Transcription factors

Transcription factors (TFs) are proteins that are involved in the initiation of transcription (the transcription of hereditary information from a gene (from DNA (nucleic acid) to RNA).

They bind to individual *cis*-regulatory elements (specific very short DNA sequences in the promoters, enhancers, silencers), thus facilitating the binding of the corresponding RNA polymerase, or other proteins. Binding of TFs to a specific region of DNA is accomplished by hydrogen and ionic bonds, or hydrophobic interactions between structurally (3D) complementary regions of proteins and the large and small groove of DNA. **Prokaryotic** RNA polymerase does not require TF for its activity, transcription in **eukaryotes** is dependent on the presence of TF.

Through them, gene expression is adapted to the needs of the cell or the whole organism (e.g. hormones, hypoxia can stimulate the expression - transcription of certain genes). This adaptation can be very fast, up to several thousand transcripts of a single DNA segment can be produced within an hour. Some TFs must first be activated, e.g. by phosphorylation or inhibitor removal.

We distinguish:

1. **general TF** - occurrence in all types of cells,
2. **special TF** - occurrence only in certain cells,
3. **basal TF** - factors required for induction of basal transcription (in cells with low transcriptional activity).

Basic concepts

If the factor blocks transcription it is a *repressor*, otherwise it is an *inducer*.

DNA sequences to which regulatory proteins bind - **control** (responsive) **elements** :

- ***promoter*** - a piece of DNA near a gene (RNA sequence of bases that have specific binding sites for RNA polymerase) that is involved in regulating its expression (e.g. TATA box, CAT box)
    - *promoter strength* - the ability of a promoter to initiate transcription, it is the affinity of the promoter region for RNA polymerase. The frequency of transcription of the adjacent gene depends on the strength of the promoter. Strong promoters have sequences in the -35 and -10 region (Pribnow box) identical to the conventional sequence, the more they differ from the conventional sequence, the weaker the promoter.
- ***operator*** - a regulatory region on DNA that lies between the promoter and the beginning of transcription. Binding of an active repressor to the operator blocks transcription and thus the expression of structural genes.
- ***enhancer*** - DNA sequence that binds activation factors
- ***silencer*** - DNA sequence that binds inhibitory factors

Negative gene regulation (i.e., "switch off" of a gene) is provided by repressors - some bind to DNA only in the absence of a specific ligand, and once bound, the promoter becomes accessible to RNA polymerase. Other repressors block transcription in the presence of the ligand.

Positive gene regulation - transcription inhibited by ligand binding to the inducer, but many inducers are active only when the ligand is bound.

Function and biological role of transcription factors

Regulation of basal transcription

- **GTFs** (general transcription factors) - essential for transcription
- many of them do not bind to DNA but are part of a preinitiation complex that reacts directly with RNA polymerase II
- most common: TFIIA, TFIIB, TFIID (includes a subunit called *TATA binding protein* (TBP) - binds specifically to the TATA box sequence), TFIIE, TFIIF and TFIIH

Cellular development

- regulate cell differentiation and determination based on signals
- TF family **Hox** is important for proper body organization
- TF coded **SRY** (Sex-determining region of Y) - determination of human sex

Response to intercellular signals

- part of a signaling cascade (activation x suppression)
- e.g. **estrogen signalling**: TF is part of the estrogen receptor, which, when activated, travels to the nucleus where it regulates the transcription of certain genes

Response to the external environment

- TFs also regulate signaling cascades of exogenous origin
- **Heat shock factor** (HSF) - activates genes allowing survival at higher temperatures
- **Hypoxia inducible factor** (HIF) - survival in oxygen-deficient environments

Cell Cycle Control

- hl. TFs that are **oncogenes** (e.g. myc) and **tumorsuppressors** (e.g. p53) - role in cell growth and apoptosis

Regulation of transcription factor activity

Synthesis of Transcription Factor

- each of the transcription and translation steps TF can be regulated
- TFs have the ability to regulate their own production by negative feedback (binding of the TF to the DNA of its own gene)

