

The Importance of Being Flexible: The Case of Basic Region Leucine Zipper Transcriptional Regulators

Maria Miller*

Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD, 21702-1201, USA

Abstract: Large volumes of protein sequence and structure data acquired by proteomic studies led to the development of computational bioinformatic techniques that made possible the functional annotation and structural characterization of proteins based on their primary structure. It has become evident from genome-wide analyses that many proteins in eukaryotic cells are either completely disordered or contain long unstructured regions that are crucial for their biological functions. The content of disorder increases with evolution indicating a possibly important role of disorder in the regulation of cellular systems. Transcription factors are no exception and several proteins of this class have recently been characterized as premolten/molten globules. Yet, mammalian cells rely on these proteins to control expression of their 30,000 or so genes. Basic region:leucine zipper (bZIP) DNA-binding proteins constitute a major class of eukaryotic transcriptional regulators. This review discusses how conformational flexibility “built” into the amino acid sequence allows bZIP proteins to interact with a large number of diverse molecular partners and to accomplish their manifold cellular tasks in a strictly regulated and coordinated manner.

Keywords: bZIP proteins, transcription control, intrinsic disorder, molecular recognition motifs, induced protein folding.

INTRODUCTION

Sequencing of the whole genomes from diverse specimens together with the application of automated high-throughput molecular biology techniques has allowed an accumulation of a vast systematic data in proteomics, interactomics, transcriptomics, and several other “omics”. The central issue of the ongoing efforts is the identification of function of genes and of the proteins they encode. These studies became possible thanks to powerful techniques for studying protein structure and protein–protein interactions, mapping regulatory elements within gene promoters utilizing gene chips and chromatin precipitation methods, and the development of bioinformatics. This wealth of data, combined with tools for gene-based analysis opened new avenues for systematic discoveries of functionally important features, posed new questions and challenges. The challenge of understanding the organization and dynamics of cellular networks in time and space has shifted the focus of interest from molecular biology toward systems biology. In the case of gene regulatory networks, attempts are being made to delineate all transcription factors (TFs), chromatin regulators and signaling pathways necessary to activate genes in specific tissues and cell types, and to determine when specific gene expression occurs and how it is regulated.

Intrinsically Disordered Proteins

Concomitantly with the widespread employment of genetic methods to identify proteins with specific functions, the

application of improved spectroscopic techniques (mainly CD and NMR) to examine macromolecular structures in solution revealed that many proteins lack a well-ordered tertiary structure under physiological conditions. It has been recognized that such proteins (or regions), termed “natively unfolded”, “intrinsically unstructured” or “intrinsically disordered” can exist as dynamic ensembles of interconverting conformers in three disordered structural states: fully unfolded extended random coils, premolten globules that retain some amount of residual secondary structure, and compact but disordered molten globule-like ensembles [1,2]. Disordered regions can be inserted within the single structural domain, usually in the form of long surface loops [3], or serve as flexible linkers between independently folded globular domains. Many examples of such proteins have been described in an excellent review by Dyson and Wright [4]. Surveys of whole genomes showed that the occurrence of long (> 50 residues) unstructured regions is very common in functional eukaryotic proteins [5]. This observation changed the general view on the native structure and its relation to protein function. As accumulated evidence showed the importance of intrinsic disorder (ID) for protein function, in a landmark paper, Wright and Dyson called for reassessing the classic protein structure–function paradigm to include all possible native conformations [6]. Function may arise from any of the conformational states or interconversion between states [5].

In a recent paper, Radivojac *et al.* [7] proposed to classify a protein as intrinsically disordered if it contains at least one disordered region whose length is sufficient for experimental characterization. Intrinsically disordered proteins (IDPs) have been studied by several methods: NMR (reviewed in [8]), fluorescence, CD and Raman spectroscopy, hydrodynamic and calorimetric methods, and small-angle X-

*Address correspondence to this author at the Macromolecular Crystallography Laboratory; National Cancer Institute at Frederick, Frederick, MD, 21702-1201, USA; Tel: (301) 846-5342 ; Fax: (301) 846-6128; E-mail: millerm@ncifcrf.gov

ray scattering, as well as other biophysical techniques [9]. Proteins and protein domains experimentally confirmed to be unstructured fall into several functional categories including effectors, scavengers, assemblers and entropy chains [5,10], and have been implicated in the regulation of cell cycle control, transcription and signal transduction [11]. IDPs were subjects of extensive recent studies, focused reviews [2,4,7,10-15] and commentaries [1,5,6].

Intrinsic Disorder Can be Predicted from the Amino Acid Sequence

The large volume of accumulated sequence data made possible the development of computational techniques that utilize the information contained in amino acid sequences to predict structural disorder. A comparison of sequences from ordered and disordered regions of proteins revealed significant differences in their respective amino acid composition. Several artificial neural network based predictors of natively disordered regions (PONDRs) have been developed based on training of the selected sequence feature (Ref. [2] and references therein). The work of Vucetic and colleagues [16] indicated that the amino acid sequence underlying ID also specifies distinct properties of the polypeptide that correlate with its function. This observation raised the possibility of assigning IDPs to functionally specific classes (flavors) based on their amino acid composition alone. Implementation of a machine learning method for predicting protein function from sequence indeed demonstrated that the inclusion of disordered features improved prediction accuracy for certain functional categories, and made possible the annotation of several orphan human proteins [17]. The manifestations of disorder at the primary structure level include a low sequence complexity, in particular a low content of bulky, hydrophobic residues combined with a high proportion of charged/polar residues, resulting in a large net charge of the protein at neutral pH. Uversky *et al.* [18] showed that it is possible to predict whether a given amino acid sequence encodes a folded or a natively unfolded protein based on a charge-hydrophobicity plot. This principle has been employed in development of the protein disorder predictor server FoldIndex [19], to identify local disordered regions on a per residue basis. Other predictors based on specific sequence attributes are PreLink [20] (low hydrophobic cluster content), NORSp [21] (solvent-accessible regions devoid of secondary structure), GlobPlot [22], which uses a scale for propensities for secondary structures and random coils, and based on sequence profiles, DISOPRED2 program. A different approach is employed by the IUPred algorithm, which evaluates the energy resulting from interresidue interactions and has not been trained against a dataset of disordered protein regions. For reviews of the currently available disorder predictors, see [23,24]. Each of these predictors suffers from different limitations and performs better for one type of disorder than for another (e.g., DISOPRED2 and PreLink are able to predict short disordered regions in the context of globally ordered protein, whereas the charge-hydrophobicity method usually indicates unstructured random coils). Therefore, a reliable prediction of protein disorder has to combine outputs from several predictors based on different physical principles and/or definitions of disorder [23-25]. The current estimate of accuracy for the *ab initio* methods such as PONDR is about 70%

[23]. It is anticipated that further improvement of computational techniques for sequence-based ID prediction could increase the level of accuracy to its upper limit of 85%–90% [7].

Application of these techniques in the whole genomes-wide studies revealed an abundance of proteins containing ID regions, their evolutionary conservation and increasing content in higher organisms, suggesting the importance of protein disorder for regulation of cellular processes [1,17,26-28]. ID is prevalent among proteins involved in cell signaling and DNA/RNA recognition in eukaryotic cells [15,29]. Recently, Romero *et al.* [30] showed that the majority of alternative spliced peptide fragments were associated with disordered protein regions, and the researchers postulated a relationship between increased ID content and alternative splicing phenomena in multicellular eukaryotes. By affecting mainly natively unstructured regions, alternative splicing may increase functional diversity of the proteome without disrupting the structural integrity of protein components.

The Role of Intrinsic Disorder in Protein Function

The dynamic flexibility associated with the intrinsic lack of structure may confer multiple benefits for protein function, particularly for biomolecular recognition and organization of multicomponent cellular networks. Unstructured regions enable proteins to contact their ligands over a large binding surface [31] and permit binding to multiple targets. Structural plasticity allows a protein to adopt more than one conformation, depending upon the ligand, and to recognize diverse targets (reviewed in [4]).

What are the advantages for molecular recognition between flexible macromolecules compared to rigid systems? In both systems the bimolecular reaction requires diffusional formation of an encounter pair that further evolves into a stable complex in a number of steps. According to the early model of “conformational selection”, a multitude of possibilities for conformational searches offered by the presence of flexible regions in interacting protein(s) result in much better selectivity compared to stably folded molecules. It has been anticipated that binding which involves many isomerization steps is much more efficient in terms of rate and specificity than the “lock-and-key” type of recognition with a large number of unsuccessful fittings. Furthermore, specific biological function often requires significant conformational reorganization of the interacting proteins. The best example is the regulation of enzymatic activity by the allosteric effect. Small adjustments to the conformation were successfully described by the “induced fit” model. However, there are very many examples of proteins that are partially or completely disordered in the unbound state and undergo function-related disorder–order transitions upon binding to their respective partners [12,18,32]. The most relevant explanation for this phenomenon comes from the requirement for highly specific but weak, readily reversible interactions in very many biological processes. The importance of transient interactions between both proteins and nucleic acids is most evident in transcription, wherein transcription regulatory proteins exchange contacts with a variety of cofactors (see below). In contrast to interactions between rigid molecules, in which stability of the complex usually correlates

with specificity, in the binding associated with large structural changes, high specificity (the best fit) is achieved through conformational adjustments, often at the expense of binding affinity. The molecular mechanisms for target-assisted protein folding and the role of flexibility in determining the binding specificity have recently been addressed by a number of experimental and theoretical studies [12,33-39].

Coupling of folding and binding facilitates control and regulation of the binding thermodynamics (reviewed in [4,40]) and increased rates of macromolecular associations [41]. These abilities can be attributed to the large solvated surface area of the polypeptide chain in an extended conformation. It is well documented that the burial of hydrophobic surfaces and the consequent release of hydration waters to the bulk solvent can significantly contribute to the thermodynamics of binding [42]. The ordering of structure, which leads to a decrease in protein surface area and exclusion of a large amount of solvent, will enhance this effect. It has been shown that higher levels of structural ordering are present in specific complexes, when compared to nonspecific complexes [43,44]. Protein folding and protein-ligand interactions often involve entropy-enthalpy compensation mechanism [45]. The entropy change that modulates the free energy of binding depends mainly on the relative contributions from two opposing components, one being solvent release, which increases entropy, and the other being change of conformational entropy. Thus the loss of entropy associated with folding of a flexible domain, which results in the reduction of binding affinity, can be to some extent compensated by a significant entropy gain that arises from ordering of structure upon *specific* binding. The mechanisms by which folding is coupled to binding are not well understood. It has been suggested that the interplay between structural ordering and desolvation may allow the action of multiple mechanisms of sequential and selective enhancement of interactions [40].

Flexibility in molecular recognition has been analyzed by molecular dynamics simulations initiated from crystal structure of ternary complexes [34], and by simulating the association of various monomers into higher oligomeric complexes [36]. Several models of folding-associated binding were proposed based on the energy landscape theory of protein folding [35]. The process of folding (binding) occurs via searches for energetically favorable intra- (inter-) molecular interactions on an energy landscape shaped like a funnel, with a narrow lower part representing the folded (bound) state. The reaction proceeds downhill along the free energy gradient. Wang and colleagues [35,38] derived the thermodynamic free energy expression for the flexible binding. They showed that inherent hydrophobic interactions and cooperativity between folding and binding lead to dynamic fluctuations between the “non-native” partially folded and closely bound conformational states. To avoid trapping within local energy minima, the binding energy landscape has to be biased (funneled) toward the native binding state. This implies certain distribution of protein apolar residues between the hydrophobic core (for the purpose of folding stability) and surface (for the purpose of binding). The balance between binding affinity and specificity depends on the ratio of the binding transition temperature versus the trap-

ping transition temperature. This criterion implies loss of some affinity to enable the molecules to reach the best fit.

Yet another advantage of coupling binding to folding may be the increased rate of complex formation. The large surface area of IDPs that is available for interactions provides a greater capture radius, compared to a compact globular protein. Using the same methodology, based on analogy between folding and binding, Shoemaker *et al.* [41] proposed a “fly-casting” mechanism for binding interactions of IDPs. According to this model, the unstructured polypeptide binds weakly over a long distance and then “reels in” the target during the protein folding process [41].

A recently published theory implies that protein disorder optimizes allosteric coupling via ensemble-mediated mechanism [46]. The presence of ID in segments containing one or both of the coupled binding sites generates an ensemble of states that is “optimally poised to respond” to binding. Upon binding to the ligand, the ensemble of states is redistributed. Significantly, such a mechanism depends only on the relative stabilities of the domains and not on the specific structural basis of that stability and can be related to experimentally obtainable values such as stability and binding affinity.

Direct experimental evidence for the coupled binding and folding was provided by Sugase *et al.* [33], who demonstrated that the binding of the phosphorylated kinase inducible activation domain of CREB (pKID) to the KIX domain of CBP involves the formation of an ensemble of unstructured encounter complexes, which are stabilized by non-specific hydrophobic contacts as well as the formation of partially folded intermediate states. Using NMR titration and ¹⁵N relaxation dispersion techniques, the researchers characterized kinetics of binding process, the populations and structure of intermediate states, and showed that the formation of the final fully folded complex with the structure of pKID stabilized by intermolecular interactions occurs without dissociation from KIX.

Interaction-prone regions with the potential to fold upon complexation have been named “molecular recognition features” (MoRFs) [32,47]. The flexibility of binding by IDPs depends on the inherent conformational preferences of disordered regions. In some cases the residual structure of those peptides is not completely random but exhibits local predominance of a particular secondary structure, usually helical, that is later stabilized in the bound state (inherent-structure mechanism). It has been proposed that such nascent, transiently formed structures serve as initial inter-chain contact points and contribute to the reduction of entropic penalty and increased affinity of binding. In an alternative mode of binding (induced-structure mechanism), the intrinsic conformational preferences of disordered regions are suppressed by interactions with the partner. Thus disordered chains with low local structural preferences may adopt multiple conformations upon binding to different partners [47].

In addition to MoRFs, short sequences that form “linear motifs” recognized by globular domains (e.g., phosphoSP binding to WW domains, PXXP binding to the SH3 domain), are usually located within unstructured regions [48-50]. These motifs are key mediators of crucial biological interactions. It has been suggested that specific binding oc-

curs via short primary contact site(s) and subsequent adhering of the longer flexible chain to the surface of the binding partner [51]. The extended polypeptide chain often displays several binding motifs, which enable the protein to bind simultaneously to multiple ligands or to form multivalent interactions with one.

