 19].

Haemoglobinopathies

Hereditary persistence of foetal haemoglobin (HPFH), in which γ-globin expression persists into adult life, is another example of human disease resulting from mutations in *cis*-acting elements. Point mutations have been identified in the promoter region of the ^{A}γ-globin gene (nt –175 of the initiation site) to which the GATA-1 transcription factor binds. The ^{A}γ-globin gene cannot be repressed [20]. GATA-1 binding sites are also present in the LCR. This could partly explain what occurs in Hispanic (γδβ) thalassaemia in which a deletion of most of the LCR leaves the gene cluster in a chromatin conformation that is inaccessible to DNase I and leads to the absence of globin gene expression [21]. Mutations in the TATA box (–28 to –31) and in the CACC box (–92 to –105) have also been described in β-thalassaemias, characterized by a decrease in β-globin gene expression [22].

Progressive myoclonus epilepsy

Progressive myoclonus epilepsy of the Unverricht-Lundborg type (EPM1) is a rare autosomal recessive disorder with onset between 6 and 13 years followed by variable progression to mental deterioration and cerebellar ataxia. This disease is characterized in the majority of the cases by alleles containing expansions of a dodecamer (12-mer) sequence located about –70 from the transcription initiation site (more than 60 copies instead of the 2 or 3 in the normal alleles) of the cysteine proteinase inhibitor gene cystatin B (CSTB). Reduction of CSTB gene expression is thought to be due to increasing distance between *cis*-acting binding sites for cell specific transcription factors and the transcription initiation site. Another possibility is that this dodecamer sequence is being recognized by a repressor and the presence of many copies of this sequence results in a multiplicity of negative regulatory effects [23].

Promoter polymorphisms and human diseases

Promoter polymorphisms affecting the expression levels of certain genes may be implicated in a variety of pathologies. Differential expression of different MHC class II alleles may be associated with promoter polymorphisms. For example, the finding that some alleles of the HLA-DQ gene differ in promoter strength in response to cytokines such as TNFα suggests a mechanism for the association of certain alleles of this gene with susceptibility to auto-immune diseases [24].

A rare allele of the human TNFα promoter called TNF2 which is located close to a well-defined HLA-A1 haplotype is associated with autoimmunity and high TNFα production. A high plasma level of TNFα is correlated with severity in malaria and leishmania infections. In addition, patients homozygous for the TNF2 allele exhibit a significantly higher incidence of death from cerebral malaria [25].

Reports have correlated progression of AIDS with a promoter variant of the CCR5 gene. The CCR5 gene encodes a cell surface chemokine receptor molecule that serves as the principal coreceptor with CD4 for macrophage-tropic (R5) strains of human immunodeficiency virus-type 1 (HIV-1). Genetic association analysis of five cohorts of people with AIDS revealed that infected individuals homozygous for the promoter allele called CCR5P1, progress to AIDS more rapidly than those with other CCR5 promoter genotypes, particularly in the early years after the infection [26].

Trans-acting factors and disease

The vital role of transcription factors together with the fact that a single factor can affect the expression of many genes, suggest that the inactivation of transcription factors as a result of an inherited mutation is incompatible with survival. This is probably true in many cases, but mutations of some transcription factors are compatible with survival and result in specific diseases. In the past few years a growing number of mutations in genes encoding transcription regulators have been shown to be associated with a number of congenital syndromes [27]. The consequences are malformations, disruption of physiological pathways or tumourigenesis. The abnormalities observed are often confined to a subset of tissues that express the affected gene. In most cases, the phenotypes are pleiotropic, reflecting the fact that transcription factors control the expression of many genes. Mutation analysis, comparison with homologous mouse models, and the uncovering of the molecular mechanisms by which these proteins act together, give unparalleled insight into the major physiological processes controlled by these genes.

In addition to congenital syndromes, a number of somatic mutations of certain transcription factors contribute to the multi-step processes of tumourigenesis and lead to the growing list of cancers, in which these processes may play a role. Chromosomal translocations observed in certain tumours such as human leukaemias, result in a rearrangement of the regulatory and coding regions of a variety of transcription factor genes [28] (Fig. 2).