Nuclear localization

- TFs are transcribed in the nucleus but translated in the cytoplasm
- contain nuclear localization signals - they are directed to the nucleus
- many TFs, including nuclear receptors, must first bind a ligand before moving into the nucleus

Activation by chemical modifications or ligand binding

- ligand binding to TF affects its intracellular localization, activation and ability to bind to DNA or cofactors
- chemical modification is another option for TF activation (e.g. STAT proteins must be phosphorylated to bind to DNA)

DNA-binding site accessibility

- DNA of heterochromatin is inaccessible to TF - heterochromatin has to be changed to euchromatin (usually by histone modifications), the DNA binding site can also be occupied (and thus blocked) by another TF (antagonism of TF in the regulation of a given gene - activator and repressor)

Availability of cofactors

- most TFs do not act independently but need cofactors

Structure of transcription factors

Transcription factors consist of the following domains:

DNA-binding domain (DBD)

- binds to specific DNA sequences, so-called **responsive elements** (enhancer, promoter) adjacent to regulated genes, examples:
  - *lambda repressor-like*
  - *srf-like* (serum response factor)
  - *GCC box*
  - *Zn2/Cys6*
  - *basic helix-loop-helix*
  - *homeodomain proteins* - role in developmental regulation, bind to homeobox DNA sequences that encode other TFs (short nucleotide sequence, identical in different genes and organisms, role in expression of relevant genes)
  - *multi-domain Cys2His2 zinc fingers*
  - *basic leucine zipper* (bZIP)

Trans-activating domain (TAD)

- contains binding sites (so-called **AFs - activation functions**) for other proteins functioning as transcriptional coregulators

Ligand binding domain (ligand, binding domanin, signal sensing domain, SSD)

- may not be present
- responds to external signals and transmits them to the rest of the transcription complex - resulting in **up/down regulation of gene expression**

Binding sites of transcription factors (responsive elements)

Transcription factors usually **interact** with their binding sites using **hydrogen bridges and Van der Waals forces**. Due to the nature of these chemical interactions, most TFs bind to DNA **specifically**.

However, not all binding site bases need to actually interact with the TF. Moreover, some of these interactions may be weaker than others. TFs are able to bind several related sequences, each with different strengths (e.g., although the main binding site for TATA binding protein (TBP) is the TATAAAA sequence, TBP can also bind to TATAT or TATATAA). However, binding to all of these compatible sequences is unlikely, as DNA accessibility and the number of cofactors also interact. Thus, it is still difficult to predict where the TF will eventually bind.

Classes of transcription factors

Mechanical cutting

1. **Transcription factors of the preinitiation complex** (general transcription factors) - TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH - are ubiquitous and interact with the promoter (often the TATA box) of structural genes, important in vertebrate and invertebrate development
2. **Upstream transcription factors** (UTF) - upstream - towards the 5' part, proteins that bind to the regulatory part of the RNA polymerase I promoter in the -110 to -180 position, presence is not necessary for transcription initiation, but increases its efficiency many times (may also act repressively)
3. **Inducible transcription factors** - same as UTF, but need to be activated or inhibited

Functional division

1. **Constitutive** - present (and active) in the cell all the time - general transcription factors, Sp1, NF1, CCAAT
2. **Conditionally active** - activation required
    - ***Developmental*** (cell-specific) TFs - expression is tightly controlled, but after expression they do not require further activation - GATA, HNF, PIT-1, MyoD, Myf5, Hox, Winged Helix
    - ***Signal-dependent*** TF - need an external signal to activate
        - *extracellular ligand-dependent* - nuclear receptors
        - *intracellular ligand-dependent* - activated by small ic. molecules - SREBP, p53, orphan nuclear receptor
        - *cell membrane receptor-dependent* second messenger cascades causing phosphorylation of TF
            - resident nuclear factors - are in the nucleus regardless of activation state - CREB, AP-1, Mef2
            - latent cytoplasmic factors - in the cytoplasm in inactive form, after activation translocated to the nucleus - STAT, R-SMAD, NF-kB, Notch, TUBBY, NFAT

Occurrence of transcription factors

Transcription factors are essential for the regulation of gene expression and are therefore found in all living organisms. The number of TFs in an organism is **directly proportional to the size of the genome**. There are approximately 2600 proteins with DNA-binding domains in the human genome and most of them function as TFs. They encode about 10% of all genes and thus represent the largest single family of human proteins. About 2000 TFs are capable of cooperating with each other, allowing sensitive regulation of each human gene.