Furthermore, IDPs often mediate formation of multi-protein assemblies as well as self-assembly of biological macromolecules [52]. These attributes make ID regions of proteins indispensable in organizing cell signaling and gene regulatory networks. Cellular processes carried out by the extensive protein interaction networks are constructed like electrical and engineering control systems, where most switches and feedback mechanisms rely on assembly and disassembly of multi-component complexes [53]. A majority of the proteins in cellular networks make only one or two connections (date hubs), while, on the other hand relatively few hubs mediate a large number of links (party hubs). Party hub proteins interact simultaneously with many partners, while date hubs interact with different partners at different times [15,54]. The ability of party hub proteins to bind multiple partners with high specificity depends on their conformational plasticity. The study of Singh *et al.* [55] revealed the significant enrichment of ID in party hub proteins, underscoring the importance of disorder for transient binding interactions. Moreover, a majority of the experimentally determined phosphorylation sites in eukaryotic proteins, as well as sites of proteolytic digestion and other modifications, are located within segments with sequence features indicative of intrinsic disorder [56,57]. Thus, ID promotes binding diversity, renders highly specific but reversible interactions, facilitates tight regulation, rapid turnover, provides a kinetic advantage for macromolecular associations, and ensures quick responses to external signals [6,14].

Intrinsic Disorder and Transcriptional Regulation

Activation of transcription involves complexes of many different proteins that recruit components of the transcriptional machinery to the promoter region and also induce changes in the structure of chromatin. At the DNA level, gene expression is directed by sequence-specific DNA binding TFs, which specifically recognize and bind to *cis*-regulatory sequences in target genes and subsequently activate or repress their transcription. It was recognized early on into the studies of transcriptional activation [58,59] that critical events regulating the initiation of transcription cannot depend on specific and rigid complementary surfaces of participating proteins. The variety of arrangements and stoichiometries observed in the preinitiation complexes, the undefined position of *cis* regulatory sequences in respect to the promoter that they activate, as well as the modular nature of specific DNA-binding proteins suggested that regulation of an assembly and activity of transcriptional complexes could not be achieved if every component had an extremely high affinity or specificity for another [60]. Also, lack of sequence similarity among trans-activating domains (TADs), which seemed to consist mainly of “acid blobs” and “negative noodles” with undefined conformation supported this conclusion [59]. Subsequently, many TADs were characterized as mostly unfolded when studied as separate peptide units [61]. Moreover, it has been demonstrated (using NMR

and small-angle X-ray scattering) that TADs of p53 and herpes simplex virus VP16 remain unstructured in the full-length protein context [62,63].

Although the principles of gene regulation are the same in case of prokaryotes and eukaryotes, the corresponding transcription regulatory proteins exhibit striking differences in their molecular architecture. Eukaryotic TFs are on average twice as long as their prokaryotic counterparts and contain extended regions of protein disorder [29,64]. Recently, Liu *et al.* [64] reported that at least 82% of eukaryotic TFs possess extended regions of ID as predicted by the PONDR software and charge-hydrophobicity plots. In a parallel study, Minezaki *et al.* [29] compared the content of disorder (assessed by the DISOPRED2 program) in transcriptional activator proteins from different species. They found that up to 49% of the entire sequence of human TF proteins is occupied by ID regions. This is in contrast to their prokaryotic counterparts where long stretches of unstructured region are rare, and for which several 3-dimensional structures of full-length proteins are available in the Protein Data Bank (PDB). This discrepancy probably reflects differences in the complexity of transcriptional apparatus, chromatin structure, and regulatory pathways used to regulate gene transcription in eukaryotes and prokaryotes [29].

A salient difference between transcriptional regulation in the two kingdoms arises from the densely packed structure of chromatin in eukaryotic cells, which restricts access of the basal transcriptional machinery to the promoter region [65]. Furthermore, in contrast to prokaryotic gene regulation whereby transcription can be modulated by a single monomeric protein, transcriptional activation in eukaryotes is generally controlled by the concerted action of many signal-regulated TFs and cofactors, which form multiprotein–DNA complexes and modify chromatin structure in the promoter and enhancer regions. This combinatorial regulation is organized in a highly hierarchical manner and allow for tightly controlled integration of independent signaling pathways in biological responses [66,67]. Clearly, the complexity of the nuclear transcriptional machinery required evolution of more sophisticated regulatory molecules that would be capable to handle tasks of overcoming the repressive effects of chromatin and suppressor proteins, as well as the competitive recruitment of coactivators.

This review focuses on the role of structural disorder in the function of the basic region:leucine zipper (bZIP) transcriptional regulators. bZIP TF proteins constitute the largest and the most conserved [68] superfamily of TF proteins that operate exclusively in eukaryotes. bZIP TFs are involved in vital cellular functions and regulate development, metabolism and responses to environmental changes [69]. This functional versatility arises from their ability to interact with each other and with structurally unrelated TF proteins [70].

BZIP BASICS

Members of the bZIP superfamily bind to target DNA sites as homodimers or heterodimers and recognize related but distinct palindromic sequences (Fig. 1A). The prototypical bZIP protein, CCAAT/enhancer binding protein α (C/EBP α), was discovered over 20 years ago and was one of the first mammalian transcription factors to be purified and

cloned [71]. These studies established that the bZIPs DNA-binding domain (DBD) consists of a positively charged segment, the basic region (BR), linked to a sequence of heptad repeats of leucine residues called the leucine zipper (LZ) [72]. Soon after, many proteins were categorized in this class, including proto-oncogenes Fos and Jun. As shown by a number of crystal structures [73-77], bZIP peptides bind to their cognate DNA duplexes as dimers of uninterrupted α helices which form a chopsticks-like structure (Fig. 1A). Dimerization is mediated by the LZ segments, which form two parallel coiled-coil α -helices wrapped around each other. The two helices, which are positioned nearly perpendicularly to the DNA double helix, diverge smoothly toward their amino termini and each BR segment contacts one-half of a palindromic site in the DNA major groove. This is the simplest known protein–DNA recognition motif.

In contrast to many instances of stably folded DNA-binding domains (e.g. GAL4 [78], and p53 tumor suppressor protein [79]), the bipartite DBD of bZIP proteins assumes a well-ordered, stable structure only when bound to a specific DNA site [80,81]. At physiological concentrations, in the absence of DNA, unfolded monomers are in dynamic equilibrium with the folded dimers [82]. A concentration-dependent folding/unfolding transition is characterized by a fast rate of the subunit exchange. Binding to a specific DNA duplex stabilizes the coiled-coil dimer and induces helical folding of the basic region.

Based on their DNA-binding specificities, bZIP proteins have been traditionally arranged into several families, each of which recognizes a unique DNA motif (Fig. 2A). The structural basis of specific DNA recognition has been discussed in several review articles [75,83] and will not be addressed here in detail. Briefly, bZIP TFs recognize their cognate DNA sites through base contacts made by five residues within the basic region motif characteristic for each family. These five positions are highly conserved among bZIP transcription factors and contain invariant Asn and Arg residues. Crystal structures of bZIP domains bound to their cognate DNA duplexes have revealed functional variability of the conserved residues in DNA recognition (Fig. 2B). For that reason, no universal code relating the BR sequence motif to its preferred DNA binding-site sequence has been established; however members of the same family are believed to recognize their cognate sites in the same manner. bZIP proteins show relaxed DNA binding specificity beyond their respective core consensus sequences, and they can recognize similar sequences with lower affinity. Selected bZIP proteins can form stable DNA complexes as cross-family heterodimers [84,85]. Heterotypic dimerization generates an expanded repertoire of specific DNA sites (composed of different half-sites) recognized by bZIP factors. Alternative methods of classification therefore have been proposed based on sequence similarity and dimerization properties [85], and recently on the phylogeny [68] of their BR-LZ domains. In contrast, members of the superfamily differ widely in their

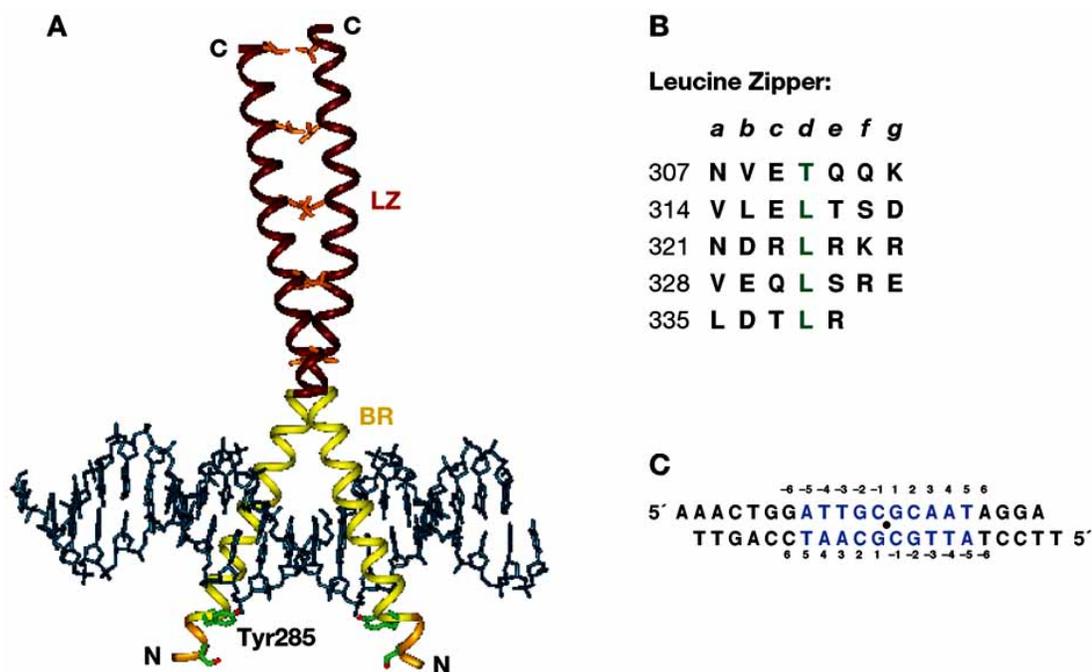


Fig. (1). (A) Overall structure of the C/EBP α bZIP–DNA complex (PDB code 1NWQ). LZ, leucine zipper; BR, basic region; EBR, extended basic region. (B) The LZ sequence displayed as heptad repeats. Numbering refers to the complete rat C/EBP α protein sequence. Lower-case letters (top line) show positions of leucine zipper residues as they would appear on a standard helical wheel representation of a coiled-coil dimer; “d” corresponds to the position of leucine residues. Coiled-coil structure is stabilized by the van der Waals interactions between residues at a and d positions from both helices and by electrostatic interactions between residues located at g and e positions. (C), The DNA sequence in crystals. Consensus C/EBP recognition site is indicated in blue, and the center of symmetry as a solid circle.

amino acid sequence outside the conserved bZIP domain. Only short conserved subregions could be identified, even in the case of proteins belonging to the same family (see Fig. (7)).

Mammalian bZIP Proteins

By 1995, 82 distinct DNA-binding bZIP proteins from a variety of eukaryotic sources had been cloned and characterized (reviewed by Hurst [69]). Complete sequencing of several genomes provided opportunities to examine the whole array of bZIP TFs contained in given organisms. Vinson and his colleagues identified and classified bZIP proteins encoded by the *Drosophila* and human genomes [85,86]. Most recently a database bZIPDB, which contains information on protein interactions and TF-target gene relationships for 49 human bZIP TFs, has been developed [87].

Mammalian bZIPs comprise eight major families, which recognize C/EBP, TRE, CRE, CRE-like and PAR sites (Table 1). Maf bZIPs form a separate class of TFs as they recognize a substantially longer DNA sequence, MARE. Maf proteins contain extended homology region (EHR) preceding BR, which enables them to make a broader contacts with DNA [88]. Three members of the family, referred to as “small Mafs” (MafG, MafF and MafK), serve as obligatory heterodimeric partners for the members of CNC-bZIP family unique for mammals. CNC–small Maf heterodimers recognize the asymmetrical NF-E2 site composed of a half site of TRE or CRE and half site of the MARE [89]. Alignment of BR:LZ sequences of human bZIP TFs and their dimerization properties can be found in [85].

STRUCTURAL CHARACTERIZATION AND DOMAIN ORGANIZATION

Early biophysical studies in solution (based mainly on CD and NMR spectroscopy) revealed the lack of ordered

structure beyond the DNA-binding domain for several bZIP proteins [90]. Only few members of this superfamily contain regions that are stably folded in the absence of the ligand (Table 1). The three dimensional structure was determined for the zinc finger-like domain identified in the three factors ATF2, ATF7 and CREB5, comprising a subgroup of ATF family [91] (PDB ID code 1BH1), the HER domain of MafG [92] (PDB code 1K1V), and the BTB/POZ domain from BACH1 (PDB code 2IHC).

In general, in order to perform their biological functions, sequence-specific DNA binding TF proteins must translocate to the nucleus, and, once bound to DNA, either activate or repress transcription. The structural basis and molecular mechanisms explaining how these proteins are able to carry on their dual and sometimes opposing functions in the regulation of gene expression, depending on the cellular context, has just begun to be revealed. Transcriptional regulatory proteins are composed of modules that may act independently of each other. The order in which these segments are arranged into functional protein molecules varies between the factors (Fig. 3). The functional domains have been identified by a combination of biochemical and deletion mutagenesis experiments and their borders are usually poorly defined.

Functional domains of bZIP factors other than DBDs may include multiple transactivation domains (TADs), nuclear import/export segments and diverse regulatory elements. Distinct repression domains have also been identified [69,94]. For example, each, c-Fos and c-Jun, has a number of domains which activate transcription and docking sites for several kinases, such as JNK or ERK, which modulate their activities [95,96]. Often, transdominant repressors are truncated bZIP variants, which retain LZ domain but either lack the activation region, or have a modified BR sequence [69].

