1 *Superclass* : Basic domains

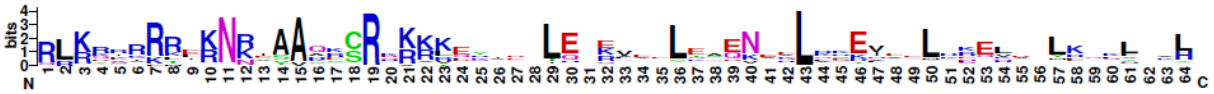
Proteins belonging to this superfamily contact the DNA through a basic region which free in solution has a random structure, but becomes alpha-helically folded upon binding to DNA (PMID 2120592, PMID 2145515, PMID 1312255, PMID 8479534, PMID 7926781). Most prominent classes of this group are the bZIP and bHLH proteins. There is no 'consensus' motif for all basic domains of this superfamily. Even the basic regions of bZIP and bHLH factors diverge greatly, although their mode of interacting with the DNA is strictly homologous: A specific alpha-helical dimerization domain provides the linkage between two DNA-contacting basic regions which adopt helical conformation when their positively charged side chains are neutralized by the phosphate backbone of the DNA. (TRANSFAC class description C0033)

1.1 *Class* : Basic leucine zipper factors (bZIP)

Description:

A DNA-binding basic region is followed by a leucine zipper. The leucine zipper consists of repeated leucine residues at every seventh position and mediates protein dimerization as a prerequisite for DNA-binding. The leucines are directed towards one side of an alpha-helix. The leucine side chains of two polypeptides are thought to interdigitate upon dimerization (knobs-into-holes model). The leucine zipper dictates dimerization specificity. Upon DNA-binding of the dimer, the basic regions adopt alpha-helical conformation as well. Possibly, a sharp angulation point separates two alpha-helices of the subregions A and B leading to the scissors grip model for the bZIP-DNA complex. The DNA is contacted through the major groove over a whole turn.
(TRANSFAC® class description C0008)

Logo plot:



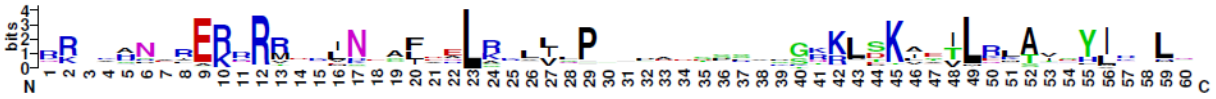
1.2 *Class* : Basic helix-loop-helix factors (bHLH)

Description:

A DNA-binding basic region is followed by a motif of two potential amphipathic alpha-helices connected by a loop of variable length, possibly an omega loop. This putative helix-loop-helix (HLH) motif mediates dimerization as a prerequisite for DNA-binding. Upon dimerization, the alpha-helical content may significantly increase supporting the model of a four-helix bundle dimerization interface. Probably, the HLH domains of both dimer constituents are arranged in a parallel

orientation. The basic region dictates DNA-binding specificity. Upon DNA-binding of the dimer, the basic regions adopt alpha-helical conformation as well.
(TRANSFAC® class description C0010)

Logo plot:

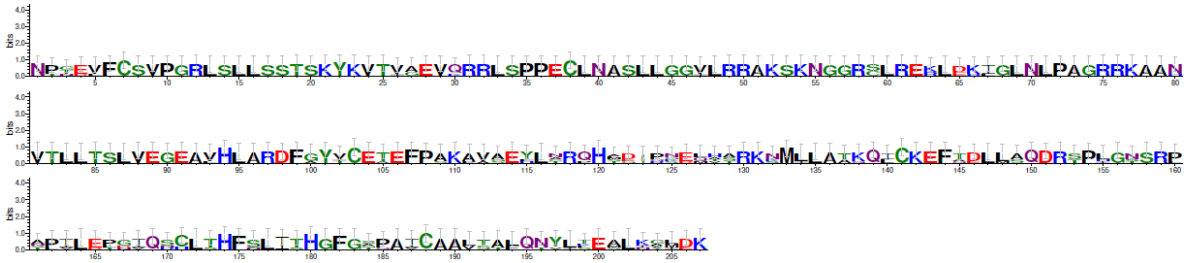


1.3 *Class* : Basic helix-span-helix factors (bHSH)

Description:

TRANSFAC® class description C0032: Examining human AP-2 as paradigm of this structural class, a DNA-contacting domain was identified which comprises several clusters of basic amino acid side chains. C-terminally adjacent to it, a dimerization domain was found which consists of two putative alpha-helices separated by a span of over 80 amino acids. This HSH dimerization domain can functionally replace the ZIP dimerization domain of c-Jun. (PMID 1998122)

Logo plot:



© geneXplain GmbH, 2019

2 *Superclass* : Zinc-coordinating DNA-binding domains

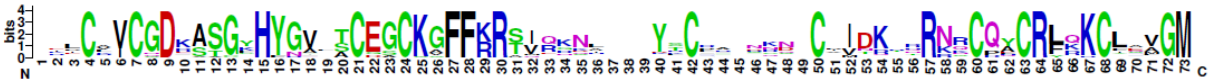
In polymerase III transcription factor TFIIIA, a repetitive pattern of cysteine and histidine residues was discovered and, together with the known zinc-dependence of TFIIIA, led to the model of 'zinc fingers' of the Cys2His2 type (PMID 3191912, PMID 6196359). Shortly after, sequences coding for the estrogen receptor were cloned and revealed a somewhat similar arrangement of cysteine residues only (Cys4), with no additional homologies (PMID 3755102, PMID 4040853). Nevertheless, a zinc finger model was proposed for these factors as well and was subsequently proven, with a minor alteration in the Cys pattern (PMID 3780678). In addition to these two large classes of zinc finger proteins, there are many factors revealing Cys4 zinc fingers of variable composition, such as the GATA-factors, some fungal regulators and adenovirus E1A. Finally, the yeast GAL4-like regulators are grouped together as Cys6 'zinc clusters'.
(TRANSFAC class description C0034)

2.1 *Class* : Nuclear receptors with C4 zinc fingers

Description:

Zinc finger motif of nuclear receptor type. Two such motifs differing in size, composition and function are present in each receptor molecule. Each finger comprises 4 cysteine residues coordinating one zinc ion. The second half including the second cysteine pair has alpha-helix conformation, the helix of the first finger binds to the DNA through the major groove. The sequence between the first two cysteines of the second finger mediates dimerization upon DNA-binding.
(TRANSFAC® class description C0002)

Logo plot:



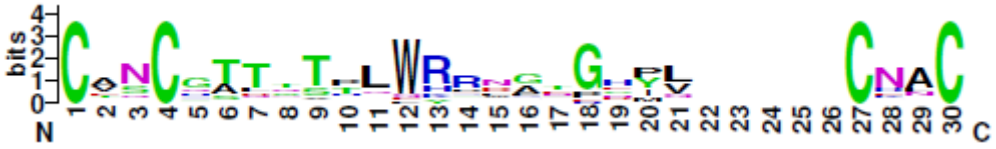
2.2 *Class* : Other C4 zinc finger-type factors