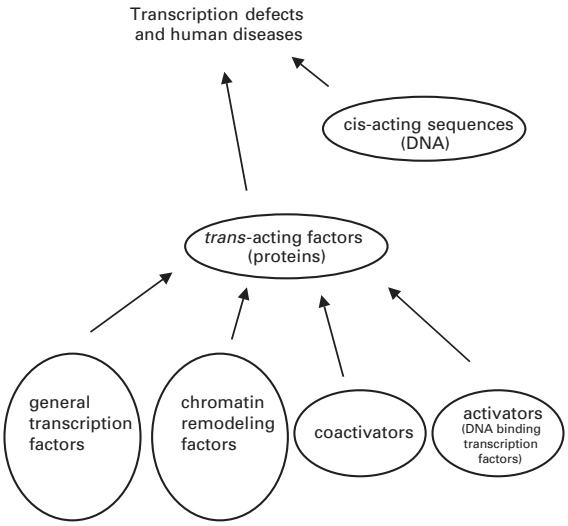


Figure 2

Transcription regulation defects and human disorders. A compilation of components that can be affected in mutations is shown.

General transcription factors mutations and human diseases

Among the large number of general transcriptions factors implicated in the initiation stage of transcription by RNA polymerase II, TFIIH is of special interest [30]. This multi-subunit protein complex is deficient in certain patients suffering from xeroderma pigmentosum. Xeroderma pigmentosum (XP) is characterized by an extreme sensitivity to sunlight-induced skin damage, hypo- and hyperpigmentation, and a marked predisposition to skin cancers. The disease is due to a deficiency in one of the mechanisms implicated in DNA repair. Two subunits of TFIIH (XP-B and XP-D) are essential for DNA nucleotide excision repair. Both proteins have helicase activity. These two subunits are mutated in some patients with XP [30]. In addition, a subset of XP patients exhibit ocular or neurological abnormalities such as mental retardation and poor physical and sexual development, but are not cancer prone (Cockayne syndrome [CS] and trichothiodystrophy [TTD]). These clinical symptoms are difficult to rationalise as DNA repair defects. The patients with mutations in the XP-B or XP-D genes have XP and CS or TTD clinical characteristics. The finding that TFIIH plays a role in both DNA repair and transcription led to the hypothesis that the unusual variety of clinical symptoms exhibited by patients with mutations in genes that encode subunits of TFIIH, do not result from defective DNA repair, but rather from defects in transcription [31].

Mutations of chromatin remodelling factors and co-activators in human diseases

Trans-acting proteins may affect gene expression by influencing chromatin structure. A group of genes in yeast, termed SWI/SNF, have emerged as strong candidates to encode proteins that directly alter the chromatin structure. The exact mode of action of these proteins, which form a large multi-protein complex, is still not clearly understood [4]. Some of these proteins have well characterized domains such as the ATPase/helicase domain, a putative DNA binding domain (zinc finger domain). In humans, several homologues of the SWI/SNF genes have been described. One of these is ATRX, α gene found to be mutated in an X-linked human syndrome, characterized by mental retardation, α-thalassaemia, genital abnormalities and facial dysmorphism [29].

The co-activator CREB binding protein (CBP) is also implicated in chromatin remodelling and has been found to be mutated in a rare human syndrome [32, 33]. One of the features that distinguishes active from inactive chromatin, is the acetylated state of histones. Histones are the principal structural proteins of eukaryotic chromosomes and play a crucial role in the packing of the DNA into chromatin. In transcriptionally active regions, the chromatin is less condensed and the histones are highly acetylated. The Rubinstein-Taby syndrome is a well defined syndrome characterized by facial abnormalities, broad thumbs, broad big toes and mental retardation. The disease is due to mutations in the gene coding for CBP which is a nuclear protein with histone acetylase activity [32, 33].