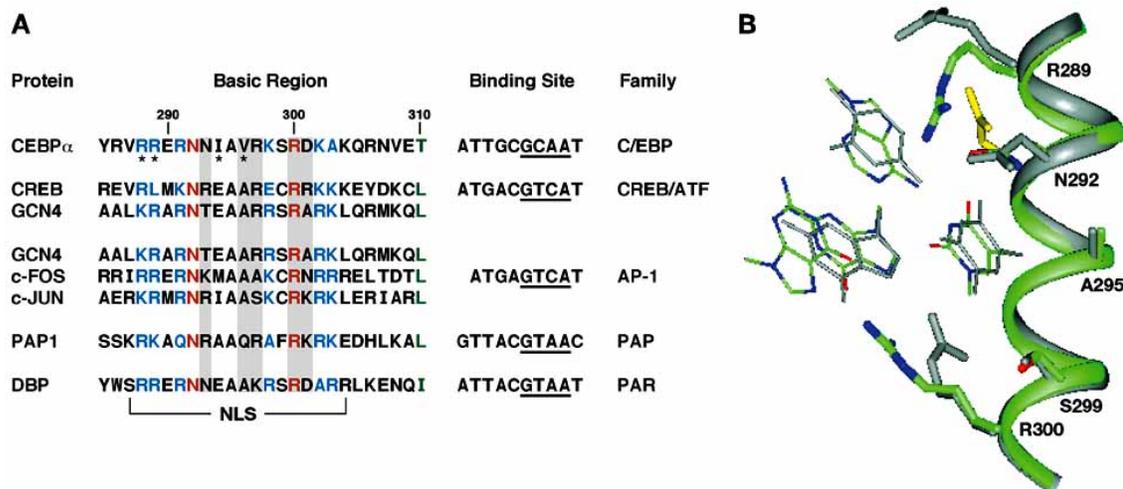


Fig. (2). DNA recognition. **(A)** Basic region amino acid sequences and DNA binding sites for representatives of bZIP families. The invariant Asn and Arg are shown in red, conserved basic residues in blue, and the first residue of the leucine zipper is green. Residues that form the KIM motif are marked by asterisks. The cores of the half-sites are underlined. **(B)**, Comparison of conformations of the conserved side chains in the basic regions of C/EBP α (green) and GCN4 (gray). The side chain of the invariant Asn residue from the PAP1 structure is shown in yellow. (Adapted from [77]).

Table 1. Human bZIP Proteins

Protein	Length	Preferred Binding Sites for Homodimers or Obligatory Heterodimers	Structural Domains ^a		Dimerization Properties
			BR:LZ	Other	
C/EBP family; reviewed in [181]					
C/EBP α	358	<i>C/EBP</i> : GATTGCGCAATC Also <i>CRE</i> and <i>PAR</i> sites	280-344		Homodimers and heterodimers within family, as well as heterodimers with Fos, Jun, ATF2 and ATF4 proteins
C/EBP β	345		269-333		
C/EBP δ	269		189-253		
C/EBP ϵ	281		202-266		
C/EBP γ	150		60-124		
AP-1 family; reviewed in [96]					
Jun subfamily:		<i>TRE (AP-1)</i> : ATGAG/CTCAT Also <i>CRE</i> sites			Jun proteins form heterodimers with Fos and ATF proteins to generate AP-1 TFs. In addition, Jun proteins form homodimers and heterodimers with CNC, and Maf proteins; ATF4 and ATF5 heterodimerize with C/EBPs, and Nrf2; whereas CREB5 with ATF2
c-Jun	331		250-314		
JunD	347		266-330		
JunB	347		266-330		
Fos subfamily:					
c-Fos	380		135-199		
FosB	338		153-217		
Fra1	271		103-167		
Fra2	326		122-186		
BATF	125		24-88		
ATF3	181		84-148		
ATF2&4 subfamily:					
ATF2	487		350-414	25-49 ^b	
ATF7	494		341-405	7-31 ^b	
CREB5	508		373-437	16-40 ^b	
ATF4	351	276-340			
ATF5	282	206-270			
CREB/ATF family; reviewed in [209]					
CREB	341	<i>CRE</i> : TGACGTCA	280-331		Homodimers
CREM	345		287-308		
CREB-H	461		240-304		
CREB3	395		172-236		
OASIS	519		288-352		
BBF2H7	520		291-355		
ATF1	271		211-268		
ATF6 family; reviewed in [210]					
ATF6	670	<i>UPRE</i> : TGACGTGG/A Other <i>CRE-like</i> sites	304-368		Homodimers
CREBL1	703		323-387		
CREB4	395		215-279		
Xbp-1	261		68-132		
PAR family; ref. [211] and references therein					
DBP	325	<i>PAR</i> : ATTACGTAAT	253-317		Homodimers
TEF	303		231-295		
HLF	295		223-287		
E4BP4	462		71-135		
MAF family; reviewed in [88]					
cMAF	403	<i>MARE</i> : TGCTGACTCAGCA, TGCTGACGTCAGCA	284-350		Large Mafs (cMAF, MAFB, NRL) form heterodimers with Fos or Jun; small Mafs (MafF, -G, -K) form homodimers or heterodimers with CNC or Fos family members
MAFB	323		234-300		
NRL	237		157-221	133-172 ^c	
MafF	164		49-113	25-62 ^c	
MafG	162		46-113	25-62 ^c	
MafK	156		49-113	25-62 ^c	

(Table 1) contd...

Protein	Length	Preferred Binding Sites for Homodimers or Obligatory Heterodimers	Structural Domains ^a		Dimerization Properties
			BR:LZ	Other	
CNC family; reviewed in [89]					
BACH1	736		553-619	34-130 ^d	Heterodimers with small Mafs
BACH2	841	<i>NF-ET</i> :	646-708	37-133 ^d	
NF-E2	373	TGCTGACTCAT	264-328		
NRF1	772		671-735		
Nrf2	605	Also <i>ARE</i> , <i>MARE</i> and	495-559		
Nrf3	694	<i>CRE</i> sites	576-640		

^a identified by SMART; ^bZinc-finger, ZNF_C2H2; ^c EHR (see text); ^d BTB/POZ

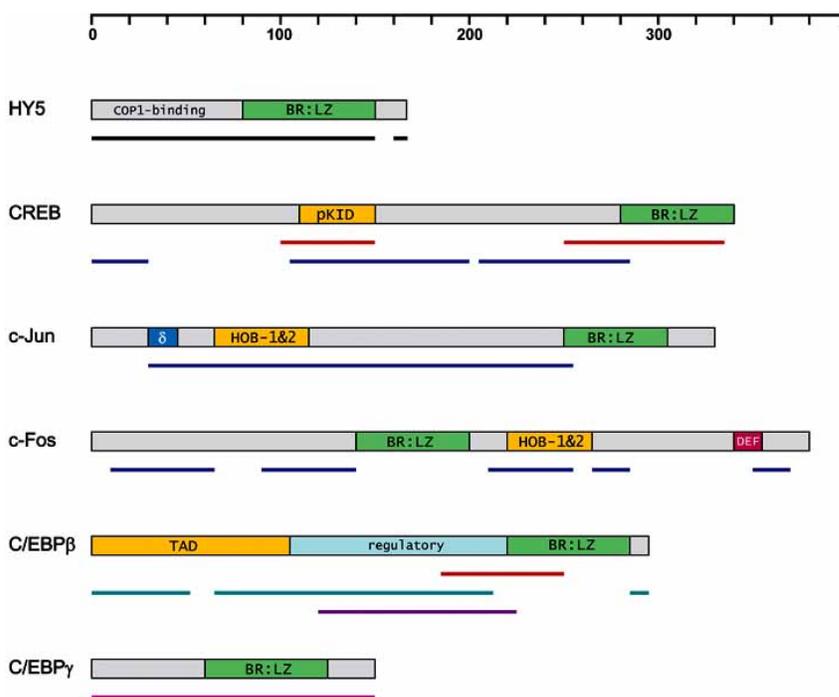


Fig. (3). Schematic representation of domain organization and predictions of ID regions (referred in various publications) for selected bZIP TFs. Regions predicted as disordered by a variety of prediction programs are marked by colored lines: black, POND [97]; dark blue, DISOPRED2 [29]; red, FoldIndex [98]; cyan, Loops/Coil definition; violet, NORSp [100]; magenta, charge-hydrophaty plot [18]. Depicted functionally relevant features are discussed in the text. Boundaries for BR:LZ domain (green) are shown as defined by SMART [102]. Regions necessary for transactivation are shown in orange. δ , and DEF domains contain docking sites for JNK and ERK, respectively. The indicated scale corresponds to the number of amino acids.

A characteristic feature of bZIP TFs is that their nuclear localization signal (NLS) is contained within the BR, which performs other functions in addition to sequence-specific DNA recognition (see below).

Dimerization, DNA-binding and transactivation/repression activities are localized to autonomous modules, which have been studied as separate polypeptide chains. Structural and functional characterization of these segments is described in the subsequent sections of this review. The information on structural properties of bZIPs in the full-length

protein context or on their putative intramolecular interactions is limited. Biophysical characterization was reported for the cAMP-responsive element binding protein (CREB) [90] and a plant bZIP TF, HY5 [97]. Studies of whole CREB by CD spectroscopy revealed that, when unbound to DNA, the protein contains 20% α -helix, 9% β -strand, and 34% β -turn, whereas 37% of its residues exist in the random coil state. DNA-binding induced a 5% increase of the α -helical content consistent with folding of the BR sequence. However, an isolated N-terminal fragment encompassing the entire region required for transactivation has a significantly

lower content of secondary structure than could be predicted by subtracting values measured for the C-terminal bZIP domain. This loss of structure in the truncated version of CREB could indicate that the bZIP domain affects the structure or stability of the rest of the molecule [90]. It was recently demonstrated that sequence-specific DNA binding of CREB induces a global conformational change within the CREB monomer, which affects structure of the kinase inducible domain (KID), rendering it refractory to the action of protein phosphatase 1 [98]. The comprehensive structural and functional characterization of HY5, a plant bZIP TF, by experimental and theoretical methods has been reported by Yoon *et al.* [97]. Using limited proteolysis in combination with mass spectrometry, CD and NMR spectroscopy, the authors demonstrated that HY5 contains a stable helical LZ domain located at the C-terminus, a molten-globule like BR region, and intrinsically disordered 77 residues on the N-terminal part. The full-length HY5 exhibited a noncooperative CD melting profile characteristic of proteins lacking stable tertiary structure.

The intrinsic structural disorder in HY5 was also implied by the primary structure analysis. The sequence of the whole protein, as well as that of investigated fragments (including the segment corresponding to a bZIP peptide), fell into the disordered region of the charge-hydrophobicity plot [18] and displayed a bias toward disorder-promoting residues in amino acid composition. The PONDR suite [99] also predicted that, except for 10 residues at the C-terminal region, HY5 is intrinsically disordered. According to a survey performed by Minezaki *et al.* [29], human bZIP-activating proteins, which do not possess functional domains other than TADs, may contain up to 70% disordered residues. Predictions of ID regions in members of the C/EBP family were reported by Miller [100]. Based on the charge-hydrophobicity plot [18], C/EBP γ was originally predicted to be completely unfolded, and more recently, 125 out of its 150 residues were predicted to be disordered by PONDR [64]. CREB, MafF and BATF are among the human TFs with the highest levels of predicted disorder, based on PONDR genome-wide analysis, with 87.68, 83.54% and 80.80% of predicted overall disorder respectively [64]. Examples of regions predicted as disordered in several bZIP proteins are depicted in Fig. (3). Despite this abundance of predicted ID regions, bZIP TFs are underrepresented in a database of protein disorder (DisProt) [101], which lists only experimentally verified cases of intrinsically unstructured proteins, such as HY5.

LEUCINE ZIPPER: MUTUALLY INDUCED FOLDING BY SELF-ASSOCIATIONS

Monomeric amphipathic helices are unstable and LZ segments fold into coiled-coil helical structures upon association with another subunit(s). bZIP peptides which cannot homodimerize (e.g., Fos and ATF4) are disordered in solution in the absence of their complementary partner [82,103]. The coil-to- α -helix transition upon dimerization is a reversible, concentration-dependent process. Dimerization constants of bZIP proteins are in the μ M range, and the estimated lifetime of dimers at 25 °C was less than 1 sec for the GCN4 homodimer [80] and less than 10 sec for the Fos-Jun heterodimer [82]. Unfolding and reassembly of double-stranded coiled-coil structures facilitates subunit exchange

between dimers. Such exchange is thought to provide a general mechanism for selective regulation of gene expression.

Thermodynamic Characterization

In the best studied case, that of yeast TF GCN4, the thermal unfolding transition occurs at 70 °C and 50 °C for concentrations of 5×10^{-4} M and 5×10^{-6} M, respectively [104]. Initial experiments utilizing CD and NMR spectrometry [80,105], calorimetry [104] and stop-flow kinetics [106] indicated that temperature- or denaturant-induced unfolding of a LZ coiled-coil structure is a cooperative, two-state transition. A series of further studies have revealed that it is a much more complicated process [107,108]. Combined optical and differential scanning calorimetry investigations of the thermal denaturation of several LZ peptides related to GCN4 show that the reaction involves several distinct steps [108]. The melting of the molecule starts at 0 °C with the fraying of the N-terminus of the leucine zipper, followed by structural changes that presumably involve repacking of the coiled-coil structure. These two concentration-independent steps are followed, at higher temperatures, by cooperative dissociation and unfolding of the two strands. Equilibrium and kinetic CD experiments demonstrated that both the stability of the dimer and the unfolding and refolding rate constants depend on the helical propensity of LZ sequences [109]. A proposed folding mechanism for GCN4-derived peptides assumes rapid equilibrium between fully unfolded peptides and partially helical (predominantly at the C-terminus) conformers, which are able to form the dimeric transition state populated by partial coiled-coil structures. The presence of nucleating helices is critical for the formation of a transition state. The increased helical propensities of variant peptides lead to the acceleration of the folding reaction, and possibly, to lowering of the transition state free energy. Residual elements of secondary structures displayed by ID regions in their unbound state are thought to play a pivotal role in the thermodynamics of folding-associated binding interactions [110].