Description:

Zinc finger motif of C4 type. Each finger comprises 4 cysteine residues presumably coordinating one zinc ion. TRANSFAC® class description C0003: Zinc finger motif of GATA-type. Can be subclassified into two sub-types, one of which is found in animal factors, the other one in plant factors. In fungi factors from both groups have been found. Most known animal GATA factors contain two zinc finger domains; most known fungal, plant and slime mold GATA factors only one

(PMID 9581501). In case of the two-fingered factors, as far as it has been investigated, only the C-terminal zinc finger plus an adjacent basic region are essential and sufficient for DNA-binding (PMID 2276623, PMID 1406646, PMID 8446581). Each finger comprises 4 cysteine residues coordinating one zinc ion. The DNA-binding zinc finger is composed of two irregular antiparallel beta-sheets and an alpha-helix; similar in structure, although not in sequence to the zinc finger of the glucocorticoid receptor DNA binding domain. The alpha-helix and the loop connecting the two antiparallel beta-sheets interact with the major groove of the DNA. The basic region carboxyl-terminal to the zinc finger wraps around into the minor groove (PMID 8332909).
(TRANSFAC® class description C0004)

Logo plot:



2.3 *Class* : C2H2 zinc finger factors

Description:

Zinc finger motif of TFIIIA/Krueppel type. Each finger comprises 2 cysteine and 2 histidine residues coordinating one zinc ion, in some cases one histidine is replaced by another cysteine. The zinc ion is essential for DNA-binding. The first half of the finger sequence is generally organized in two antiparallel beta-strands, the second half as an alpha-helix and partly as an 3_{10}-helix. Hydrophobic contacts between the beta-strands and the alpha-helix are formed by the conserved phenylalanine and leucine residues. The alpha-helix binds to DNA through the major groove. The finger link frequently is TGEKPY.
(TRANSFAC® class description C0001)

Logo plot:
No logo plot at Class level available.

2.4 *Class* : C6 zinc cluster factors

Description:

Zinc finger motif of GAL4-type. Present to date in many fungal regulators, **but no human example is known so far**. The proteins are typically, though not exclusively transcriptional activators. Six cysteine residues coordinate two zinc ions, i. e. two of the thiol groups coordinate two zinc ions (six cysteines binuclear cluster DNA-binding domain). All GAL4-type proteins possess a single N-terminal zinc finger with an adjacent basic region, which is critical for sequence-specific recognition. Each C3-repeat unit of the structural motif of six cysteines forms short alpha-helical structures separated by a loop with a proline-associated turn. A number of zinc cluster proteins contain leucine zipper-like heptad repeat motifs C-terminal to the zinc-cluster. These motifs have

© geneXplain GmbH, 2019

been shown to form coiled-coil structures involved in protein-protein interactions (often homodimerization). The GAL-4 protein binds as a dimer to DNA sites characterized by the presence of two CGG triplets separated by 11 base pairs. The zinc(II)-binding cluster lies in the DNA major groove and contacts three base pairs.
(TRANSFAC® class description C0005)

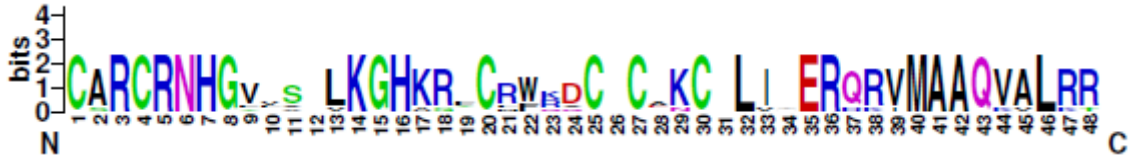
Logo plot:
No logo plot at Class level available.

2.5 *Class* : DM-type intertwined zinc finger factors

Description:

The known transcription factors exhibiting a DNA-binding domain of intertwined zinc fingers belong to the DM group of proteins, named after Drosophila Doublesex and C. elegans Meb-3. Intertwined are one zinc finger of C2HC and one of HCC2 type, each binding one zinc ion. Together, they follow the consensus pattern C--C--h--------H---C----c-c-c, where upper and lower case letter refer to either zinc finger, and dashes symbolize any other amino acid. The most C-terminal cysteine is part of a short alpha-helical structure, which extends into a carboxy-terminal tail that stably folds into an alpha helix upon binding to DNA. This recognition helix binds in the minor groove of the DNA. In contrast to other minor-groove binders, it does not induce a sharp bending of the DNA. (PMID 16382145).

Logo plot:

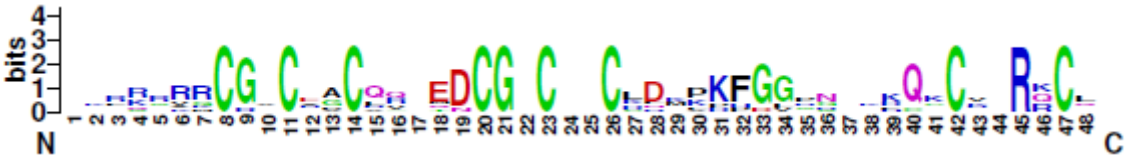


2.6 *Class* : CXXC zinc finger factors

Description:

CXXC zinc fingers contain two zinc ion binding 4-cysteine patterns, each of them comprising a CGxCxxC motif and a distantly located cysteine. One of these patterns (with lower-case Cs) is embedded within the other (with upper-case Cs): C-x(2)-C-x(2)-C-x(5,6)-c-x(2)-c-x(2)-c-x(9,15)-c-x(4)-C. The second and third cysteine of each motif lie within a short alpha-helix. The overall finger structure also forms a two-stranded antiparallel beta-sheet, the two beta-strands being located immediately N- and C-terminally adjacent to the CXXC zinc finger pattern. The CXXC zinc finger binds to DNA at CpG-containing sequences, most of the contacts are exerted by residues of an extended loop in the x(9,15) region of the above signature. (PMID 16990798)

Logo plot:

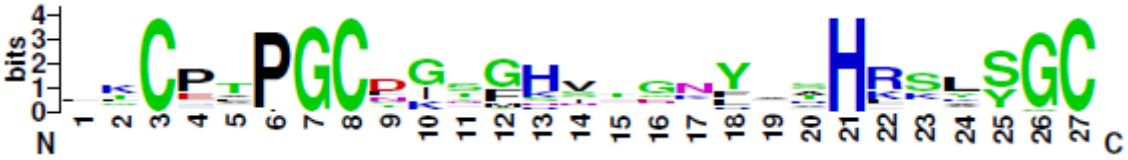


2.7 *Class* : C2HC zinc finger factors

Description:

The individual zinc finger modules of this class are characterized by a tetrahedrical coordination of one zinc ion by three cysteines and one histidine, arranged according to the pattern C-X(2,4)-C-X(12)-H-(3,5)-C, similar to C2H2 zinc fingers. In contrast to them, they form a rather irregular, but compact structure without recognizable secondary structure elements. At least some C2HC zinc fingers bind sequence-specifically to DNA, where the whole C2HC zinc finger sites in the major groove. (PMID: 14744132, 18073212) – This class should not be mixed up with CCHC zinc fingers, which follow the signature C--C----H----C (i.e., having much shorter spacing between the leading two C and the H), and are mostly RNA-binders.