Activators and human inherited syndrome

Transcription activator mutations and development

The prototypical example of a transcription factor that is mutated in a human disease is Pit-1. This is a transcription factor characterized by a POU-type homeodomain (POU domain). It is expressed in the anterior pituitary gland and is required for the proliferation and survival of cells that produce growth hormone (GH), prolactin hormone (PRL) and thyroid stimulating hormone (TSH). Mutations in Pit-1 lead to mental retardation with combined pituitary hormone deficiency (CPHD). A number of humans with CPHD and Pit-1 gene mutations have been described. Since Pit-1 has different functions, phenotypic variability seen in patients with Pit-1 mutations is consistent with the different locations and types of Pit-1 mutations. In certain cases, the transmission is autosomal dominant. The mutated version of Pit-1 binds to its site and not only fails to stimulate gene expression but also prevents the binding of the normal protein. This mutated Pit-1 acts as a dominant negative [34, 35].

A growing list of transcription factors turn out to be mutated in developmental syndromes. Mutations in most of these factors lead to complex and pleiotropic phenotypes. Detailed study, as it has been the case for Pit-1, of how mutations elicit their effect will provide important insight into the mechanisms of developmental regulation by transcriptional control. For some of them, the clinical features are simpler; for example, the mutation of POU4F3 in a case of autosomal dominant inherited progressive hearing loss recently described. The study of POU4F3 allows a better understanding of certain physiological features of the sense of hearing and of the physiopathological processes that can lead to hearing loss [36].

Transcription activator mutations and steroid hormone receptors

Steroid and thyroid hormones, as well as vitamin D, retinoids and some nutrient metabolites such as fatty acids and prostaglandins, exert their effect by binding to members of the zinc-finger superfamily of nuclear hormone receptors. All members of this family share a poorly conserved amino-terminal region containing the transactivation domain responsible for gene activation, a zinc finger DNA binding domain, and a highly conserved carboxyl-terminal domain involved in ligand binding. The family is composed of the androgen receptor, the oestrogen receptor, the glucocorticoid receptor, the vitamin D_3 receptor, the thyroid hormone receptor, the retinoic acid receptor and several orphan receptors. Nuclear hormone receptor proteins regulate expression of specific target genes by ligand-dependent activation of the receptors with subsequent dimer formation and DNA binding [37, 38]. Hereditary mutations in nuclear hormone receptors lead to rare inherited syndromes of hormone resistance, characterized by abnormal hormone responses either in adults or during development. Among the best documented examples are androgen receptor mutations which can lead to a severe form of testicular feminization. There are other mutations with a less severe phenotype such as the Reifenstein syndrome which is characterized by a male phenotype with hypogonadism, cryptorchidism and hypospadias. Some mutations lead to an intermediate phenotype with incomplete testicular feminization. Interestingly, there is no correlation between the localization of the mutations (DNA binding domain or hormone binding domain) and the severity of the phenotype. However, the severity of the syndrome is correlated with the extent to which the function of the DNA binding domain or the hormone binding domain is disrupted. In addition, a very different syndrome called the Kennedy disease, characterized by spinal and bulbar muscular atrophy, is due to an expansion of a CAG (Q) trinucleotide repeat in the activation domain (amino-terminal part) of the androgen receptor [39].

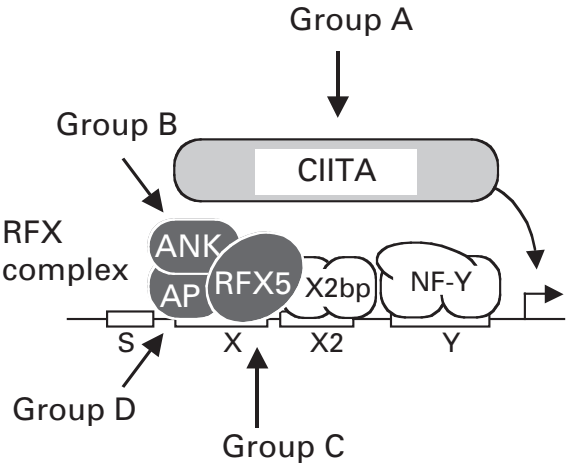


Figure 3

Molecular analysis of patients with MHC-II deficiency has led to the identification of four *trans*-acting proteins essential for MHC-II gene expression. MHC-II promoter contains three conserved *cis*-acting boxes, X, X_2 and Y. The four *trans*-acting factors mutated in each four complementation groups are indicated. Co-activator CIITA is finely regulated and plays a master role in the specificity and in the inducibility of the cells. RFX5, RFXAP and RFXANK are the three subunits of the RFX complex which bind the X box. Each mutated gene is essential and specific for MHC-II expression. X_2BP and NF-Y are also multi-protein complexes but not specific for MHC-II gene expression.