Dimerization Specificity

Depending on the sequence of their LZ domain, bZIP proteins can form homodimers and/or form heterodimers with bZIP proteins from their own or different subfamilies. Selective dimerization is critical for bZIP TF's biological function. Heterodimers within the same family retain their DNA-binding specificity but exhibit different transactivating potential and synergy with other regulatory proteins. Cross-family heterodimerization affects DNA binding activity as well as the stability of the DNA-protein complexes, and is fundamental for the diversification of binding specificities of bZIP TFs. Heterotypic dimerization generates an expanded constellation of dimers with distinct functional properties from a small number of monomers [111]. For example, 53 unique bZIP domains identified in the human genome have the potential to create 1,405 unique dimers [85,112]. This powerful regulatory mechanism depends on the proper assembly of the specific partners. For these reasons, the dimerization properties of LZ-containing TFs have been a subject of extensive investigation. Analysis of available crystal and solution structures [113], molecular modeling [114] and

mutagenesis studies [84] has helped to establish sets of rules, which make it possible to predict interaction preferences for bZIP peptides with great precision [85,112,115]

Dimerization selectivity depends on the “complementarity” of the putative partners, which is defined by the precise distribution of hydrophilic and hydrophobic residues in the LZ regions. The dimer interface is formed by the residues located at the *a*, *d*, *e*, and *g* positions of the *(abcdefg)_n* heptad repeat (see Fig. 1). Residues at *a* and *d* positions are usually aliphatic and comprise the hydrophobic core of the duplex. An exception is the *a* position of the second heptad of many bZIP proteins, which is often occupied by asparagine to prevent higher order oligomerization [116]. The *e* and *g* positions are typically occupied by polar amino acids and influence both stability and specificity of the coil-coil. Residues located in the *g* position of one subunit and residues at the *e* position of the second subunit, which is situated five steps closer to the carboxyl terminus, are poised to interact in one direction by a “knob to holes” packing arrangement of the two helices, and often form electrostatic inter-helical bridges *g*↔*e*' (denotes the second subunit of the dimer). Homodimerization and heterodimerization are prevented when the facing side chains in position *e* and *g* create steric or charge repulsion [74]. Analysis of crystal structures of the GCN4 homodimer and Fos/Jun heterodimers bound to their cognate DNA sequences, and mutagenesis studies have es-

tablished that inter-helical salt bridges are the most decisive determinants of dimerization selectivity [85]. In choosing the dimerization partner, preventing repulsive interactions appears to be more important than increasing the number of attractive interactions. This so-called “i+5” rule explains dimerization preferences for members of the Fos and Jun subfamilies (Fig. 4). A more precise determination of dimerization specificity involves consideration of the coupling energy for *g*↔*e*' pairs, which can be calculated based on a double-mutant thermodynamic cycle [85,105].

Partnering selectivity for 49 human bZIP derived peptides were tested experimentally using coiled-coil arrays [117]. These studies confirmed most of the earlier predictions, but also identified several unexpected interactions. Analysis of fully sequenced proteomes has uncovered many more details that have to be taken into consideration in order to understand the mechanisms of diverse bZIP dimerization patterns [118]. New aspects of regulatory mechanisms underlying biological activities of bZIP TFs that emerged from proteomic studies were reviewed in [112]. As noted by the authors, the current challenge is to understand how the specific dimerization and intermolecular interactions of these proteins are regulated in changing cellular environments.

In vivo, the repertoire of bZIPs capable of associating in a given cell depends on the expression pattern of the respective proteins and on their responses to external stimuli. Fur-

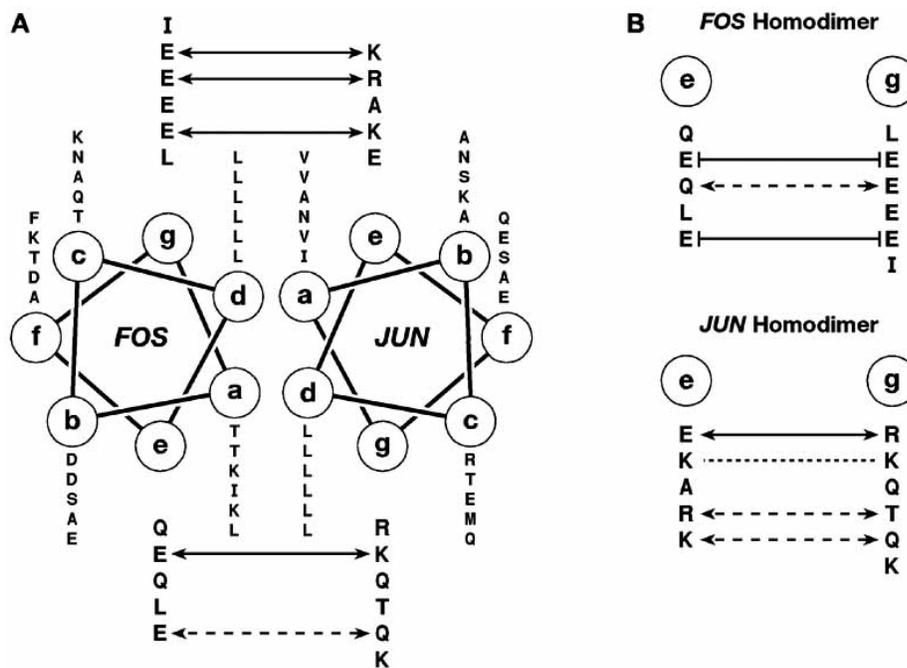


Fig. (4). Mechanism of dimerization specificity for Fos and Jun peptides. **(A)** A helical wheel diagram of the Fos–Jun coiled-coil. The sequence is read from N to C termini outwards from the wheel. The Fos–Jun heterodimer is stabilized by four electrostatic attractive interactions (solid arrows) and by a hydrogen bond interaction (dashed arrows) between the *g* and *e* positions. **(B)** Potential interactions within homodimer interfaces of Fos and Jun. The formation of Fos homodimer is prevented by four electrostatic repulsive interactions (solid lines). The formation of Jun homodimer is driven by two positive electrostatic interactions and it is additionally stabilized by hydrogen bonds and the interactions between Lys residues (dotted line), wherein the repulsive electrostatic energies between like charges are overcome by the favorable Van der Waals interactions between the methylenes of lysine side chains and hydrophobic amino acids in the *a* and *d* positions [85].

thermore, the reversible nature of the dimerization process facilitates the interconversion between early and late activated complexes, and contributes to time and stimuli-dependent differential activation of specific genes. Such modulation of biological activities by control of dimer composition has been observed in the C/EBP family [119] and in the Fos/Jun family [70]. Through the change of dimeric partner, small Mafs can switch transcriptional activity from repression to activation [89].

THE MANY FACES OF THE BASIC REGION

The short BR performs a remarkable number of functions critical for the biological activity of bZIP proteins: cytoplasm-to-nucleus translocation, specific DNA recognition and protein-protein interactions possibly contributing to the regulation of gene expression. These functions can be regulated by intramolecular interactions of BRs with other regions, their phosphorylation and binding to auxiliary proteins.

In the unbound state, the BR elements of bZIP proteins exhibit differential, temperature-dependent helical content. It has been shown by CD and NMR spectroscopy that in the absence of DNA, BR of GCN4 and C/EBP β are largely unfolded in solution [80,120] and have been described as a dynamic ensemble of transiently formed helical structures. The population of transient helical conformations increases with decreasing temperature [121]. The BR of plant TF, HY5, has been characterized as a molten globule [97]. In contrast, in the absence of DNA, the ATF4 BR adopts a stable helical conformation, as revealed by the crystal structure of the ATF4-C/EBP β heterodimer (see Fig. 5A) [103]. The propensity to form a helical structure is defined by the amino acid sequence. Conserved BR sequences of selected bZIP proteins are preceded by helix-capping residues, which may be responsible for stabilizing their partial α -helical structures [122].

The DNA-dependent folding transition of the basic domain has been documented by CD, fluorescence, and NMR spectroscopic techniques. The requirement of coupling of local folding to site-specific DNA binding has been deduced from theoretical analysis of calorimetric data. Spolar and Record [44] proposed that the significant gain in total association entropy observed in specific protein-DNA interactions is mainly due to changes in water-accessible surface area, and can be explained by the mechanism of protein folding coupled to binding. During such a process, the loss of conformational entropy resulting from structural ordering may be compensated by gain of entropy arising from the burial of large nonpolar surfaces on the complex formation. Non-specific binding, which involves only electrostatic interactions with the DNA phosphodiester backbone, was shown to not produce this ordering effect [43,44]. It has subsequently been suggested that nascent helices transiently formed in the unbound state may significantly reduce the entropy penalty associated with DNA binding by a flexible domain [121], and contribute to the increased affinity of specific binding [122]. Thus, the balance between the specificity and affinity of DNA binding depends on the amino acid sequence and can be regulated by the length of partial helical

structures existing in the basic domain prior to complex formation.

The BR can also maintain specific interactions with viral and cellular proteins that function as transcriptional coactivators. The multiple factor-bridging protein 1 (MBF1) stimulates transcription through selective associations with distinct subclasses of bZIP proteins, linking them to the general transcription factor TATA-box binding protein (TBP) [123]. It has been demonstrated that bZIP of Jun (but not Fos) binds directly to the N-terminus of the human TBP-associated factor-1 (hTAF1), causing a de-repression of TFIID-driven transcription [124]. In certain cells, CRE-dependent transcription is enhanced by TORC (transducer of regulated CREB activity) protein, which mediates interactions of CREB bZIP with the TAF4/TAF_{II}130 component of TFIID [125]. Specific BR sequences are also targeted by the human T-cell leukemia virus Tax, whereas the hepatitis B virus pX interacts with a broad range of bZIP proteins [126]. Miotto and Struhl [127] showed that the selective interactions with MBF1 and the related Chameau histone acetylases (or HBO1, the human homolog) are mediated by BR residues that face away from the DNA in the protein-DNA complex. Chameau interacts with AP-1 during *Drosophila* development via two arginine residues (R221 and R232 in JunD), whereas binding of MBF1 to JunD requires glutamate or glutamine residues located in the fork region, in addition to the two arginines. MBF1 and Chameau compete for binding to the BR of JunD and do not cooperate in its transcriptional activation. However, either MBF1 or Chameau can synergistically coactivate AP-1 genes with pX.

The disordered structure of the BR prior to specific DNA binding facilitates a variety of protein-protein interactions, which regulate bZIP activity. All information necessary to direct bZIP translocation to the nucleus is contained within the BR [69,128]. The NLS motif of bZIPs consists of two clusters of Lys/Arg residues separated by a linker of 10–12 residues (Fig. 2). This type of classic NLS is recognized by a component of transport receptors, importin- α /Kap α [129,130]. The NLS-binding domain of importin- α comprises 10 helical repeats (armadillo motifs), which form a concave surface containing multiple pockets for NLS cargoes. The crystal structure of the NLS peptide-importin- α complex [131] revealed that the bipartite NLS peptide binds to the receptor in an extended conformation (Fig. 5B). The two basic clusters are located in separate specific binding pockets of the receptor, whereas residues from the linker are poorly ordered.

Nuclear import/export as well as DNA-binding of many bZIP proteins, including Jun, C/EBP β , and several plant factors, is regulated by phosphorylation [132]. For example, phosphorylation of the conserved Ser239 in C/EBP β impairs its binding to the cognate DNA sequence and induces nuclear export [133]. It has been proposed that the SXEY sequence motif contained in the BR, which is conserved in this family, may be a site of phospho-dependent interactions with BRCA1 [100]. Other functional sites present in this region include the kinase interaction motif (KIM) [100], and a cysteine residue which is a sensor of oxidative stress in c-Jun, JunD, c-Fos, and EB1 [134,135].

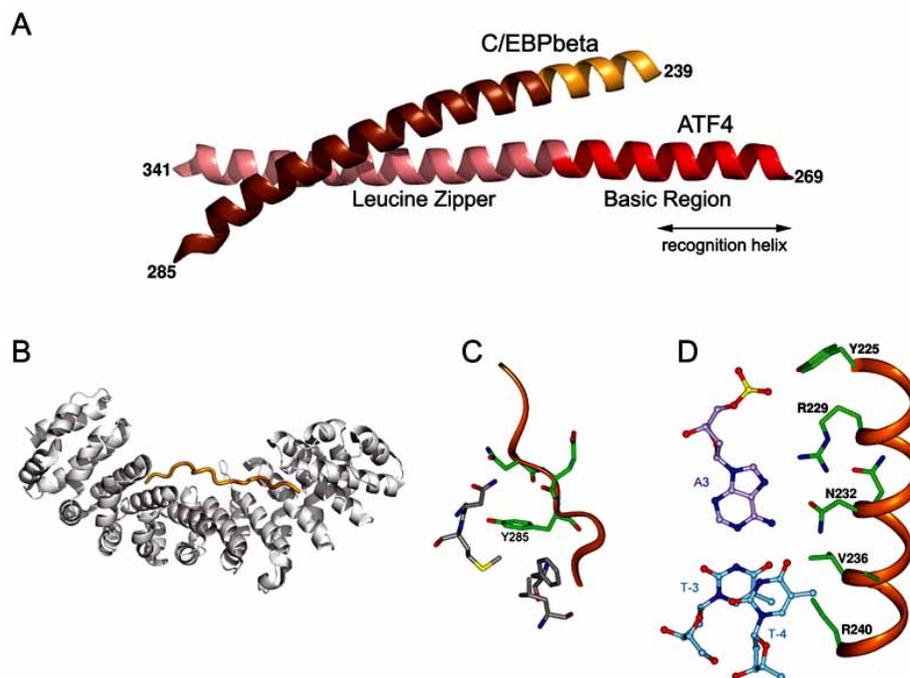


Fig. (5). Conformational flexibility and binding polymorphism of BR from C/EBPs (orange). **(A)** Structure of C/EBPβ(224-285)–ATF4(280-341) heterodimer in the absence of DNA; (PDB code 1CI6). Note that the entire region corresponding to the C/EBPβ recognition helix is missing. **(B)** Structure of the classic, bipartite NLS peptide (orange) bound to armadillo domain of importin α (gray); (PDB code 1EJY). **(C)** Hypothetical interactions of C/EBPα BR peptide with the BRCT domain of BRCA1. Y285 corresponds to Y225 in C/EBPβ. **(D)** C/EBPβ recognition helix (residues 225-240) bound to cognate DNA (PDB code 1H8A).

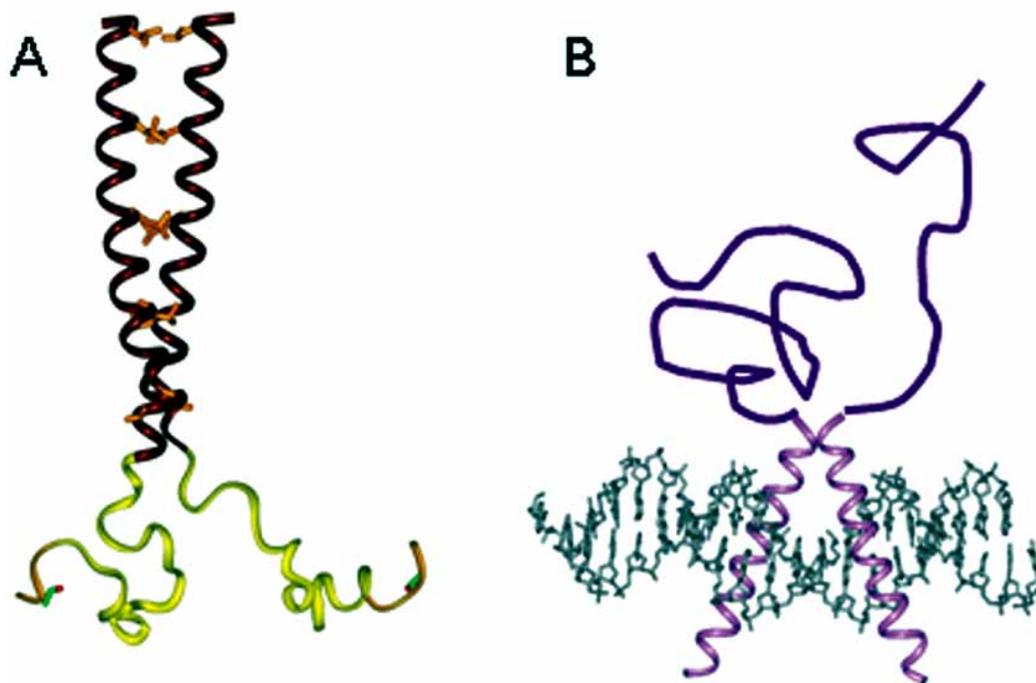


Fig. (6). Possible intermediates in the monomer and dimer pathways during assembly of bZIP–DNA complexes.