Logo plot:



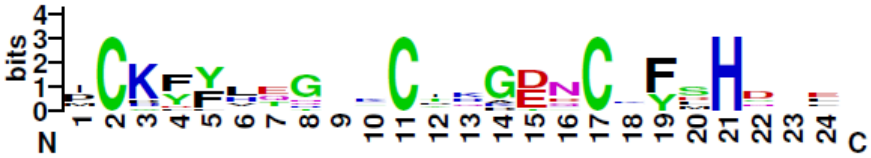
2.8 *Class* : C3H zinc finger factors

Description:

These factors comprise a putative zinc finger motif with the signature C------C-----C---H. DNA-binding has been proven only for three proteins of this class, ZC3H8 (2.8.1.0.1), of which closely related proteins have been included for similarity, ZGPAT (2.8.2.0.1), with no close relatives found, and RC3H2 (2.8.4.0.1). Most, if not all, other proteins of this class appear to be RNA-binders and be involved more likely in splicing rather than transcriptional regulation.

© geneXplain GmbH, 2019

Logo plot:

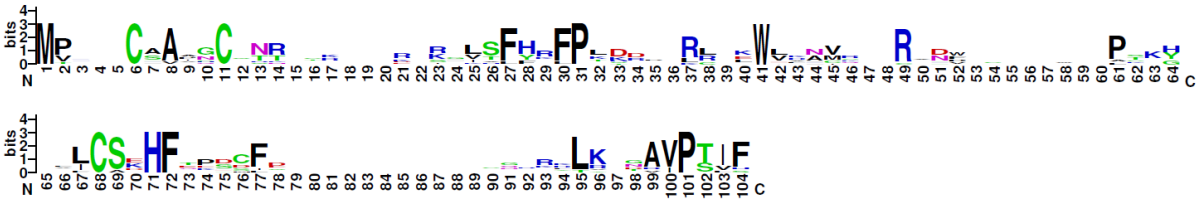


2.9 *Class* : C2CH THAP-type zinc finger factors

Description:

THAP-type zinc fingers follow the signature C- x(2-4)-C-x(35-50)-C-x2-H. Compared with 'canonical' C2H2 zinc fingers, those of THAP type exhibit an unusually long spacing between the first pair of cysteines and the following cysteine-histidine pair, which together coordinate one zinc ion. This long spacer adopts a beta-alpha-beta structure, with the second beta-strand right in front of the third cysteine. An additional short 3,10-helix comprises the zinc-coordinating histidine residue. The two beta-strands form a beta-sheet that inserts into the major groove, whereas the loop following the zinc finger exerts minor groove contacts. (PMID 18073205, PMID 20010837)

Logo plot:



3 *Superclass* : Helix-turn-helix domains

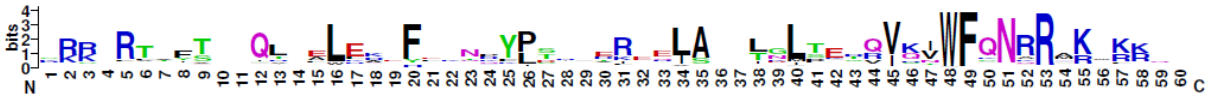
The helix-turn-helix superclass is a particularly large and heterogeneous family of transcription factors. Their DNA-binding motif appeared very early during evolution since it is found in prokaryotic and bacteriophage regulators as well as in mammalian transcriptional activators and repressors (PMID 6236744, PMID 6330566, PMID 6429549). Presumably due to this property, most of them fulfill elementary functions in eukaryotes such as developmental regulation and determination of differentiation processes. In general, they comprise a DNA-recognition helix which inserts into the major DNA groove, another one which exerts unspecific stabilizing backbone interactions, and in most cases additional helices which help to stabilize the overall structure of the DNA-binding module.
(TRANSFAC class description C0035)

3.1 *Class* : Homeo domain factors

Description:

Three consecutive alpha-helix structures, helix 3 contacts mainly the major groove of the DNA, some contacts at the minor groove are observed as well. Helix 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators.
(TRANSFAC® class description C0006)

Logo plot:



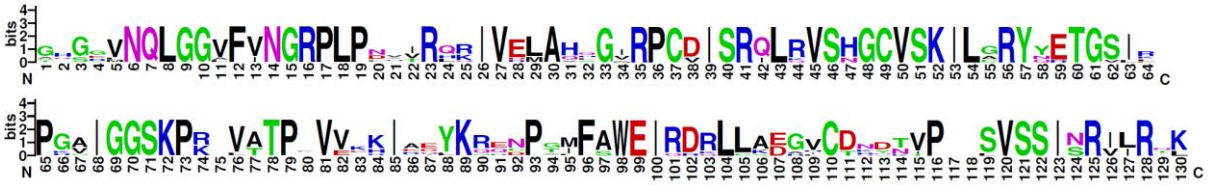
3.2 *Class* : Paired box factors

Description:

DNA-binding domain of approximately 130 amino acid residues. Its N-terminal half is basic, its C-terminal half highly charged in general. It probably comprises 3 alpha-helices.
(TRANSFAC® class description C0017)
A paired box is combined with a C-terminally located homeo domain. The paired box comprises approximately 130 amino acid residues. Its N-terminal half is basic, its C-terminal half highly charged in general. It probably comprises 3 alpha-helices. Three consecutive alpha-helix structures, helix 3 contacts mainly the major groove of the DNA, some contacts at the minor groove are observed as well. Helix 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators.
(TRANSFAC® class description C0018)

© geneXplain GmbH, 2019

Logo plot:



3.3 *Class* : Fork head / winged helix factors

Description:

Identified by homology between HNF-3A and fkh. The domain comprises approx. 110 AA. Analysis of the crystal structure has revealed a compact structure of three alpha-helices, the third alpha-helix being exposed towards the major groove of the DNA. The domain also exerts minor groove contacts. Upon binding to DNA, it induces a bend of 13 degree.
(TRANSFAC® class description C0023)

Logo plot:

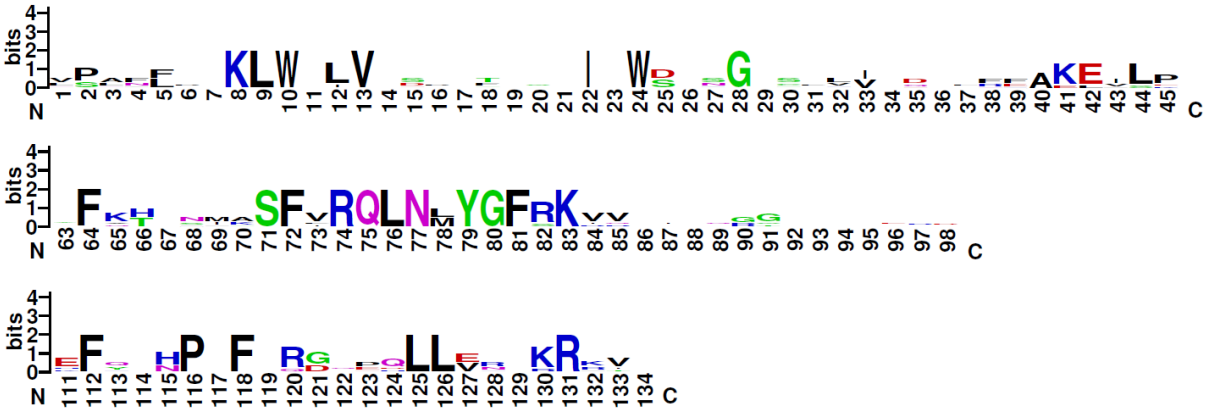
No logo plot at Class level available.

3.4 *Class* : Heat shock factors

Description:

Modular structure with an N-terminal DNA binding domain characterized by a central helix-turn-helix motif, an adjacent domain with heptad hydrophobic repeats (HR-A/B) involved in oligomerization, a cluster of basic amino acid residues essential for nuclear import (NLS), and an acidic C-terminal activation domain. (PMID 8689565, PMID 9869631)
(TRANSFAC® class description C0050)

Logo plot:



3.5 *Class* : Tryptophan cluster factors

Description:

The tryptophan clusters comprise several tryptophan residues with a spacing of 12-21 amino acid residues; the subclass of myb-type DNA-binding domains typically exhibit a spacing of 19-21 amino acid residues. All IRF family members, except IRF-1 and IRF-2, contain conserved IAD, IRF association domain that is responsible for protein-protein interactions with other IRF family members or with Ets factor PU.1 and/or bHLH factor E47. IRF-1 and IRF-2 contain IAD2 that is conserved between them and, like IAD, involved in protein-protein interactions. (PMID 10586038)
(TRANSFAC® class description C0022)

Logo plot:

No logo plot at Class level available.

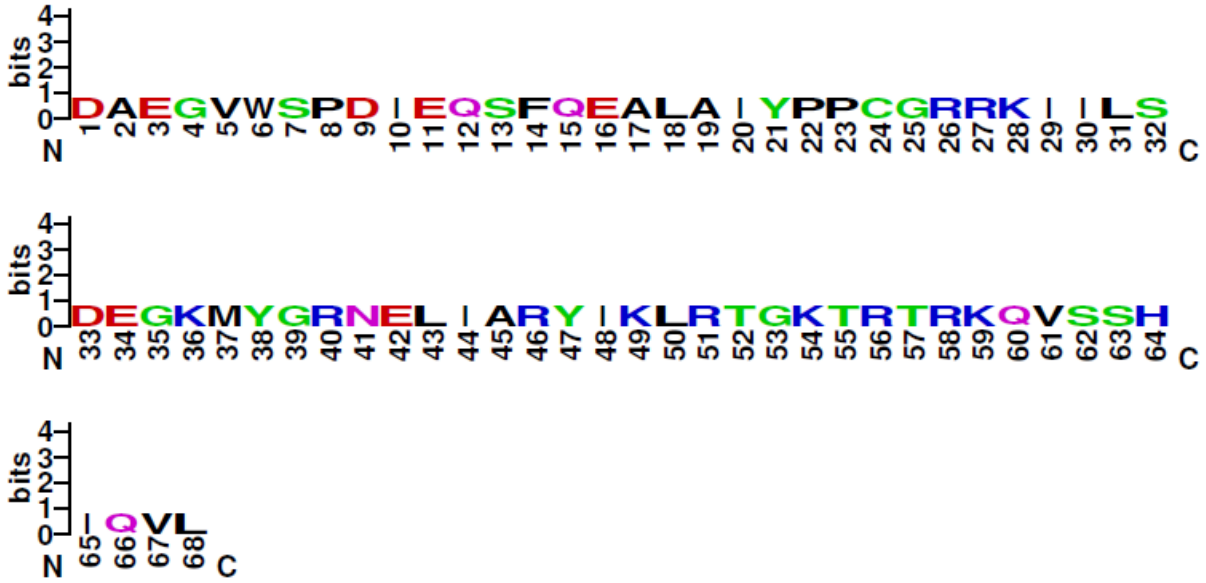
3.6 *Class* : TEA domain factors

Description:

The TEA domain has been identified as a region which is conserved among the transcription factors TEF-1, TEC1 and abaA. This domain in TEF-1 has been shown to interact with DNA, although two additional regions may also contribute to DNA-binding. It is predicted to fold into three alpha-helices, with a randomly coiled region of 16-18 amino acid residues between helices 1 and 2, and a short stretch between helices 2 and 3 of 3-8 residues.
(TRANSFAC® class description C0024)

© geneXplain GmbH, 2019

Logo plot:



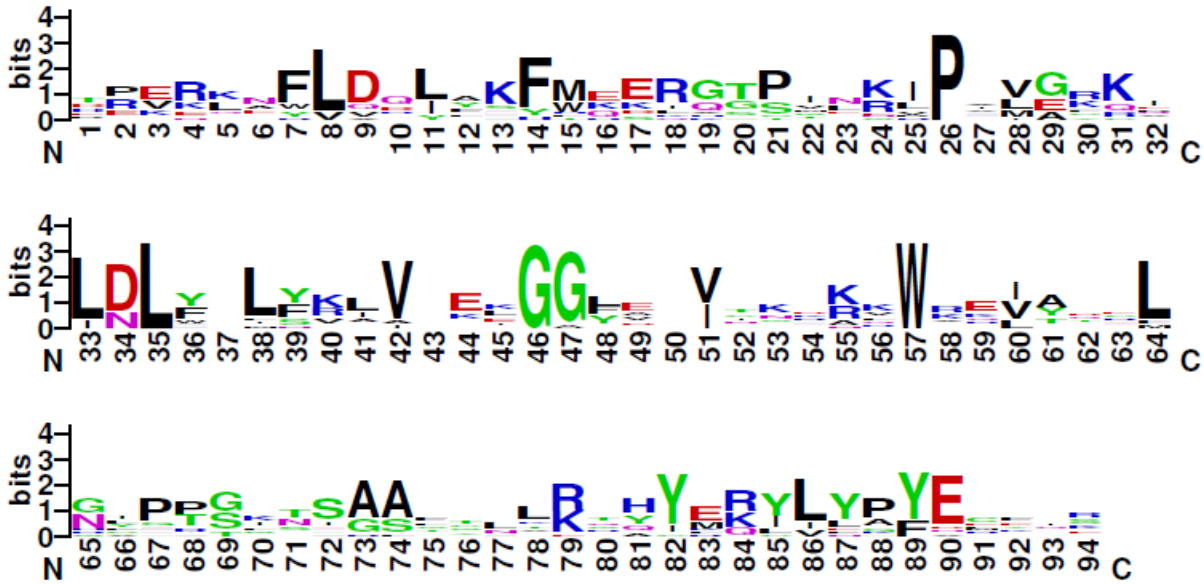
3.7 *Class* : ARID domain factors

Description:

The AT-rich interaction domain ARID comprises a variant helix-turn-helix motif, the DNA-binding of which is accompanied by additional DNA-contacting loop structures and requires further contacts for sequence-specific binding. Among the seven mammalian subfamilies, only ARID3 and ARID5 were reported earlier to bind DNA in a sequence-specific manner (PMID 15922553 and references therein). Later on, a preferred binding sequence was also reported for JARID1A (PMID 18270511), and since there may be also subtle affinity differences for the other ARID factors, and all of them seem to have an impact on transcriptional control, all known factors have been listed here. They have been classified under superclass 3 since the recognition helix of the ARID domain inserts into the major groove in a similar manner as the 'classical' helix-turn-helix motif.

Highly conserved DNA-binding domain that encompass 108-132 amino acids. The solution structure of the Mrf-2 binding domain contains six helices, two loops and a flexible C-terminus. In the DNA complex Mrf-2 contacts both the major and minor grooves.
(TRANSFAC® class description C0048)

Logo plot:



© geneXplain GmbH, 2019

4 *Superclass*: Other all-α-helical DNA-binding domains

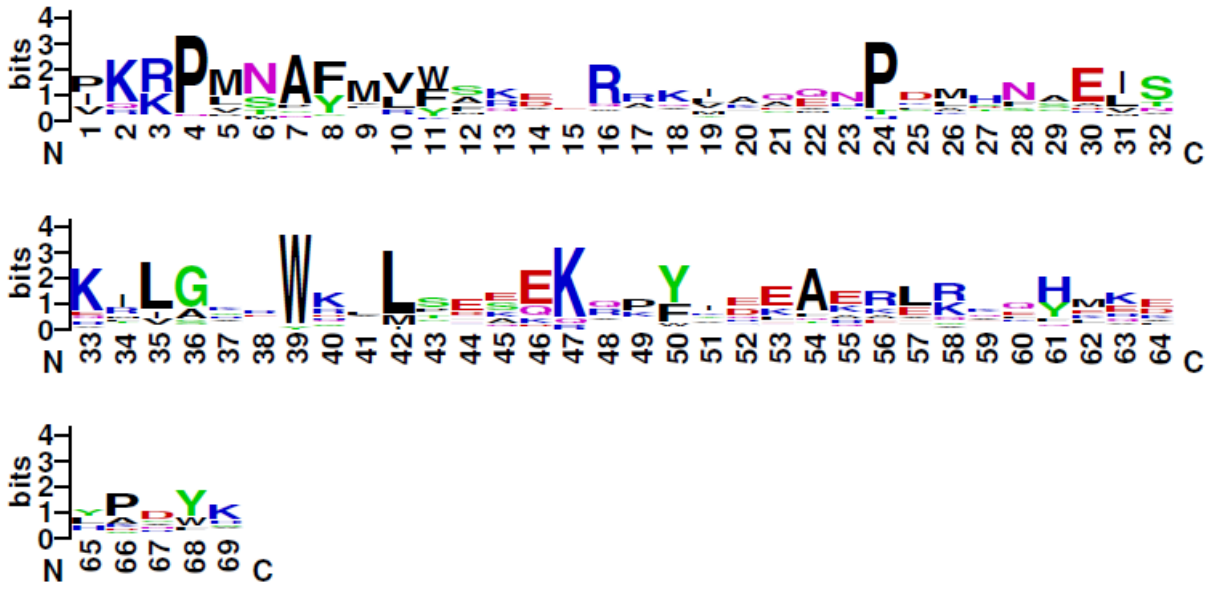
This superclass comprises transcription factors with DNA-binding domains that exhibit alpha-helically structured interfaces interacting with the DNA. So far, only the HMG proteins are structurally well characterized, whereas the DNA-binding mode of NF-Y proteins is mainly based on modeling by homology. The proteins of this group show significant similarities to some non-specific DNA-binding proteins.

4.1 *Class*: High-mobility group (HMG) domain factors

Description:

Proteins of this class comprise a region of homology with the chromosomal non-histone HMG proteins such as HMG1. This region comprises the DNA-binding domain which in some instances such as HMG1 mediates sequence-unspecific, in other cases such LEF-1 sequence-specific binding to DNA. This domain exhibits a typical L-shaped conformation made up of 3 alpha-helices and an extended N-terminal extension of the first helix. The latter together with helix 1, which contains a kink, form the long arm of the L, whereas helices 1 and 2 form the short arm. Binding to the minor groove induces a sharp bending of the DNA by more than 90 degree, away from the bound protein. The overall topology of the DNA-protein complexes resembles somewhat that of the TBP-TATA box complex.
(TRANSFAC® class description C0015)

Logo plot:



4.2 *Class* : Heteromeric CCAAT-binding factors

Description:

The heteromeric CCAAT-binding factor known from mammalian systems is NF-Y (CBF, CP1), a homolog of yeast HAP2/3/4. Like the latter, it is a heterotrimeric complex of which all three subunits NF-YA, NF-YB and NF-YC are required for binding to the CCAAT-element on the DNA. NF-YB and NF-YC both exhibit a histone-fold domain, showing some homology with histones H2B and H2A, respectively. These regions are required for heterodimerization, which is followed by docking of NF-YA. Modeling of the NF-Y complex and its binding to DNA according to the homology with histone and NC2 complexes suggests that the heterotrimeric complex provides an alpha-helically folded interface to interact with the DNA. Contribution of some loop structures is also conceivable. Upon binding through major and minor groove contacts, the NF-Y complex induces considerable bending of the DNA.

Logo plot:

No logo plot available.

© geneXplain GmbH, 2019

5 *Superclass*: α-Helices exposed by β-structures

Members of this superclass possess a DNA-contact interface of alpha-helices, which are exposed by a scaffold of beta-strands. Different to members of superclasses 1-4, the DNA-binding alpha helices do not insert in either groove but rather are packed against the DNA.