Transcription activator mutations in other miscellaneous human diseases

In contrast to most of the human diseases described above, some other diseases are characterized by less pleitropic phenotypes. The transcription factors mutated in these diseases are more specific to more or less well defined physiological processes. The identification and the mutation analysis of such transcription factors are of great interest for a better understanding of the biological pathway in which they are involved.

Maturity-onset diabetes of the young (MODY) is a genetically heterogeneous monogenic form of non-insulin dependent (type 2) diabetes mellitus (NIDDM). It is characterized by early onset, usually before 25 years of age and often in adolescence or childhood, and autosomal dominant inheritance. MODY subjects have normal insulin sensitivity but suffer from a defect in glucose-stimulated insulin secretion, suggesting that pancreatic β-cell dysfunction rather than insulin resistance is the primary defect [40]. Linkage studies have localized at least four genes that can be mutated in MODY. The hepatic nuclear factor (HNF) 4-alpha gene is mutated in MODY1 [41]. The HNF 1-alpha gene is mutated in MODY3 [42]. The insulin promoter factor-1 (IPF1) gene is mutated in MODY4 [43]. All three proteins are transcription factors. Auto-immune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) is a rare autoimmune disease. This is the first report of a single-gene defect causing a systemic human auto-immune disease. AIRE, the novel gene mutated in this disease, encodes a putative transcription factor with zinc finger motifs [44].

MHC class II deficiency syndrome is a rare primary immunodeficiency and is a very interesting example of transcription factor defects in human disease. Although the phenotype is homogenous, the disease is genetically heterogeneous. The MHC class II genes are intact and the disease has been shown to be caused by several regulation defects. Four specific transcription factors have been isolated, corresponding to four groups of patients. In group A, the CIITA gene is mutated. In groups B, C and D, three subunits of the multiprotein complex, RFX, RFXANK, RFX5, RFXAP, that bind to MHC class II promoters are defective. RFXANK is mutated in group B, RFX5 is mutated in group C, and RFXAP is mutated in group D. In MHC class II deficiency, these four transcription factors are not only essential but also specific for MHC class II expression (Fig. 3) [45, 46].

Transcription activators mutation and cancer

Cancer often results from the aberrant activation of specific genes known as oncogenes, which encode components of the cellular machinery that regulate normal growth processes. Either over expression of these genes or mutations, which lead to the formation of a more active product, can result in deregulated control of cellular proliferation

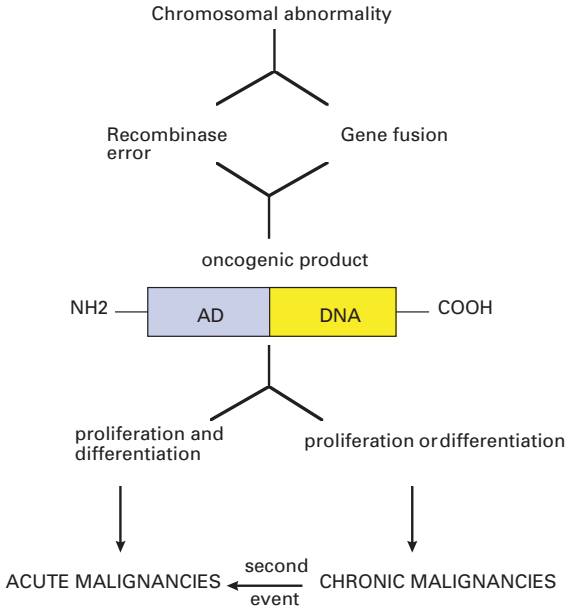


Figure 4

Schematic view of a recurrent molecular consequence of chromosomal abnormalities in human cancer. The creation of a chromosomal abnormality results in an oncogenic protein in the mutant cell. The structure of such proteins is usually composed of a transcriptional activation capacity (AD) associated with a DNA-binding domain (DNA-BD). As a result of this chimeric protein, cell proliferation and cell differentiation control can be affected in leukaemias and malignant solid tumours.