The binding polymorphism and conformational plasticity exhibited by BRs underscore the importance of structural disorder for promoting transient interactions with multiple partners. Several of these interactions are mutually exclusive and may be utilized for the sequential ordering of cellular events in response to external signals. Importantly, the binding site for transcriptional coactivators such as MBF1, TORC, and Chameau/HBO1 is created only upon binding of bZIP protein to cognate DNA sequence. This ensures occurrence of these associations in the proper time and location.

ASSEMBLY OF bZIP-DNA COMPLEXES

Dimerization is considered a prerequisite for DNA-binding activity, as LZ mutants that cannot dimerize fail to form stable protein-DNA complexes [72]. However, the sequence of events leading to the formation of a dimeric protein-DNA complex is a matter of debate. In principle, two main pathways are possible: (1) initial formation of a bZIP dimer followed by DNA binding (dimer pathway), or (2) initial formation of a bZIP monomer-DNA complex that subsequently binds a second bZIP monomer (monomer pathway) [136] (see Fig. 6). Several bZIP peptides have been shown to readily dimerize in the absence of DNA, whereas there are only few examples of monomeric bZIP-DNA complexes. On the other hand, the bZIP dimer is relatively unstable, and when not bound to DNA it rapidly dissociates to monomers [82,104]. bZIPs, which cannot form homodimers will be present in solution as monomers in the absence of a complementary partner, as shown for Fos. Wu *et al.* [137] observed that homodimerizing CREB protein also exists primarily as a monomer in solution, and its dimerization is DNA-dependent. It has further been suggested that the weak association constants of dimerizing TFs ensure that stable protein-DNA complexes can assemble only upon binding to a specific sequence [138]. The rate of assembly via the dimer pathway is expected to be limited by the rate of protein dimerization. In contrast, protein-DNA association is directed by long distance electrostatic interactions and should be much faster. Importantly, if the monomer-DNA intermediate is unstable relative to the dimer-DNA complex, it is less likely to become kinetically trapped at a nonspecific site. The monomer pathway thus offers kinetic advantages and provides a more efficient means for discrimination against nonspecific binding [139]. Schepartz and colleagues provided compelling kinetic and spectroscopic evidence that certain bZIPs (e.g., ATF-2, Fos and Jun) and related helix-loop-helix zipper families (e.g., Max) follow an assembly pathway in which two monomers sequentially bind to DNA and form their dimerization interface while bound [139,140]. The obligatory dimer pathway of DNA recognition has also been challenged by others [141,142].

The shortcoming of these studies is the lack of data on the behavior of full-length bZIP proteins, in which intramolecular interactions may influence dimerization and DNA-binding properties. In many instances DNA binding activity of bZIP proteins (e.g., C/EBP β), is autorepressed in their inactive state [143]. Also, dimerization is critical for the stability of C/EBP proteins, which in monomeric form undergo rapid ubiquitination-dependent degradation by the 26S proteasome [144]. Moreover, interactions of several TFs

with importins have been observed to occur in the dimeric form [128,145]. These considerations indicate that both pathways, and sometimes a combination of two may be utilized *in vivo*. The mechanism of assembly may depend on many factors such as the helical propensity of the BR and LZ sequences (see above), the expression pattern of suitable partners, and the presence of auxiliary proteins [126] and may vary on a case-by-case basis.

THE ELUSIVE TRANSACTIVATION DOMAIN

Sequence-specific DNA binding TFs (transcriptional activators and repressors) stimulate transcription of their target genes by regulating the assembly and/or activity of transcriptional initiation complexes. They elicit their effect by recruiting members of a diverse family of coactivators, which initiate local opening of chromatin structure and mediate recruitment of the RNA polymerase II complex (Pol II) to the transcriptional start site. Transcription-activating proteins possess specialized distinct domain(s) responsible for transactivation, which interact with the basal transcriptional machinery either directly or via mediator proteins. Activation is a complex, cooperative process that requires dynamic rearrangement of contacts between TADs, general transcription factors, various coactivators, and chromatin remodeling factors.

TADs themselves often have modular structure and may be composed of multiple small transactivating elements (TEs). Some of these elements show little independent activity, but can synergistically activate transcription. At the sequence level, TAD regions are usually enriched in a single amino acid, a property that was the basis for an early classification scheme that divided TADs into groups comprising acidic, basic, glutamine-rich, proline-rich, serine/threonine-rich and isoleucine-rich domains (reviewed in [146]). However, further biochemical studies have indicated that the preponderance of one or two amino acids does not correlate with specific function, which rather seems to depend on short sequences with specific patterns of hydrophobic and aromatic residues [147]. TADs show very little sequence similarity, even among members of the same family, and yet they often compete for a common target protein and may inhibit each other by a mechanism referred to as "squelching" [148].

When expressed as recombinant proteins, TADs appear to be disordered. Structural disorder was first experimentally documented in the TAD of the Herpes Simplex Virus transactivating protein VP16 [149]. Among bZIPs, the TAD region of CREB [90], the C-terminal part of the ATF-2 TAD [91], the C-terminal TAD of Fos [150], and most recently TAD of the plant bZIP TF HY5 [97] have been experimentally characterized as disordered. These results suggested that TAD regions may adopt ordered structures only when bound to their molecular partners. The observed low-sequence similarity, combined with the lack of well-defined structural features of these modules has made it difficult to understand their apparent ability to maintain common, specific interactions with multiple coregulators. Furthermore, in many cases the domain boundaries were poorly defined [151]. Characterization of the conformational propensities of different TADs and their specific contacts with coactivators

have been further complicated by their involvement in complex interdependent networks of intermolecular interactions regulated by phosphorylation and other posttranscriptional modifications (see below).

It is now well established that the majority of TADs rely on relatively short segments with specific patterns of hydrophobic residues to recognize their target proteins. The commonly accepted model is that of folding-associated recognition. Increased binding specificity at the expense of thermodynamic stability, which characterizes binding coupled to folding of ID regions (see above), facilitates transient interactions with a range of different proteins with major structural and functional differences. For example, depending on cellular cues, the disordered C-terminal TAD of c-Fos may interact with TBP, with the general coactivator CREB-binding protein (CBP), or with the regulatory factor Smad3 [150,152]. Similarly, c-Myc transactivating activity is regulated by binding to variety of factors including TBP, MM-1 and p21 (Ref. [151] and references therein). In this way, the flexibility of cellular response is a direct consequence of the conformational plasticity of TADs.

The versatility of binding and the affinity toward target proteins depend on the presence of specific binding motifs, and the structural and dynamic properties of the polypeptide. Both direct induction of secondary structure and wrapped folding-on-binding onto the target protein surface have been observed [4,12,33]. The functional recognition motifs are often flanked by clusters of charged amino acids. Hermann *et al.* [153] proposed a two-step model for the association of such TADs with their partners. In the first step, initial low-affinity complex formation is driven by ionic interactions, and is followed by slow interconversion into a more stable form stabilized by hydrophobic interactions. The second step is accompanied by folding of the TAD into a structure that closely fits the molecular surface of the ligand. Recent work of Sugase *et al.* [33] unraveled the mechanism by which the phosphorylated TAD of CREB (pKID) folds on the surface of KIX domain of CBP (see below). Initially, the disordered pKID forms an ensemble of transient encounter complexes at multiple sites on the KIX surface, which is followed by the formation of an intermediate complex, in which the pKID is only partially folded. In this case, transient encounter complexes are stabilized predominantly by non-specific hydrophobic contacts.

TE segments containing short protein recognition motifs often display residual secondary structure features in their unbound states. These partially formed structures can facilitate formation of complexes and contribute to a decrease in the entropic penalty of binding, in the same manner as discussed earlier in the case of BR-LZ domains [47,110]. Hydrophobic-solvent inducible amphipathic-helical segments have been observed in several acidic TADs [154] and could be predicted by sequence analysis.

Examples of common protein recognition motifs identified within TADs of various TFs are shown in Fig. (7). The TADs regions of Jun subfamily contain two adjacent motifs, HOB1 and HOB2, which serve as cooperating activation modules. Related sequences are also present in c-Fos, but not in other subfamily members. Similarly, activating C/EBP proteins have conserved regions named Box A and Box B

embedded in their common acidic TADs [155]. Segments corresponding to HOB2 from Fos/Jun and Box B from C/EBP proteins share weak sequence similarity. The entire fragment encompassing Box A and B is required for interactions of C/EBPs with the TAZ2 domain of CBP/p300 [155,156] and possibly the same segment also mediates interaction with TBP [155]. Homology BOX B comprises the L/FXXLF motif and corresponds to a “signature helix” found in TADs of many transcriptional activators, including the tumor suppressor protein, p53 [157]. The L/FS/ADLF sequence, conserved among C/EBP family members, has been found in other TFs (e.g., ALL1, NFAT1) and has been identified as a critical component of the TAF9 binding motif [158].

Regulation by Phosphorylation

Accumulated evidence indicates that the transactivating activity of many TFs is regulated by phosphorylation. Phosphorylation of TADs may modulate direct binding to protein ligands, as well as intramolecular interactions. All the serine residues present in the p53 TAD were found to be phosphorylated by several distinct kinases. Depending on the phosphorylation status, p53 may interact with MDM2, TAF9, and/or the TAZ2 domain of p300 (reviewed in [159]). The transactivation capacity of Jun, JunD, and ATF-2 is stimulated by the Jun N-terminal kinase (JNK), which specifically phosphorylates Ser63 and Ser73 within the HOB1 motif (see Fig. (7)), whereas the analogous Thr232 in HOB1 of Fos is phosphorylated by ERK [160], reviewed by Wagner [96]. Activation of CREB transcriptional activity depends on the recruitment of the transcriptional coactivator CBP. Association of CREB with the KIX domain of CBP is induced by protein kinase A (PKA) mediated phosphorylation of Ser133, which is located in the kinase-inducible domain (KID) of CREB [161]. In response to ionizing radiation, ATM-dependent phosphorylation of CREB on Ser121 inhibits CREB–CBP interactions [162]. A conserved serine in homology BOX A of C/EBP β has been shown to be phosphorylated by CDK1 in a cell cycle-dependent manner [163]. Furthermore, the primary phosphorylation often generates new phosphodependent protein–protein interaction motifs. Fos phosphorylated by ERK on multiple residues within its C-terminal TAD becomes a target of prolyl isomerase Pin1 activity. It is thought that conformational changes induced by isomerization of the peptidyl-prolyl bond lead to further enhancement of Fos transcriptional activity [164]. It was noted [100] that phosphorylation of the serine residue within BOX A that is conserved among activating members of the C/EBP family would generate the pSXXI/L (pS denotes phosphorylated serine) motif, which could be recognized by a pair of BRCT repeats from the C-terminus of the PAX-transactivation-domain-interacting protein (PTIP) [165]. Conformational flexibility enables these TADs to interact with both modifying enzymes and their cognate coactivators, thus allowing for context-dependent recognition of TFs by signaling and transcriptional machineries, respectively [166].

Constitutive and Inducible Recognition of Activators by the KIX Domain of CBP/p300

Structural and mechanistic insights into folding of unstructured peptides induced by binding have been revealed

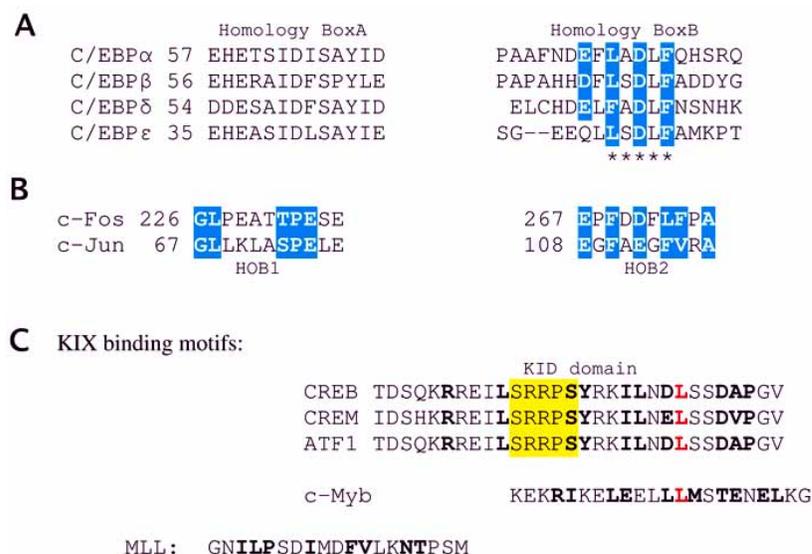


Fig. (7). Protein recognition motifs identified in TADs. **(A)** Sequence alignment of Box A and B homologies present in C/EBPs. Residues comprising TAF9 binding motif are marked by asterisks. **(B)** Homology boxes, HOB1 and HOB2, from c-Jun/c-Fos. Identical residues representing consensus and sequence similarities with Box B are highlighted in cyan. **(C)** KIX binding sequences. Residues contacting KIX are shown in bold print. The conserved phosphorylation motif in KID domains is highlighted in yellow and the critical Leu (see text) is shown in red.

by solution structures of complexes of distinct CBP/p300 domains (see below) with several classes of TADs [167-170]. Particularly well characterized by structural and thermodynamic methods are interactions involving the KIX domain, which can recognize TADs of many different TFs. These studies reveal the role of the propensity to form helical structures of unbound peptides in constitutive binding and a structural basis for phosphorylation-regulated inducible binding of activators to KIX. Solution structures of KIX in complex with different peptides are shown in Fig. (8). The KIX domain adopts a compact helical bundle structure composed of three α helices (H1, H2 and H3) and two short 3_{10} helices held together by an extensive hydrophobic core. Two patches of hydrophobic residues are located on the opposite sides of the KIX surface, which serve as binding sites for distinct classes of TADs. The KIX domain of CBP can participate in both, phosphorylation-dependent and phosphorylation-independent, interactions with TFs. Association of CREB with the KIX domain of CBP requires phosphorylation of Ser133 located in the kinase-inducible domain (KID) of CREB [161]. The solution structure of pKID bound to the KIX domain [171] is shown in Fig. (8B). As demonstrated by NMR studies [171], free phosphorylated KID (pKID) is mostly unstructured in solution. Upon binding to KIX, the pKID peptide folds into two separate helices (N- and C-terminal) joined by a loop and positioned almost perpendicularly to each other. Both helices are stabilized by packing against the surface of KIX. The C-terminal helix is situated in the shallow hydrophobic groove of the KIX domain formed by the H1 and H3 helices and contributes most of the interactions with KIX. The phosphorylated Ser133 is located at the N terminus of the C-terminal helix of pKID in the vicinity of Tyr658 and Lys662 of KIX. Mutagenesis studies indicated that hydrogen bonding interaction between

pSer133 and Tyr658 plays decisive role in stabilization of the pKID-KIX complex. Basal affinity of unphosphorylated KID to KIX is two orders of magnitude lower than the affinity of KID in the phosphorylated state.