5.1 *Class*: MADS box factors

Description:

Proteins of this class comprise a region of homology. The DNA-binding domain also comprises the dimerization capability. In the DNA-bound dimer (shown for SRF), two antiparallel amphipathic alpha-helices (alpha-I), form a coiled coil and are oriented approximately parallel on the minor groove. These helices make minor and major groove contacts, the N-terminal extensions form minor groove contacts. The bound DNA is bent and wrapped around the protein. It exhibits a compressed minor groove in the center and widened minor groove in the flanks. (PMID 7637780) (TRANSFAC® class description C0014)

Logo plot:

5.2 *Class*: E2-related factors

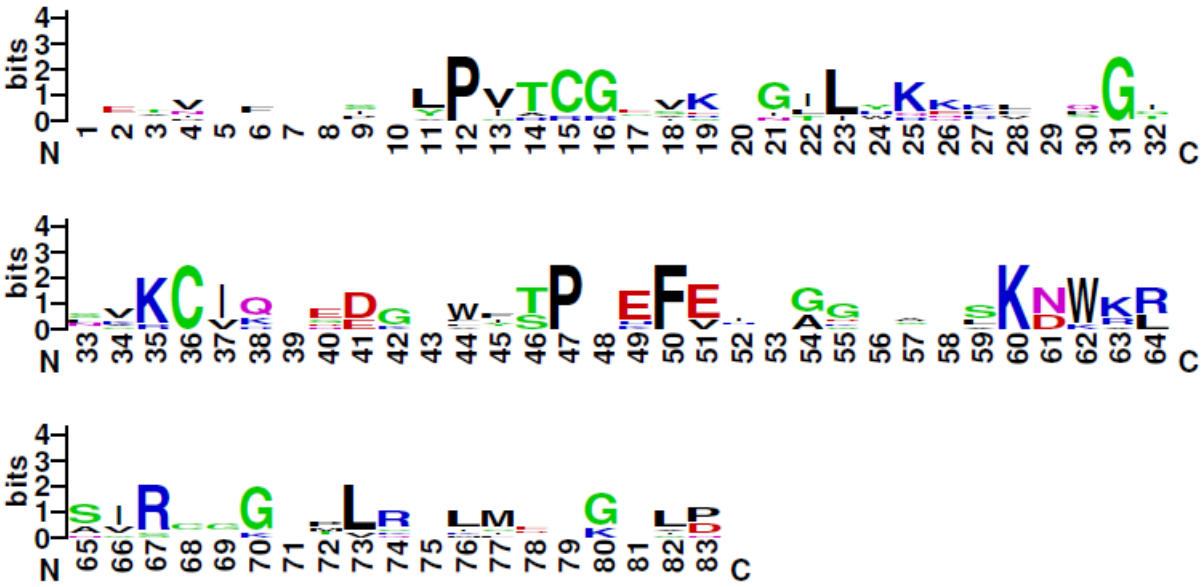
No human instances yet.

5.3 *Class* : SAND domain factors

Description:

A twisted beta-sheet exposes four alpha-helices that are packed against one side of the sheet. Most important for DNA-binding is a KDWK motif at the N-terminus of the third alpha-helix. Residues of this alpha-helix and the KWDK motif contact bases through the major groove, but the alpha-helix does not insert into the major groove. The domain structure of some of the class members is supported by an overlaying zinc coordinating motif. (PMID 11427895, PMID 15649886)

Logo plot:



© geneXplain GmbH, 2019

6 *Superclass*: Immunoglobulin fold

DNA-binding domains of this superclass exhibit an immunoglobulin-like fold. It is characterized by a beta-core structure, usually a beta-sandwich, which exposes a DNA-contact interface of mostly loops, but also other secondary structure elements, from which the DNA-binding amino acid residues are protruding.

6.1 *Class*: Rel homology region (RHR) factors

Description:

The structure of the Rel-type DBD exhibits a bipartite subdomain structure, each subdomain comprising a beta-barrel with five loops that form an extensive contact surface to the major groove of the DNA. Particularly, the first loop of the N-terminal subdomain (the highly conserved recognition loop) performs contacts with the recognition element on the DNA, but other loops are involved. The fact that the main DNA-contacts are made through loops has been suggested to provide a high degree of flexibility in binding to a range of different target sequences. Augmenting interactions are achieved by two alpha-helices within the N-terminal part that form strong minor groove contacts to the A/T-rich center of the B-element. In p65, the sequence between both alpha-helices is much shorter and even helix 2 is truncated. The second, C-terminal domain is necessary mainly for protein dimerization.
(TRANSFAC® class description C00120)

Logo plot:

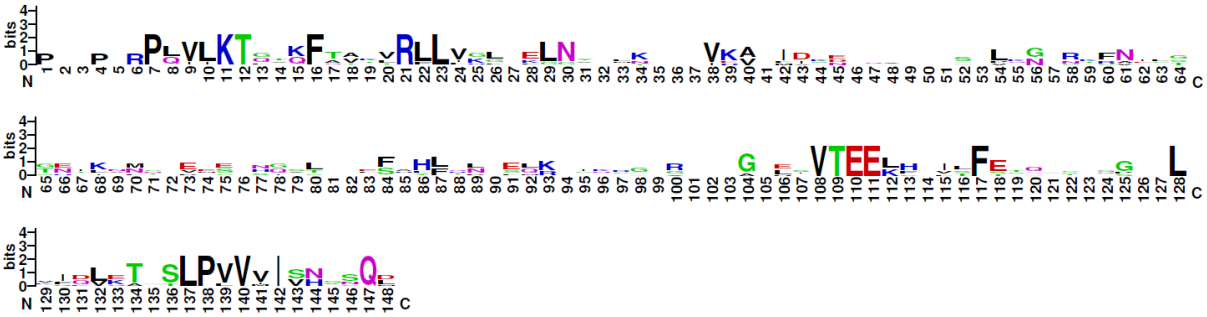
No logo plot at Class level available.

6.2 *Class*: STAT domain factors

Description:

STAT proteins bind to DNA as dimers. The DNA-contacting interface is organized by an eight-stranded beta-barrel, N-terminally preceded by a four-helix bundle and C-terminally followed by a mostly alpha-helical connector region. The residues that bind to the major groove of the DNA are mostly exposed by loops connecting the beta-strands of the beta-barrel and the one linking the beta-barrel and the first helix of the 'connector' region. The STAT dimer nearly completely embraces the DNA double helix, compared with a 'pair of pliers' (PMID 9671298), with the DNA-binding interface as jaws and the four-helix bundles as handles. Bound by STAT, the DNA undergoes a moderate bending of about 40 degrees. (PMID 9671298)

Logo plot:



6.3 *Class* : p53 domain factors

Description:

A beta sandwich constitutes a scaffold that exposes several loops and a loop-sheet-helix motif. One of the loops forms a contact of an arginine residue in the minor, side chains of the loop-sheet-helix motif in the major groove of the DNA. (PMID 8023157, PMID 21464285)

Logo plot:

No logo plot available.