and conversion to the malignant phenotype. Considering the importance of transcription factors in the control of all cellular processes, it is not surprising that a large group of oncogenes encode transcription factors engaged in cell cycle control. Moreover, a number of developmental decisions as well as cell type differentiation, are under the control of oncogenic transcriptions factors, which are often the final signal targets of signal transduction pathways [47].

The role of oncogenic transcription factors in contributing to malignancy by altering programs of cell growth, differentiation and development is most obvious in the case of human leukaemia. Abnormal expression or inappropriate activation of specific oncogenes encoding transcription factors cause a variety of leukaemias which are characterized by genetic rearrangements induced by particular chromosomal translocations and gene amplifications. The translocation can result in the relocalisation of the gene encoding the transcription factor to a position where it is adjacent to a highly expressed gene (an immunoglobulin gene in B-cell leukaemia or a T-cell receptor gene in T-cell leukaemia). A second type of defect results when the breakpoints of the translocations fuse portions of two genes and create a hybrid protein comprised of domains derived from both genes. Such rearrangements presumably result in leukaemia because the hybrid protein exhibits properties that are distinct from those of either protein and it has the capacity to transform the cell into a malignant phenotype. Such fusion proteins give rise to hybrid transcription factors that possess the DNA-binding specificity of one parental protein and the activation characteristics of another [28, 48]. One example of gene fusion has been demonstrated in acute promyelocytic leukaemia (APL) with the translocation t(15;17). The translocation breakpoints consistently occur within the retinoic acid receptor α (RARα) gene (a member of the nuclear receptor superfamily), which is fused to the zinc finger-containing PML transcription factor in all cases of promyelocytic leukaemia (Fig. 4) [49].

In addition to leukaemia, the link between oncogenesis and transcriptional deregulation has been further strengthened by the recent cloning of chromosomal abnormalities associated with solid tumours. In Wilms' tumour, for example, the amino-terminal domain of the EWS transcription factor is fused to the last three zinc fingers of the DNA-binding domain of WTI. In Wilms' tumour, but also in melanoma of soft tissue, where EWS is implicated, the oncogenic conversion of EWS follows a common scheme of activation, exchanging its putative RNA binding domain with a part of different DNA binding domains: WTI in Wilms' tumour and ATF-1 in soft tissue melanoma. Thus, the DNA binding domains appear to be tumour specific [50].

A number of anti-oncogenes or tumour suppressor genes, like p53, encode transcription factors. The mutational inactivation or deletion of such genes can result in cancer. Germ-line mutations of transcriptions factors with tumour suppressor function are responsible for inherited cancers (Li-Fraumeni syndrome for p53, Wilms' tumour for WT1 or retinoblastoma for Rb) [51]. Somatic mutations of the same genes appear to play prominent roles in the development of a wide variety of more common sporadic human cancers including various carcinomas, brain tumours, sarcomas, and leukaemias [52].

Conclusion

Mutations in gene encoding proteins involved in the initiation of gene transcription play a key role in a wide variety of human diseases. Amongst these transcription factor mutations are the largest group. They represent more than 60 inherited human diseases, most of which are characterized by malformation syndromes. Mutation of transcription factors leads in most cases to pleiotropic effects because these factors control the expression of many genes. There are hundreds of regulatory transcription factors that function by binding DNA sequences within their target genes. Most of these proteins are encoded by multi-gene families with well conserved motifs (particularly the DNA binding domain). Members of a given family display the same structural fold for binding DNA and recognize similar DNA sequences.

How can specificity be obtained within such a complex world?