On the other hand, c-Myb peptide interacts with KIX via the same binding surface as pKID in a constitutive fashion. Upon binding to KIX, c-Myb peptide folds into a single 16 residue-long α helix that bends to optimize interactions with KIX. Nearly half of the interactions between the two proteins are provided by a critical leucine residue from c-Myb (Leu302) located at the kink of the helix, which inserts its side chain deeply into the hydrophobic pocket of KIX. c-Myb binds to KIX with an affinity sevenfold higher than unphosphorylated KID, and 20-50 fold lower than phosphorylated KID. These differences enable c-Myb to be a constitutive activator, whereas upon phosphorylation, CREB competes efficiently with c-Myb and other constitutive activators for binding to a limited amount of CBP [172]. Structured-based sequence alignment (see Fig. 7C) shows the lack of sequence similarity between KID domains of CREB/ATF proteins and the c-Myb peptide. It is believed that the ability of the activator to form an amphipathic helix is necessary to form a constitutive low affinity complex with KIX. Analysis of interactions by isothermal calorimetry (ITC) is consistent with enthalpy driven complex formation between pKID and KIX, and with binding of c-Myb to KIX depending on both enthalpy and entropy components. These different modes of binding arise from the differences in the length and distribution of nascent helical structures present in unbound peptides.

There is another deep hydrophobic groove created by side chains of three helices located on the opposite side of KIX, which can bind mixed lineage leukemia protein (MLL)

and possibly Jun, HTLV-1 Tax, and HIV Tat-1. Recent studies showed that KIX is able to bind simultaneously to two different factors in a cooperative manner (Fig. 8A) [170]. ITC experiments demonstrated that binding of MLL increases the affinity of KIX for either pKID or c-Myb; conversely MLL binds with a twofold higher affinity to the KIX–c-Myb binary complex than to KIX alone. The structure of the KIX–c-Myb–MLL ternary complex revealed allosteric changes in the MLL binding site of KIX, which resulted in the formation of favorable electrostatic interactions between KIX and c-Myb that may account for the synergy between two factors participating in the regulation of the same gene.

Phosphorylation of CREB Ser133 creates a potential phosphorylation site for glycogen synthase kinase 3 (GSK-

3 β) at CREB Ser129. GSK-3 β recognizes and modifies substrates that are phosphorylated at position +3 with respect to the target phosphorylation site (consensus SXXXpS, where pS denotes phosphorylated serine residue) [173]. Molecular modeling studies [174] indicate that, in the ternary GSK-3, ATP, pKID complex, two helices of pKID are almost collinear. pKID peptide interacts with GSK-3 β residues: Phe67, Gln89, and Asn95, whereas the binding pocket for phosphorylated Ser133 consists of side chains of Arg96, Arg180 and Lys205 of GSK-3 β (Fig. 8D).

FLEXIBLE REGULATORY REGIONS

The functional domains containing a residual structure are usually separated by stretches of disordered regions lacking any elements of the regular secondary structure. In con-

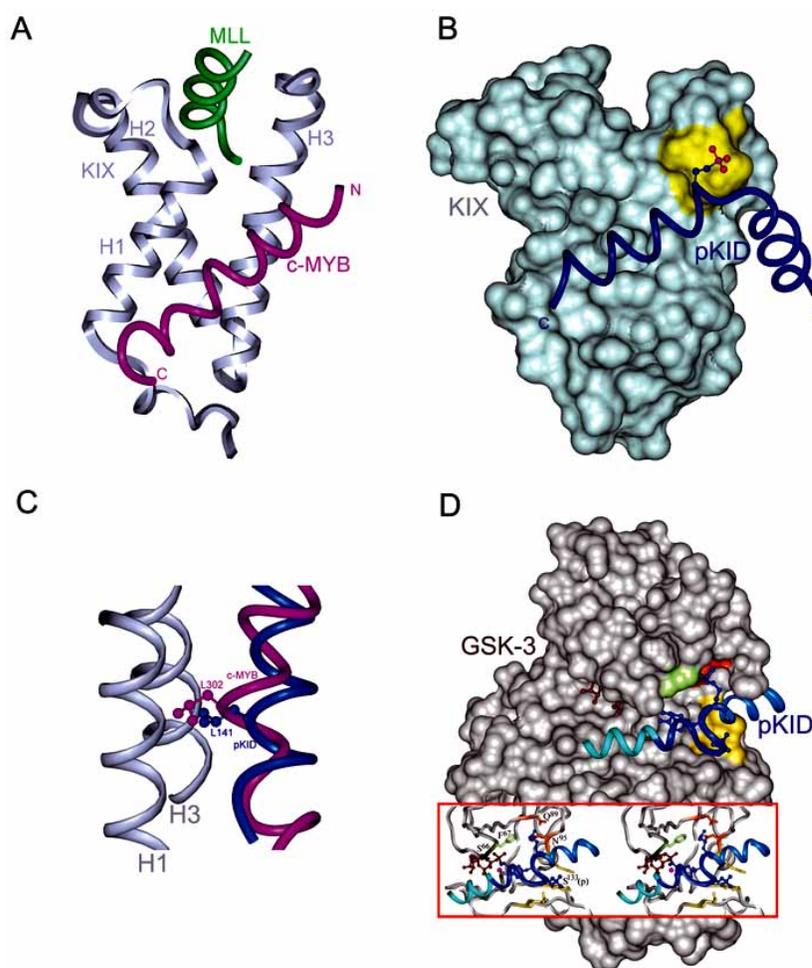


Fig. (8). Binding commonality and binding diversity displayed by pKID domain of CREB. (A) Synergistic binding of c-MYB (purple) and MLL (green) to KIX (gray); (PDB code 2AGH). (B) pKID (navy) bound to KIX; (PDB code 1KDX). Solvent accessible surface of KIX is shown in gray, Tyr658 in yellow, side chain of phosphorylated Ser133 is represented as *balls-and-sticks*. (C) Superposition of pKID and c-MYB peptides bound to the same site on KIX (competitive binding). The bend of the α helix of c-Myb facilitates optimal interactions with KIX and enables a critical Leu side chain to penetrate much more deeply than the equivalent Leu from pKID does into the hydrophobic pocket of KIX. (D) Model structure of pKID bound to GSK-3 β (beige). The phosphate binding site is depicted in yellow; (adapted from [174]).

trast to regions with the potential to fold into the helix upon binding to the proper ligand, these regions remain unstructured, irrespective of their environment. Their role can be as simple as spacers or they may provide the large conformational freedom required for the activation and/or repression domains of the DNA bound TFs to reach other components of transcriptional apparatus. Very often, however, these unstructured linkers harbor important functional sites and play critical role in the activation and regulation of biological functions of numerous proteins [57]. For example, two regulatory segments termed RD1 and RD2, which regulate DNA binding and transactivating activity of C/EBP β [143] and C/EBP ϵ [175], as well as the δ domain, which control TAD of c-Jun, were predicted to be disordered by a variety of prediction methods [100] (see Fig. 3).

Polypeptide chains in extended coil form are characterized by lower sequence complexity and higher net charge than other types of disorder and often have a high content of proline residues [16]. According to recent analysis, most significant amino acid patterns associated with protein disorder are clusters of three or four glutamic acids and those containing prolines [176]. Consistent with this type of sequence bias, solvent-accessible, unstructured regions are sensitive to proteolysis [177] and are often enriched in PEST sequences (e.g., regions rich in proline, glutamic acid, serine and threonine residues) that confer multiple protein degradation signals and phosphorylation targets [178]. Significantly, in several cases unstructured regions were directly recognized and cleaved by the 20S proteasome [14]. Structureless regulatory regions also serve as platforms for docking sites and recognition motifs facilitating interactions with specific modifying enzymes.

Transcriptional regulatory proteins are end points of many signal transduction cascades and their activities are subjected to multiple modes of regulation [166,179]. In response to external stimuli TFs have to be inactivated as quickly as they are induced. A number of short-lived factors such as NF- κ B, AP-1 and C/EBPs, undergoes selective degradation by ubiquitin proteasome pathway [144,180]. Activity of TFs is also controlled by phosphorylation, sumoylation and acetylation on lysine residues and methylation on arginine and lysine residues. Flexible regulatory regions of C/EBP proteins are targeted for sumoylation [175] in addition to phosphorylation and contain putative phospho-Ser/Thr recognition sites for WW and Polo-box modules of Pin1 and Polo-like kinases, respectively [100]. Majority of bZIP proteins are downstream targets for mitogen-activated protein (MAP) kinases (e.g., the components of RAS pathway) [181], which often act in concerted manner with Ser/Thr phosphatases to regulate TFs in a graded fashion (reviewed in [166]). MAP kinases phosphorylate very similar motifs containing minimal consensus sequence Ser/Thr-Pro and recognize specific substrates through binding surfaces (docking grooves) located outside the catalytic active site. These docking grooves bind to short peptide motifs (docking sites) residing in target proteins. Depending on the number and sequence of these recognition motifs, substrates are phosphorylated by specific subsets of MAPKs (reviewed in [182]). For example, a D docking site located within the δ domain of c-Jun (Fig. 3) mediates phosphorylation of c-Jun by JNK, whereas a DEF motif is required for ERK phos-

phorylation of c-Fos. On the other hand, JunD, which contains both D and DEF motifs, undergoes phosphorylation either by ERK or JNK. Certain proteins (e.g., JunB) contain the docking site but lack the phospho-acceptor motif and act as scaffolds for assembly of complexes containing components of MAP kinase cascades. Such a mode of specific recognition requires large area of interactions between the kinase and its substrate that can not be achieved by interactions of two globular proteins and depends on the ability of the unstructured region of one to wrap around the other.

MULTIPROTEIN-DNA COMPLEXES IN COMBINATORIAL GENE REGULATION

Transcriptional regulation of the target gene requires the synergetic action of multiple TFs bound to cis-regulatory elements, and cooperation of diverse families of coregulators [183,184]. Activator-binding sites are often clustered into enhancers that function as separate regulatory modules. Specificity in transcription is achieved by cooperative recruitment of several sequence-specific DNA-binding TFs and formation of multiprotein-enhancer DNA complexes termed enhanceosomes [185]. The gene-specific architecture of such a complex is dependent on the spatial arrangement of cis-regulatory elements and the correct array of bound activators which together generate a network of protein-protein and protein-DNA interactions unique to a given enhancer. TFs bound to composite regulatory sites modulate each other's activity, thus, a particular factor can assume different functions when bound to different enhancers. Moreover, cooperative interactions of TFs increase stability of the multiprotein-DNA complex, enhance the recognition specificity, and further extend the combinatorial potential of transcription regulation. Importantly, concerted action of numerous signal-activated TFs allows for integration of independent signaling pathways at a specific promoter.

Direct Protein-Protein Interactions of bZIPs with Other TF Proteins at Composite Sites on Promoters

Many recognition sequences in natural promoter and enhancer regions deviate from the optimal binding sites for regulatory proteins. The weak binding affinities to these suboptimal sites impose a requirement for interactions with others TFs. The inherent flexibility of bZIP regions facilitates direct association of bZIP proteins with other structurally unrelated TFs (Fig. 9) such as NF- κ B/ Rel family members, NFAT, cMyb, and SMADs bound to a nearby regulatory element.

Synergistic regulation of transcription specificity by multiple factors is best illustrated by studies of regulatory complexes containing Jun/Fos proteins [70]. Nonconsensus AP-1 binding sites are present downstream of most NFAT sites in the promoter region of several genes in immune system cells. Both AP-1 and NFAT show only weak independent binding at such composite sites and the productive response to NFAT requires concomitant activation of AP-1 proteins. The activation of NFAT proteins is regulated by calcium and calcineurin, whereas that of AP-1 proteins by RAS/MAP kinase and PKC, thus composite sites, such as NFAT:AP-1, integrate signals from distinct signal transduction pathways. The cocrystal structure of Fos-Jun-NFAT DBDs bound to the

DNA fragment containing composite site ARRE2 from the IL-2 gene promoter provided the structural basis for transcriptional cooperativity of these TFs [186]. NFAT and Fos–Jun heterodimer bind on the same face of the DNA to adjacent binding sites (Fig. 9). Interactions of AP-1 with NFAT require substantial bending of both the coiled-coil and the DNA duplex. The majority of the protein–protein interactions are between NFAT and the Fos subunit of AP-1. These contacts help to orient Fos–Jun heterodimer in a unique orientation with respect to DNA, so that Jun binds to the half of the asymmetric AP-1 site close to NFAT. In contrast, both orientations were observed in the crystal of the Fos–Jun–DNA ternary complex [74]. Of note, protein–protein interactions are not critical for cooperative binding of c-Jun–ATF2 heterodimer and IRF-3 to the interferon β enhancer. In this case, the sites for adjacent proteins overlap and the cooperativity of binding arises from sequence-dependent conformability of the DNA [187].