6.4 *Class* : Runt domain factors

Description:

The members of this transcription factor class have been identified on the basis of their homology to a defined region within the Drosophila protein Runt. The runt domain (of RUNX1) contains twelve beta strands, seven of which shape the fold of an immunoglobulin-like beta sandwich (S-type Ig fold) also found in the DNA-binding domains of other transcription factors (NF-kappaB, NFAT, p53, STAT, and the T-domain) that is preceded by an N-terminal alpha helix (PMID 10404214, PMID 10545320, PMID 12217689). The runt domain is responsible for both DNA-binding and heterodimerization which occur at distinct, non-overlapping sites within the domain (PMID 11327761). Heterodimerization with CBFbeta leads to conformational changes in the runt domain (S-switch) that stabilize DNA binding. (PMID 10734228, PMID 11257229, PMID 12217689) (TRANSFAC® class description C0029)

Logo plot:

No logo plot available.

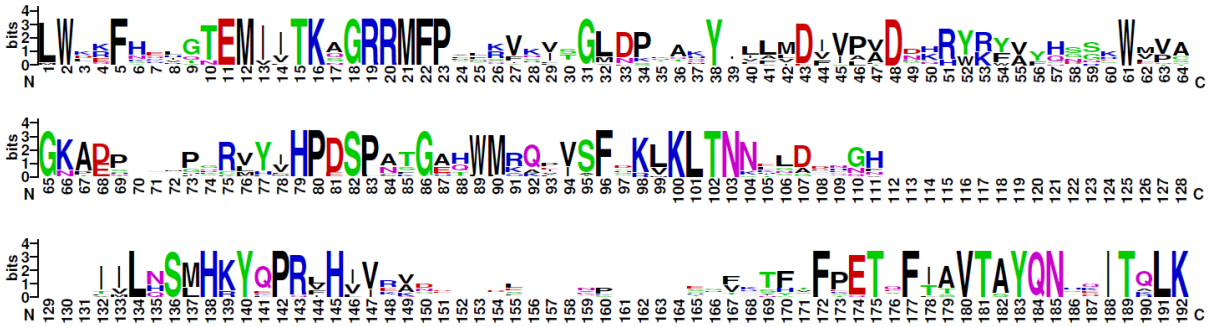
© geneXplain GmbH, 2019

6.5 *Class* : T-Box factors

Description:

The T domain has originally been discovered in Brachyury (T) the founder member of the T-box family. It encodes a DNA binding domain of about 180 amino acids. X-ray analysis of the T domain showed that it binds as a dimer to a 20 nucleotide partially palindromic sequence. A new type of specific DNA contact is seen, in which a carboxy-terminal helix is deeply embedded into an enlarged minor groove without bending the DNA.
(TRANSFAC® class description C0044)

Logo plot:



6.6 *Class* : NDT80 domain factors

Description:

A beta sandwich comprising a three-standed sheet packed against a four-stranded sheet constitutes a scaffold that exposes the DNA-contacting amino acid residues of six distinct regions. Most of these residues are located in loops, some are part of short alpha-helices or beta-strands. They form both minor and major groove contacts to the DNA. (PMID 12411490)

Logo plot:

No logo plot available.

6.7 *Class* : Grainyhead domain factors

Description:

Ancient family of factors characterized by a distinct type of DNA binding domain. Members of this class are known in C. elegans, Drosophila, and vertebrates, and involved in the regulation of

diverse biological processes (PMID 12888489 and references therein). Majority of the factors are activators, but LBP-9 and CRTR-1 are repressors (PMID 10644752, PMID 11073954). On the basis of sequence similarity, class can be subdivided into the two families: LSF/CP2 and Grainyhead itself (PMID 12888489). Members of this class are characterized by the conserved DNA binding- and oligomerization-associated regions (PMID 12888489). Consensus DNA recognition element of the LSF/CP2 family members represents a direct bipartite repeat CNRG-N(6)-CNRG (PMID 9668115). LSF/CP2 factors bind DNA as homotetramers, Drosophila Elf-1 binds as a dimer (PMID 8349681). (TRANSFAC® class description C0051)

Logo plot:

No logo plot available.

© geneXplain GmbH, 2019

7 *Superclass*: β-Hairpin exposed by an α/β-scaffold

The DNA-binding domains of this superclass have an alpha/beta-structured scaffold that exposes a beta-hairpin. This hairpin is the main DNA-contacting element and inserts into the major groove of the DNA. (PMID 12682016, PMID 20147459)

7.1 *Class*: SMAD/NF-1 DNA-binding domain factors

Description:

Significant similarities between DNA binding domains of Smad and NF1 factors have been revealed by a PSI-BLAST database search. Authors have proposed that Smad and NF1 factors belong to the same group of genes probably resulting from the common ancestry [25470]. For more details about Smad or NF1 factors see and correspondingly. TRANSFAC class description C0041: Proteins with 2 MAD-homology domains at N- and C-terminal ends, which have been named MH1 and MH2, are independently folded structures (PMID 10605817); the structure of the MH2 domain is a beta-sandwich with a 3-alpha-helix bundle at one end, and 3 large loops and an alpha-helix at the other end (PMID 10605817); the loop-helix region of one subunit packs against the 3-alpha-helix bundle of another subunit (PMID 10605817); MH1, a novel type of DBD, consists of 4 alpha-helixes, 6 short beta-strands, and 5 loops. Beta-strands form 2 small beta-sheets and 1 beta-hairpin (PMID 9741623); base-specific DNA recognition of the MH1 domain bound to a GTCT motif is via an 11-amino acid beta-hairpin, which lies in the major groove of the DNA (PMID 9741623, PMID 10605817); 3 classes of SMAD: Inhibitory SMADs (SMAD-6,SMAD-7), Receptor-regulated SMADs (SMAD-1, SMAD-2, SMAD-3, SMAD-5, SMAD-8), Common Mediator SMADs (SMAD-4, SMAD-4beta, SMAD-10). (PMID 10605817)
(TRANSFAC® class description C00059)

Logo plot:

No logo plot at Class level available.

7.2 *Class*: GCM domain factors

Description:

Within the N-terminal parts of Drosophila GCM, human and mouse GCMa and murine GCMb, a region of about 150 AA was disclosed that exhibits a high degree of homology. This gcm-motif is responsible for DNA-binding. Within this motif, three regions of identity were discovered: A, 10 AA; B, 9 AA; C, 10 AA. Additionally, there are 7 conserved Cys and 4 conserved His residues scattered over the gcm-motif.
(TRANSFAC® class description C00031)

Logo plot:

No logo plot available.