Importance of transcription factors in human disease

Because of their important roles in development, intercellular signalling and the cell cycle, TF mutations have been linked to the pathogenesis of several human diseases.

Rett Syndrome

- mutation **MECP2** (methyl CpG binding protein 2) - in all cells of the body, a repressor of transcription, in high concentrations in neurons - CNS maturation and synapse formation
- mutations - formation of excess proteins in neurons (acts as a repressor of transcription)
- a syndrome **characterized** by slow and abnormal development of apparently healthy newborn girls;

development stops between 6 and 18 months of age, skull circumference grows slowly, muscle tone disappears, involuntary movements and seizures begin, scoliosis may develop
- brain atrophy, hyperammonemia, progressive dementia
- **Inheritance** is probably dominant with linkage to the X chromosome
- MECP2 mutations also in neonatal encephalopathy, Angelman syndrome, autism

Diabetes - MODY (Maturity onset diabetes of the young)

- a rare **hereditary form of diabetes** manifesting by age 25 and not requiring (for at least 5 years) insulin
- approximately 5% of all DM cases
- inheritance is usually autosomal dominant
- different types of MODY with mutations of ***hepatic nuclear factor*** (HNF; belonging to steroid/thyroid hormones, plays an important role in embryogenesis, development and cell differentiation, ensures normal activity of entoderm of liver, intestine, kidney and ß-cells of pancreas - **MODY1** - rapid progression, hyperglycemia, insulinopenia) or ***insulin promoter factor-1*** (IPF1; 13. chr., **MODY 4**, affects early pancreatic development and expression of some pancreatic ß-bb genes, including insulin gene, glucokinase, somatostatin gene expression in pancreatic D cells)

Developmental Verbal Dyspraxia (DVD)

- mutation **FOXP2**
- individuals are not capable of the finely coordinated movements necessary for speech production
- aphasia, verbal stunting, a specific developmental deviation at the word level, i.e., a disorder in word realization, word structure, a particular part of a word

Autoimmune Diseases

- **IPEX** (*Immunodysregulation Polyendocrinopathy Enteropathy X-linked Syndrome*)
- mutation **FOXP3** (the main regulator of Treg development and function, mutations also occur in DM, RS, asthma or renal diseases)
- a rare autoimmune disease affecting suppressor T-lymphocytes

Malignancy

- many TFs are tumorsuppressors or oncogenes - their mutation or dysregulation is associated with tumorigenesis
- e.g. **Li-Fraumeni syndrome** - mutation in the tumour suppressor p53 (in 50% of cases); rare autosomal dominant tumour syndrome; individuals from such affected families are more likely to develop malignant disease often at an early age
- the predominant types of tumours are soft tissue sarcomas, osteosarcomas, CNS tumours, leukaemia, adrenal medulla carcinoma (ADCC) and premenopausal breast tumours

Sources

Related articles

- Transcription
- Regulation of gene expression in eukaryotes
- Regulation of gene expression in prokaryotes

References

- SILBERNAGEL, S – LANG,. *Atlas patofyziologie člověka.* 1. edition. Prague : Grada Publishing, 2001. 404 pp. ISBN 80-7169-968-3.
- ALBERTS, B – ALEXANDER,, et al. *Molecular Biology of the Cell.* 4. edition. Garland, 2002. pp. 404. ISBN 0-8153-4072-9.
- LATCHMAN, DS. Transcription factors: an overview. *Int. J. Biochem. Cell Biol..* 1997, y. 29 (12), p. 1305-12, ISSN 1357-2725.
- KARIN, M. Too many transcription factors: positive and negative interactions. *New Biol..* 1990, y. 2, p. 126-31, ISSN 1552-4450.
- WIKIPEDIA EN https://en.wikipedia.org/wiki/Transcription_factor