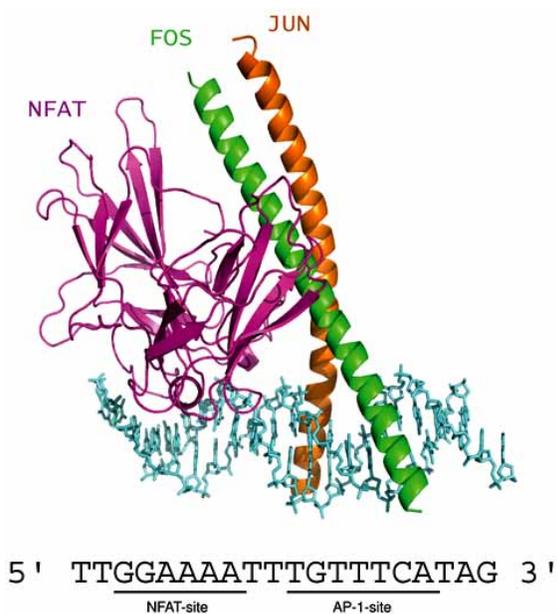


Fig. (9). The NFAT–Fos–Jun–DNA quaternary complex (PDB code 1A02). The ARRE2 DNA sequence in the crystal is shown below the figure. Note the deviation from the consensus sequence (TGAGTCA) recognized by AP-1 proteins. Protein–protein interactions between NFAT and the LZ of AP-1 enable the two TFs to bind DNA cooperatively, and regulate coordinately the IL-2 promoter.

Another well studied example of transcriptional synergy is the cooperation of c-Myb with C/EBP family members in induction of the *mim-1* gene expression during myeloid cell differentiation. The binding sites for c-Myb and C/EBP proteins on the *mim-1* gene promoter are separated by a sequence of 80 base pairs. As demonstrated by atomic force microscopy, their cooperative interaction involves DNA loop formation. X-ray crystallography provided a high resolution

view of c-Myb–C/EBP β interactions and the mechanism of DNA looping stabilization [188]. In the crystal of the c-Myb–C/EBP β –DNA complex, C-terminal portions of C/EBP β chains A and B interact with a subdomain of c-Myb (bound to another DNA fragment) to form a four-helix bundle structure (Fig. 10). The side chain of Lys332 from C/EBP β chain B makes a salt bridge with the phosphate of c-Myb-bound DNA molecule, presumably contributing to DNA loop formation. The C-terminal LZ extensions of C/EBP β , which interact with c-Myb are unstructured without c-Myb and assume helical conformation upon binding to c-Myb. Reciprocally, C/EBP β chain A stabilizes the conformation of a c-Myb loop that interacts with the DNA backbone.

Assembly of TFs on the enhancer may as well be affected by non-DNA binding proteins. For example, binding of C/EBP β to suboptimal site located on the C-reactive protein gene promoter is enhanced by protein–protein interaction with c-Rel, which is not itself bound to DNA [189]. Also, several viral proteins are known to associate with bZIPs to gain access to the promoter (see above). X-ray crystallographic studies of multiprotein complexes bound to DNA were reviewed in [190] and [191].

Architectural TFs

Long-range interactions between cis-regulatory modules (enhancers, silencers and promoters) are mediated by architectural proteins, which facilitate interactions among proteins bound to separate recognition sites by binding to multiple sequences within enhancers and inducing changes in DNA structure. For example, BACH1 helicase forms heterodimers with small Maf proteins that bind to MARE recognition elements. Oligomerization of BACH1–Maf heterodimer via the N-terminus BTB/POZ domain of BACH1 generates multimeric and multivalent DNA binding complexes that are able to mediate interactions between distant, multiple MAREs and to induce DNA loop structures [93].

Critical to the assembly of many enhanceosomes are the high mobility group A proteins, HGMA (previously known as HMG1(Y)). Members of the HGMA family are non-histone, multifunctional proteins that participate in a variety of cellular processes, including the regulation of chromatin structure (reviewed in [192,193]). The striking feature of these proteins is their enormous structural flexibility and their ability to bind to both DNA and a wide range of proteins from different classes, including many TFs. Each HGMA protein contains three repeats of peptides (AT-hooks), which enable them to bind to the minor groove of short stretches of AT-rich DNA and to induce structural alterations in bound DNA, such as looping, bending, and/or unwinding of linear DNA particles. It has been proposed that HGMA proteins recognize gene-specific arrangements of AT-rich sequences within enhancer region and coordinate incorporation of other nuclear factors to the enhanceosome [192]. Direct interactions of HGMA with ATF-2 and c-Jun were reported in the context of the IFN- β gene enhancer. HGMA also interacts with AP-1 (IL-2 promoter), Jun-B/Fra-2 (HPV18 enhancer), C/EBP β (adipocyte-specific gene promoters), and C/EBP α on the IR promoter (Ref. [193] and references therein).

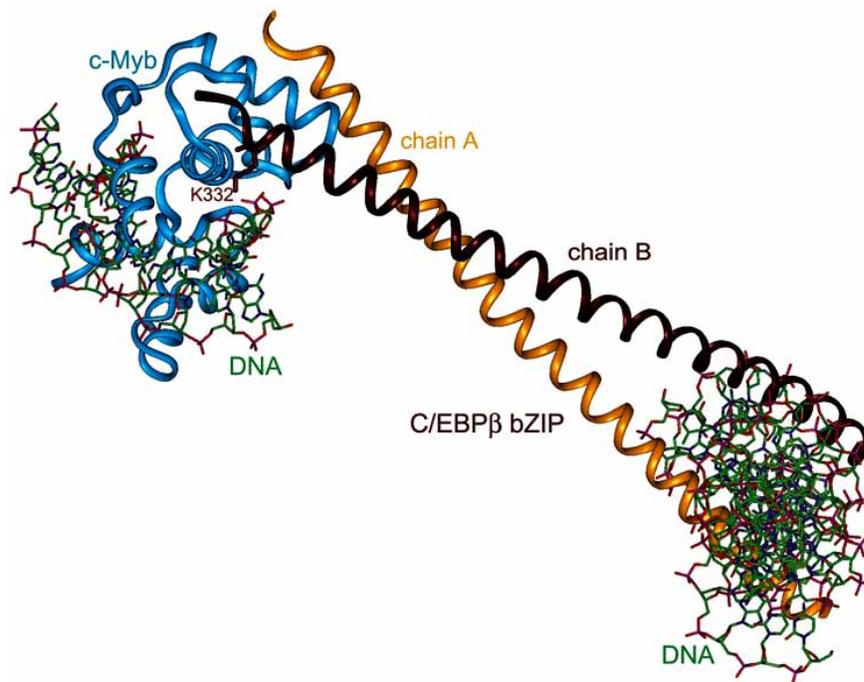


Fig. (10). The intercomplex interactions between c-Myb bound to one DNA fragment and the C-terminal portions of C/EBP β homodimer bound to another DNA fragment observed in the crystal structure of the c-Myb-C/EBP β -DNA complex [188]; PDB code 1H89. The helical fold of C-terminal extensions of C/EBP β LZ is induced by interaction with c-Myb.

Interactions with Coregulators

Enhancer-bound TFs recruit multi-subunit transcriptional coactivator complexes [194], which facilitate the binding and function of Pol II at the core promoter. The interaction of multiple coregulators with TFs in different temporal and spatial contexts provides another possible level of regulation of gene expression. The accessibility of DNA template is modulated by SWI/SNF chromatin remodeling complexes, histone acetyltransferase (HAT) and deacetylase (HDAC) complexes, histone methyltransferases (HMTs) and histone kinases. Recruitment of HATs (e.g., p300 or its ortholog CBP, GCN5/PCAF, MBF1) and HMTs by activators is crucial for activation of many classes of genes, whereas deacetylation of the histone tails is required for repression. Histone lysine methylation is involved in both gene activation and repression, depending upon the specific lysine residue that gets methylated. The possibility of interaction with distinct coregulator complexes underlies the capability of certain TFs to perform tissue and cell type dependent dual functions. Thus C/EBP proteins, which interact with the general coactivator, CBP/p300 and usually act as transcriptional activators, were shown to inhibit the expression of *PPAR β* gene in mouse keratinocytes through the recruitment of a transcriptional repressor complex containing HDAC-1 [195]. The essential role in the regulation of transcription in eukaryotes is played by the multi-subunit Mediator complex, which activates distinct expression programs via interactions with gene-specific TFs [196]. Other examples of factors found in the regulatory complexes are BRCA1 adapter protein [197]

and TORC (see above). These cofactor molecules bridge the sequence-specific DNA-binding proteins to the components of general transcription machinery. On the other hand, a range of TFs binds directly to TBP and its associated factors TAFs [194]. Surprisingly, it was recently found that the terminal kinases of signal transduction cascades which phosphorylates TFs and other components of chromatin, also play a structural role and are stably associated with the genes they activate [198]. As Pokholak *et al.* [199] demonstrated by genome-wide ChIP-Chip analyses in yeast, activated MAPK and PKA kinases can bind at target genes to factors within transcription complexes.

Examples of bZIP proteins' interactions with diverse coregulators are presented in Table 2. The individual interactions are too weak to activate target genes and the stable recruitment of cofactors to the promoter requires cooperation between regulatory TFs [200] and/or multivalent activator-coactivator interactions. For example, synergistic binding of two TADs from Nrf2 to CBP is necessary for induction of reporter gene expression, whereas activation of transcription by CREB requires cooperativity between pKID (which recruits CBP) and the glutamine-rich Q2 domain (which recruits TFIID) as well as coordination with other gene-specific TFs [98,200]. The synergy between TFs is facilitated by the CBP or p300 and other scaffolding proteins that are capable of forming simultaneous interactions with multiple TFs. The p300 protein binds to TAD of c-Myb via KIX domain and to TADs of C/EBP proteins via TAZ2 domain, and further stimulates the synergy between c-Myb and

Table 2. Association of bZIP Proteins with Transcription Coregulator Complexes

Cofactor(domain)	bZIP TF	References
CBP/p300(KIX) CBP/p300(PHD/CH2) CBP/p300(TAZ2/CH3) CBP/p300	CREB,CREM, ATF1, ATF4, c-Jun ATF2 C/EBPs (except - γ), JunB, c-Fos, Nrf2 Par proteins	Revised in [212,213] [214]
SWI/SNF(Brg1)	C/EBP α , C/EBP β	Revised in [215]
Mediator(MED23)	C/EBP β	[196]
HDAC-1 complex	C/EBP β	[195]
TORC	CREB (bZIP)	[125]
MBF1	JunD (bZIP)	[123]
BRCA1(BRCTdomain) BRCA1	BACH1 JunB, JunD	[216] [217]
TFIID (TBP) TFIID (TAF4) TFIID (TAF9) TFIID (TAF1)	C/EBP α , c-Fos, FosB, ATF4 CREB(Q2 domain) C/EBP β c-Jun(bZIP)	[89,150,155] [98] [158] [124]
TFIIB	C/EBP α , ATF4	[155,209]
TFIIF(RAP30)	ATF4	[209]

C/EBPs in expression of *mim-1*. CBP/p300 is also capable to link factors bound to different cognate regulatory modules, and is thought to act as a parser of regulatory information [95].

The cooperative recruitment of TFs, coactivators and chromatin remodeling factors to promoters produces regulatory complexes specific to tissue, cell type and external stimulus. The functions of individual components of such complexes are interdependent and require the concerted action of all the protein in the complex [66,67]. Depending on the cell type and external stimulus, distinct sets of regulatory proteins can assemble at the same promoter. An additional layer of complexity in transcriptional complexes is introduced by enzymatic activities exerted by participating proteins toward each other. HATs such as CBP/p300 and PCAF acetylate variety of TFs (e.g., p53, c-Jun, C/EBP β , mafG, and CREB) as well as HGMA proteins, altering their DNA-binding ability and transactivation potential [201,202]. Reciprocally, CBP/p300 bound C/EBP proteins facilitate its massive phosphorylation [156,203]. Accumulated evidence implies that BRCA1–BARD1 heterodimer that exhibits the E3-ubiquitin ligase activity may specifically ubiquitylate proteins involved in transcription [204]. Another example is the histone methyltransferase CARM1 that methylates CBP/p300 and disrupts its interaction with CREB, thus inactivating CREB-mediated transcription [205].

A recurring theme in the regulation of assembly and function of transcriptional complexes is structural disorder of participating proteins. Results from CD and NMR studies imply that HGMA proteins are almost entirely unfolded in solution and undergo disorder-order structural transitions upon binding to diverse ligands [192]. CBP/p300, which

associates with myriads of TFs and other protein partners, has more than 50% of its residues in predicted intrinsically disordered regions [4]. These regions are located between folded domains and function as flexible linkers. The central two-thirds of BRCA1, which separates the N-terminal RING domain responsible for interaction with BARD1 and the C-terminal BRCT domain, has been predicted to be unstructured [206].

Numerous theories have been proposed to explain the prevalence of unstructured proteins in transcription control. (1) Disordered regions may be considered a sort of “molecular glue”, which is needed to connect together all the components of multiprotein-DNA complexes [29,192]. (2) The ability of IDPs to form transient interactions characterized by high specificity [4,11,55,64] is critical for gene regulatory networks. (3) Rapid clearance and degradation of IDPs provide an additional level of control for turning on and off transcription responses to intracellular signaling [6,13]. (4) Formation of multimolecular complexes and protein–DNA/RNA interactions requires large intermolecular interfaces [207]. ID regions provide for large interfaces without causing cellular crowding or increasing the size of complexes and cells [31]. (5) The “fly-casting” mechanism for binding interactions [41] predicts an increased rate of binding and it may be particularly important in transcriptional processes when the concentrations of regulatory proteins and their targets are low [12]. (6) Conformational adaptability to environment of IDPs enables different modes of regulation [64]. In particular, ID regions are amenable for phosphorylation and other covalent modifications, which modulate activities of TFs as well as the assembly and disassembly of transcription regulatory complexes [4,64]. (7) The existence

of a disordered state prior to specific binding may prevent the occurrence of spontaneous interactions with certain partners at inappropriate times or location [64] (e.g., interaction of MBF1 with BR of DNA bound AP-1 proteins). (8) ID optimizes allosteric coupling, thereby facilitate site-to site communication and signal propagation [46]. Taken together, ID underlies a variety of critical aspects of spatial and temporal organization of interaction networks that govern the gene expression program.