As mentioned earlier, combinatorial arrays of multiple proteins add specificity and stability to DNA-protein interactions. Homodimers and heterodimers can be formed between members of the same gene family. Alternatively, partnerships can be formed between proteins that belong to unrelated groups. In addition, two activators can act synergistically by binding to DNA cooperatively. Synergism can also be obtained by indirect interactions in which the two activators do not contact each other but simultaneously bind to different sites within a single transcription complex. All this complexity renders the correlation between a particular transcription factor mutation and a human phenotype difficult to analyze. A mutation that completely disrupts the DNA binding domain can result in a phenotype completely different from a mutation, that modifies the co-operative stability between two partners, for example. This has been well studied for the Pit-1 protein. Since Pit-1 has different functions in cells producing growth hormone (GH), prolactin hormone (PRL) and the thyroid stimulating hormone (TSH), the phenotypic variability seen in patients with Pit-1 mutations is consistent with the different locations and types of Pit-1 mutations. The consequences of mutations in transcription factors seem to be, in a majority of cases, a loss of function (deletions and premature protein truncations), reinforcing the idea that correct dosage is crucial and most diseases results from haplo-insufficiency. This evidence also suggests that increased doses of some transcription factors could be deleterious. However, while the genetic defects of most of the syndromes discussed here are identified, the exact role and the different modes of action of most of the transcription factors presented are not yet well understood.

Polymorphisms in the coding region of transcription factors have not been discussed in this work. Nevertheless, it will certainly become important in the future to better understand the pathogenesis of more common diseases. The combination of a mutation in a transactivator A and a polymorphism in another transactivator B, both being implicated in the activation of one or more genes, may lead to differential phenotypes.

In contrast to mutations in *trans*-acting factors, which are implicated in a growing list of pleiotropic human syndromes and a variety of cancers, only relatively few syndromes have been shown to be due to mutations in *cis*-acting DNA sequence elements. Regulatory regions such as enhancers can be localized very distant from the coding region and the scarcity of syndromes due to mutations in *cis*-acting elements may be explained by our ignorance of the regulatory elements of most of the genes studied.

Correspondence:
Jean Villard MD, PhD
Immunology and Transplant Unit
Hôpital Cantonal de Genève
24 rue Micheli-du-Crest
CH-1211 Geneva 4
E-Mail: jean.villard@hcuge.ch

References

1 Sauer F, Tjian R. Mechanisms of transcriptional activation: differences and similarities between yeast, Drosophila, and man. Curr Opin Genet Dev 1997;7:176–81.
2 Maniatis T, Goodbourn S, Fischer JA. Regulation of inducible and tissue-specific gene expression. Science 1987;236:1237–45.
3 Kingston RE, Bunker CA, Imbalzano AN. Repression and activation by multiprotein complexes that alter chromatin structure. Genes Dev 1996;10:905–20.
4 Emerson BM. Specificity of gene regulation. Cell 2002;109: 267–70.
5 Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 1989;245:371–8.
6 Young RA. RNA polymerase II. Annu Rev Biochem 1991;60: 689–715.
7 Koleske AJ, Young RA. An RNA polymerase II holoenzyme responsive to activators. Nature 1994;368:466–9.
8 Renkawitz R. Transcriptional repression in eukaryotes. Trends Genet 1990;6:192–7.
9 Levine M, Tjian R. Transcription regulation and animal diversity. Nature 2003;424:147–51.
10 Crossley M, Orkin SH. Regulation of the beta-globin locus. Curr Opin Genet Dev 1993;3:232–7.
11 Sharpe JA, Chan-Thomas PS, Lida J, Ayyub H, Wood WG, Higgs DR. Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice. EMBO J 1992;11: 4565–72.
12 Bentley DL. Regulation of transcriptional elongation by RNA polymerase II. Curr Opin Genet Dev 1995;5:210–6.
13 Roeder RG. Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. Cold Spring Harb Symp Quant Biol 1998;63:201–18.
14 Pabo CO, Sauer RT. Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem 1992; 61:1053–95.
15 Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell 1994;77:599–608.
16 Arany Z, Sellers WR, Livingston DM, Eckner R. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. Cell 1994;77:799–800.
17 Korzus E, Torchia J, Rose DW, Xu L, Kurokawa R, McInerney EM, et al. Transcription factor-specific requirements for coactivators and their acetyltransferase functions. Science 1998;279:703–7.
18 Crossley M, Ludwig M, Stowell KM, De Vos P, Olek K, Brownlee GG. Recovery from hemophilia B Leyden: an androgen-responsive element in the factor IX promoter. Science 1992; 257:377–9.
19 Crossley M, Brownlee GG. Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. Nature 1990;345:444–6.
20 Martin DI, Tsai SF, Orkin SH. Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. Nature 1989;338:435–8.
21 Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W. Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene 1997; 205:73–94.