© geneXplain GmbH, 2019

8 *Superclass*: β-Sheet binding to DNA

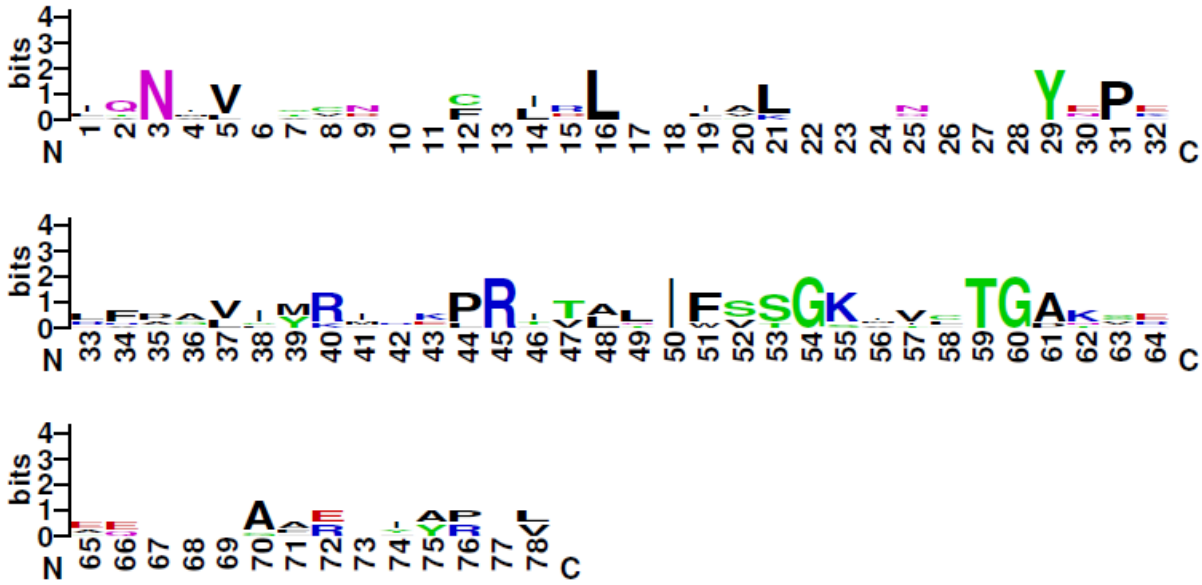
The DNA-binding domains of this superclass bind to DNA through single extended strands or beta-sheets.

8.1 *Class*: TATA-binding proteins

Description:

A 10-stranded, anti-parallel beta-sheet forms a quasi-symmetric saddle-like structure, where the convex side is backed by four alpha-helices and the concave side exerts mostly hydrophobic interactions with the minor groove of the TATA-box. Binding of TBP imposes a significant kink of the DNA double helix axis. (PMID 10617571, and references cited therein)

Logo plot:



8.2 *Class*: A.T hook factors

Description:

The A.T hook is a short sequence motif of 9-16 amino acids, occurring in multiple repeats (3 in the human HMGA factors). It is unordered when free in solution, but adopts an extended conformation upon binding to the minor groove of an AT-rich DNA sequence. In contrast to TBP, it does not hydrophobically intercalate and therefore preserves the B conformation of the bound DNA. (PMID 9253416)

Logo plot:

© geneXplain GmbH, 2019

9 *Superclass*: β-Barrel DNA-binding domains

DNA-binding domains that have a beta-barrel fold and are not covered by superclasses 7-8 are compiled in this superclass.

9.1 *Class*: TATA-binding proteins

Description:

Cold-shock domain proteins are characterized by a highly conserved region first found in prokaryotic cold-shock proteins (PMID 1622927, PMID 8031301). This domain is a single-stranded nucleic acid-binding structure interacting with DNA or RNA. It consists of an antiparallel five-stranded beta-barrel, the strands of which are connected by turns and loops (PMID 8321288, PMID 8321289). Within this structure, a three-stranded beta-strand contains a conserved RNA-binding motif, RNP1 (PMID 8321288, PMID 8321289, PMID 8967899). Not all CSD proteins are transcription factors. Those which specifically bind to a certain sequence are termed Y-box proteins. - Proteins of this class were previously called protamine-like domain proteins because of having a highly positively charged domain with interspersed proline residues.
(TRANSFAC® class description C0019)

Logo plot:

No logo plot available.

0 *Superclass*: Yet undefined DNA-binding domains

Factors of this superclass have been shown to bind to DNA in a sequence-specific manner, but their DNA-binding domain has not yet been identified and/or structurally characterized.

0.1 *Class*: AXUD/CSRNP domain factors

Description:

Factors of this family possess a conserved domain comprising a serine-rich, a basic and a cysteine-rich region. The mouse ortholog of CSRNAP-1 binds to AGAGTG. (PMID 17726538)

0.2 *Class*: NonO domain factors

Description:

A group of multifunctional factors, probably involved in transcriptional as well as in splicing regulation. Transcriptional regulation is most likely through direct DNA-binding. All factors of this class identified so far exhibit a highly charge region of 124 amino acid residues N-terminally of two RRMs (RNA-recognition motifs).

0.3 *Class*: Leucine-rich repeat flightless-interacting proteins

Description:

A 100-aa DNA-binding domain has been identified in LRRFIP1; it shows only limited homology with the corresponding region in LRRFIP2, for which DNA-binding has not yet been shown. In either case, this region contains many charged side chains.

0.4 *Class*: NFX1-type putative zinc finger factors

Description:

A large class of proteins, with most known members from the plant kingdom. The DNA-binding domain comprises a cysteine-rich region with a repeated motif of the signature C-x(1-6)-H-x-C-x3-C-h-x(3-4)-C-x(1-10)-C (with the lower-case histidine occasionally replaced by another cysteine). Although this motif has been denominated 'NFX1-type zinc finger', neither zinc ion binding nor its 3D structure have been unravelled so far. (PMID 7964459, PMID 20522174)

© geneXplain GmbH, 2019

0.5 *Class* : GTF2I domain factors

Description:

Factors of this class exhibit several DNA-binding repeats that do not show similarities with other known DNA-binding domains. Several repeats of one protein differ in their DNA-binding affinity. Their 3D structure comprises 3 alpha-helices and two short beta-folded strands, one between helix 2 and 3, one C-terminal of helix 3. (PMID 15987678, PDB 2D9B, 2DN4, 2ED2, 2EJE)

0.6 *Class* : CG-1 domain factors

Description:

The CG-1 domain has been delineated as a DNA-binding region and was named after its DNA-binding specificty for CGCG motifs. It is found in transcription factors of a wide range of species, including human and plants. The 3D structure is not yet known. (PMID 11925432)

0.0 *Class* : Uncharacterized

Description:

No DNA-binding domain has been identified in these factors, although there is evidence that they may be (DNA-binding) transcription factors.