22 Faustino P, Lavinha J, Marini MG, Moi P. beta-Thalassemia mutation at –90CÆT impairs the interaction of the proximal CACCC box with both erythroid and nonerythroid factors. Blood 1996;88:3248–9.
23 Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, et al. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature 1997;386:847–51.
24 Morzycka-Wroblewska E, Munshi A, Ostermayer M, Harwood JI, Kagnoff MF. Differential expression of HLA-DQA1 alleles associated with promoter polymorphism. Immunogenetics 1997;45:163–70.
25 Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proc Natl Acad Sci USA 1997;94:3195–9.
26 Martin MP, Dean M, Smith MW, Winkler C, Gerrard B, Michael NL, et al. Genetic acceleration of AIDS progression by a promoter variant of CCR5. Science 1998;282:1907–11.
27 Latchman DS. Transcription-factor mutations and disease. N Engl J Med 1996;334:28–33.
28 Sanchez-Garcia I. Consequences of chromosomal abnormalities in tumour development. Annu Rev Genet 1997;31:429–53.
29 Gibbons RJ, Picketts DJ, Villard L, Higgs DR. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell 1995;80:837–45.
30 Liu J, Akoulitchev S, Weber A, Ge H, Chuikov S, Libutti D, et al. Defective interplay of activators and repressors with TFIH in xeroderma pigmentosum. Cell 2001;104:353–63.
31 Lehmann AR. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. Genes Dev 2001; 15:15–23.
32 Giles RH, Peters DJ, Breuning MH. Conjunction dysfunction: CBP/p300 in human disease. Trends Genet 1998;14:178–83.
33 Hartzog GA, Winston F. Nucleosomes and transcription: recent lessons from genetics. Curr Opin Genet Dev 1997;7: 192–8.
34 Andersen B, Rosenfeld MG. Pit-1 determines cell types during development of the anterior pituitary gland. A model for transcriptional regulation of cell phenotypes in mammalian organogenesis. J Biol Chem 1994;269:29335–8.
35 Radovick S, Cohen LE, Wondisford FE. The molecular basis of hypopituitarism. Horm Res 1998;49(Suppl 1):30–6.
36 de Kok YJ, van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, et al. Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. Science 1995;267:685–8.
37 Meier CA. Regulation of gene expression by nuclear hormone receptors. J Recept Signal Transduct Res 1997;17:319–35.
38 Mangelsdorf DJ, Evans RM. The RXR heterodimers and orphan receptors. Cell 1995;83:841–50.
39 Ferlini A, Patrosso MC, Guidetti D, Merlini L, Uncini A, Ragno M, et al. Androgen receptor gene (CAG)n repeat analysis in the differential diagnosis between Kennedy disease and other motoneuron disorders. Am J Med Genet 1995;55:105–11.
40 Herman WH, Fajans SS, Ortiz FJ, Smith MJ, Sturis J, Bell GI, et al. Abnormal insulin secretion, not insulin resistance, is the genetic or primary defect of MODY in the RW pedigree. Diabetes 1994;43:40–6.
41 Yamagata K, Furuta H, Oda N, Kaisaka PJ, Menzel S, Cos NJ, et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature 1996; 384:458–60.
42 Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M, et al. Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). Nature 1996;384:455–8.
43 Stoffers DA, Ferrer J, Clarke WL, Habener JF. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet 1997;17:138–9.
44 An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. The Finnish-German APECED Consortium. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. Nat Genet 1997;17:399–403.
45 Reith W, Mach B. The bare lymphocyte syndrome and the regulation of MHC expression. Annu Rev Immunol 2001;19: 331–73.
46 Villard J, Masternak K, Lisowska-Grospierre B, Fischer A, Reith W. MHC class II deficiency: a disease of gene regulation. Medicine 2001;80:405-18.
47 Nebert DW. Transcription factors and cancer: an overview. Toxicology 2002;181–182:131–41.
48 Rabbitts TH. Chromosomal translocations in human cancer. Nature 1994;372:143–9.
49 Koken MH, Puvion-Dutilleul F, Guillemin MC, Viron A, Linares-Cruz G, Stuurman N, et al. The t(15;17) translocation alters a nuclear body in a retinoic acid-reversible fashion. EMBO J 1994;13:1073–83.
50 Gerald WL, Rosai J, Ladanyi M. Characterization of the genomic breakpoint and chimeric transcripts in the EWS-WT1 gene fusion of desmoplastic small round cell tumor. Proc Natl Acad Sci USA 1995;92:1028–32.
51 Malkin D. Germline p53 mutations and heritable cancer. Annu Rev Genet 1994;28:443–65.
52 Sherr CJ, McCormick F. The RB and p53 pathways in cancer. Cancer Cell 2002;2:103–12.

Swiss
Medical Weekly

Official journal of
the Swiss Society of Infectious disease
the Swiss Society of Internal Medicine
the Swiss Respiratory Society

The many reasons why you should choose SMW to publish your research

What Swiss Medical Weekly has to offer:

- SMW's impact factor has been steadily rising, to the current 1.537
- Open access to the publication via the Internet, therefore wide audience and impact
- Rapid listing in Medline
- LinkOut-button from PubMed with link to the full text website http://www.smw.ch (direct link from each SMW record in PubMed)

- No-nonsense submission – you submit a single copy of your manuscript by e-mail attachment
- Peer review based on a broad spectrum of international academic referees
- Assistance of our professional statistician for every article with statistical analyses

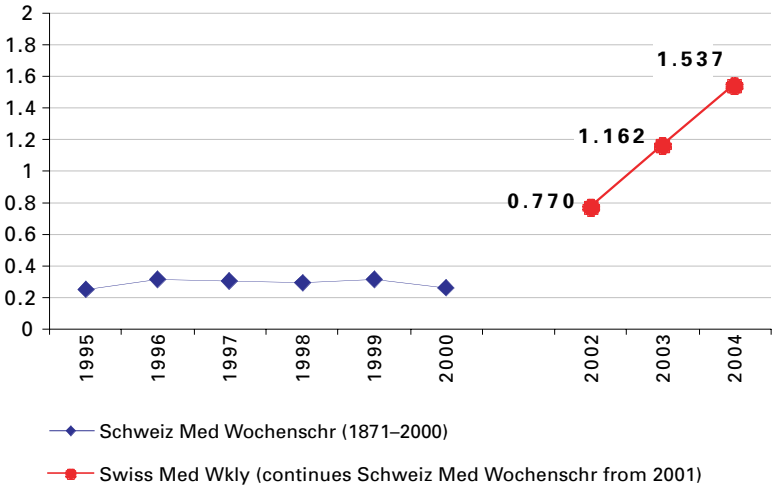
- Fast peer review, by e-mail exchange with the referees
- Prompt decisions based on weekly conferences of the Editorial Board
- Prompt notification on the status of your manuscript by e-mail
- Professional English copy editing
- No page charges and attractive colour offprints at no extra cost

We evaluate manuscripts of broad clinical interest from all specialities, including experimental medicine and clinical investigation.

We look forward to receiving your paper!

Guidelines for authors:
http://www.smw.ch/set_authors.html

Impact factor Swiss Medical Weekly



EMH FMH Schwabe
Editores Medicorum Helveticorum

All manuscripts should be sent in electronic form, to:

EMH Swiss Medical Publishers Ltd.
SMW Editorial Secretariat
Farnsburgerstrasse 8
CH-4132 Muttenz

Manuscripts: submission@smw.ch
Letters to the editor: letters@smw.ch
Editorial Board: red@smw.ch
Internet: http://www.smw.ch