SUMMARY

Eukaryotic TF proteins are central components of dynamic supramolecular complexes that control transcription in a combinatorial manner, and all seem to require a considerable level of structural disorder to perform their functions. This notion is particularly relevant to the bZIP class of activators, which do not possess preformed DBD domains. With the exception of few proteins with specialized functional domains other than TADs, the native structure of a typical bZIP protein is a premolten globule. bZIPs are composed of regions with the potential to fold upon complexation and regions that retain irregular conformations independently of their environment. Conformational flexibility is necessary for the formation of sequential, transient intermolecular interactions that regulate bZIPs' cellular compartmentalization, DNA-binding and transactivating activities and eventually their degradation. bZIP TFs rely on disorder-order transitions of their DBDs for specific dimerization and DNA-binding. Such a mechanism of DNA recognition insures a fast and reliable choice of a specific site and timely binding of coregulators. The ability of bZIP monomers to associate selectively with many different dimerization partners in regulatable manner enables the formation of multiple easily interconverted dimers with distinct DNA and protein binding properties, thus greatly broadening the range of transcriptional control exerted by these factors. Specific dimerization also underlies the dominant repressive capability of naturally occurring truncated variants of these proteins, which retain the bZIP domain, but lack other functional domains. The inherent flexibility of the LZ dimeric coiled-coil domains facilitates physical associations with other factors. The mostly unfolded regions that are responsible for transactivation contain protein recognition motifs and are able to form highly specific but reversible interactions with wide range of protein targets, interactions that can be easily modulated by phosphorylation/dephosphorylation events. Binding diversity and binding commonality with other cofactors are essential for the dynamic and competitive exchange of contacts with multiple components of transcriptional regulatory complexes. Solvent-exposed, unstructured regions serve as flexible linkers between TADs and DNA binding elements. These segments display sites for post-translational modifications, as well as docking domains and recognition motifs for specific modifying enzymes, which regulate the activities of TF proteins in a stimulus-dependent fashion. Structural disorder thus enables multiple modes of regulation of bZIP proteins' activities and girds their ability to effectively control the cellular patterns of gene expression.

USEFUL WEB RESOURCES:

- PFAM: Protein families' database [208]; (<http://www.sanger.ac.uk/Software/Pfam/>)
- DisProt: Database of Protein Disorder [101]; (<http://www.disprot.org>)
- SMART: Simple Modular Architecture Research Tool [102]; (<http://smart.embl.de>)
- bZIPDB: A database of regulatory information for human bZIP transcription factors [87]; (<http://biosoft.kaist.ac.kr/bzipdb>)

ABBREVIATIONS:

- TF = Transcription factor
- ID = Intrinsic disorder
- IDP = Intrinsically disordered protein
- DBD = DNA-binding domain
- BR = Basic region
- LZ = Leucine zipper
- TAD = Transactivation domain
- NLS = Nuclear localization signal

ACKNOWLEDGEMENTS

I thank my colleagues from MCL for help and discussions, and Michael Gribskov for many useful suggestions and thoughtful comments on the manuscript. This work was supported by the Intramural Research Program of NIH, National Cancer Institute, Center for Cancer Research.

REFERENCES

- [1] Dunker, A.K.; Obradovic, Z. The protein trinity—linking function and disorder. *Nat. Biotechnol.*, **2001**, *19*, 805-806.
- [2] Uversky, V.N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.*, **2002**, *11*, 739-756.
- [3] Fukuchi, S.; Homma, K.; Minezaki, Y.; Nishikawa, K. Intrinsically disordered loops inserted into the structural domains of human proteins. *J. Mol. Biol.*, **2006**, *355*, 845-857.
- [4] Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **2005**, *6*, 197-208.
- [5] Dunker, A.K.; Brown, C.J.; Lawson, J.D.; Iakoucheva, L.M.; Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry*, **2002**, *41*, 6573-6582.
- [6] Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, **1999**, *293*, 321-331.
- [7] Radivojac, P.; Iakoucheva, L.M.; Oldfield, C.J.; Obradovic, Z.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder and functional proteomics. *Biophys. J.*, **2007**, *92*, 1439-1456.
- [8] Dyson, H.J.; Wright, P.E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.*, **2004**, *104*, 3607-3622.
- [9] Receveur-Brechot, V.; Bourhis, J.M.; Uversky, V.N.; Canard, B.; Longhi, S. Assessing protein disorder and induced folding. *Proteins*, **2006**, *62*, 24-45.
- [10] Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.*, **2002**, *27*, 527-533.
- [11] Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.*, **2005**, *18*, 343-384.
- [12] Dyson, H.J.; Wright, P.E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 54-60.
- [13] Fink, A.L. Natively unfolded proteins. *Curr. Opin. Struct. Biol.*, **2005**, *15*, 35-41.

- [14] Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.*, **2005**, 579, 3346-3354.
- [15] Dunker, A.K.; Cortese, M.S.; Romero, P.; Iakoucheva, L.M.; Uversky, V.N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.*, **2005**, 272, 5129-5148.
- [16] Vucetic, S.; Brown, C.J.; Dunker, A.K.; Obradovic, Z. Flavors of protein disorder. *Proteins*, **2003**, 52, 573-584.
- [17] Lobley, A.; Swindells, M.B.; Orengo, C.A.; Jones, D.T. Inferring function using patterns of native disorder in proteins. *PLoS Comput. Biol.*, **2007**, 3, e162.
- [18] Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **2000**, 41, 415-427.
- [19] Prilusky, J.; Felder, C.E.; Zeev-Ben-Mordehai, T.; Rydberg, E.H.; Man, O.; Beckmann, J.S.; Silman, I.; Sussman, J.L. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics.*, **2005**, 21, 3435-3438.
- [20] Coeytaux, K.; Poupon, A. Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics.*, **2005**, 21, 1891-1900.
- [21] Liu, J.; Rost, B. NORSp: Predictions of long regions without regular secondary structure. *Nucleic Acids Res.*, **2003**, 31, 3833-3835.
- [22] Linding, R.; Russell, R.B.; Neduva, V.; Gibson, T.J. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.*, **2003**, 31, 3701-3708.
- [23] Ferron, F.; Longhi, S.; Canard, B.; Karlin, D. A practical overview of protein disorder prediction methods. *Proteins*, **2006**, 65, 1-14.
- [24] Bourhis, J.M.; Canard, B.; Longhi, S. Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Curr. Protein Pept. Sci.*, **2007**, 8, 135-149.
- [25] Oldfield, C.J.; Cheng, Y.; Cortese, M.S.; Brown, C.J.; Uversky, V.N.; Dunker, A.K. Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, **2005**, 44, 1989-2000.
- [26] Liu, J.; Rost, B. Comparing function and structure between entire proteomes. *Protein Sci.*, **2001**, 10, 1970-1979.
- [27] Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **2004**, 337, 635-645.
- [28] Chen, J.W.; Romero, P.; Uversky, V.N.; Dunker, A.K. Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J. Proteome. Res.*, **2006**, 5, 888-898.
- [29] Minezaki, Y.; Homma, K.; Kinjo, A.R.; Nishikawa, K. Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation. *J. Mol. Biol.*, **2006**, 359, 1137-1149.
- [30] Romero, P.R.; Zaidi, S.; Fang, Y.Y.; Uversky, V.N.; Radivojac, P.; Oldfield, C.J.; Cortese, M.S.; Sickmeier, M.; LeGall, T.; Obradovic, Z.; Dunker, A.K. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. USA*, **2006**, 103, 8390-8395.
- [31] Gunasekaran, K.; Tsai, C.J.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem. Sci.*, **2003**, 28, 81-85.
- [32] Oldfield, C.J.; Cheng, Y.; Cortese, M.S.; Romero, P.; Uversky, V.N.; Dunker, A.K. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry*, **2005**, 44, 12454-12470.
- [33] Sugase, K.; Dyson, H.J.; Wright, P.E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, **2007**, 447, 1021-1025.
- [34] Verkhivker, G.M.; Bouzida, D.; Gehlhaar, D.K.; Rejto, P.A.; Freer, S.T.; Rose, P.W. Simulating disorder-order transitions in molecular recognition of unstructured proteins: where folding meets binding. *Proc. Natl. Acad. Sci. USA*, **2003**, 100, 5148-5153.
- [35] Wang, J.; Verkhivker, G.M. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys. Rev. Lett.*, **2003**, 90, 188101.
- [36] Levy, Y.; Cho, S.S.; Onuchic, J.N.; Wolynes, P.G. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.*, **2005**, 346, 1121-1145.
- [37] Wang, J.; Lu, Q.; Lu, H.P. Single-molecule dynamics reveals cooperative binding-folding in protein recognition. *PLoS. Comput. Biol.*, **2006**, 2, e78.
- [38] Wang, J.; Xu, L.; Wang, E. Optimal specificity and function for flexible biomolecular recognition. *Biophys. J.*, **2007**, 92, L109-L111.
- [39] Lu, Q.; Lu, H.P.; Wang, J. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. *Phys. Rev. Lett.*, **2007**, 98, 128105.
- [40] Demchenko, A.P. Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.*, **2001**, 14, 42-61.
- [41] Shoemaker, B.A.; Portman, J.J.; Wolynes, P.G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA*, **2000**, 97, 8868-8873.
- [42] Sturtevant, J.M. Heat capacity and entropy changes in processes involving proteins. *Proc. Natl. Acad. Sci. USA*, **1977**, 74, 2236-2240.
- [43] Duan, Y.; Wilkosz, P.; Rosenberg, J.M. Dynamic contributions to the DNA binding entropy of the EcoRI and EcoRV restriction endonucleases. *J. Mol. Biol.*, **1996**, 264, 546-555.
- [44] Spolar, R.S.; Record, M.T., Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **1994**, 263, 777-784.
- [45] Dunitz, J.D. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.*, **1995**, 2, 709-712.
- [46] Hilser, V.J.; Thompson, E.B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. USA*, **2007**, 104, 8311-8315.
- [47] Mohan, A.; Oldfield, C.J.; Radivojac, P.; Vacic, V.; Cortese, M.S.; Dunker, A.K.; Uversky, V.N. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.*, **2006**, 362, 1043-1059.
- [48] Puntervoll, P.; Linding, R.; Gemünd, C.; Chabani-Davidson, S.; Mattingdal, M.; Cameron, S.; Martin, D.M.A.; Ausiello, G.; Brannetti, B.; Costantini, A.; Ferre, F.; Maselli, V.; Via, A.; Cesareni, G.; Diella, F.; Superti-Furga, G.; Wyrwicz, L.; Ramu, C.; McGuigan, C.; Gudavalli, R.; Letunic, I.; Bork, P.; Rychlewski, L.; Küster, B.; Helmer-Citterich, M.; Hunter, W.N.; Aasland, R.; Gibson, T.J. ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.*, **2003**, 31, 3625-3630.
- [49] Neduva, V.; Linding, R.; Su-Angrand, I.; Stark, A.; de Masi, F.; Gibson, T.J.; Lewis, J.; Serrano, L.; Russell, R.B. Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS. Biol.*, **2005**, 3, e405.
- [50] Apic, G.; Russell, R.B. A shortcut to peptides to modulate platelets. *Nat. Chem. Biol.*, **2007**, 3, 83-84.
- [51] Csizmek, V.; Bokor, M.; Banki, P.; Klement, E.; Medzihradzsky, K.F.; Friedrich, P.; Tompa, K.; Tompa, P. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry*, **2005**, 44, 3955-3964.
- [52] Namba, K. Roles of partly unfolded conformations in macromolecular self-assembly. *Genes Cells*, **2001**, 6, 1-12.
- [53] Blundell, T.L.; Fernandez-Recio, J. Cell biology: brief encounters bolster contacts. *Nature*, **2006**, 444, 279-280.
- [54] Han, J.D.; Bertin, N.; Hao, T.; Goldberg, D.S.; Berriz, G.F.; Zhang, L.V.; Dupuy, D.; Walhout, A.J.; Cusick, M.E.; Roth, F.P.; Vidal, M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **2004**, 430, 88-93.
- [55] Singh, G.P.; Ganapathi, M.; Dash, D. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins: Struct. Funct. Bioinformatics*, **2007**, 66, 761-765.
- [56] Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.*, **2004**, 32, 1037-1049.
- [57] Eisenhaber, B.; Eisenhaber, F. Posttranslational modifications and subcellular localization signals: indicators of sequence regions without inherent 3D structure? *Curr. Protein Pept. Sci.*, **2007**, 8, 197-203.
- [58] Ptashne, M.; Gann, A. Transcriptional activation by recruitment. *Nature*, **1997**, 386, 569-577.
- [59] Sigler, P.B. Transcriptional activation. Acid blobs and negative noodles. *Nature*, **1988**, 333, 210-212.
- [60] Frankel, A.D.; Kim, P.S. Modular structure of transcription factors: implications for gene regulation. *Cell*, **1991**, 65, 717-719.
- [61] Uesugi, M.; Nyanguile, O.; Lu, H.; Levine, A.J.; Verdine, G.L. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science*, **1997**, 277, 1310-1313.

- [62] Ayed, A.; Mulder, F.A.; Yi, G.S.; Lu, Y.; Kay, L.E.; Arrowsmith, C.H. Latent and active p53 are identical in conformation. *Nat. Struct. Biol.*, **2001**, *8*, 756-760.
- [63] Grossmann, J.G.; Sharff, A.J.; O'Hare, P.; Luisi, B. Molecular shapes of transcription factors TFIIB and VP16 in solution: implications for recognition. *Biochemistry*, **2001**, *40*, 6267-6274.
- [64] Liu, J.; Perumal, N.B.; Oldfield, C.J.; Su, E.W.; Uversky, V.N.; Dunker, A.K. Intrinsic disorder in transcription factors. *Biochemistry*, **2006**, *45*, 6873-6888.
- [65] Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*, **1999**, *98*, 1-4.
- [66] Levine, M.; Tjian, R. Transcription regulation and animal diversity. *Nature*, **2003**, *424*, 147-151.
- [67] Michelson, A.M. Deciphering genetic regulatory codes: a challenge for functional genomics. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 546-548.
- [68] Amoutzias, G.; Veron, A.; Weiner, J.; III, Robinson-Rechavi, M.; Bornberg-Bauer, E.; Oliver, S.; Robertson, D. One Billion Years of bZIP Transcription Factor Evolution: Conservation and Change in Dimerization and DNA-Binding Site Specificity. *Mol. Biol. Evol.*, **2007**, *24*, 827-835.
- [69] Hurst, H.C. Transcription factors 1: bZIP proteins. *Protein Profile*, **1995**, *2*, 101-168.
- [70] Chinenov, Y.; Kerppola, T.K. Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene*, **2001**, *20*, 2438-2452.
- [71] Landschulz, W.H.; Johnson, P.F.; Mcknight, S.L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **1988**, *240*, 1759-1764.
- [72] Landschulz, W.H.; Johnson, P.F.; Mcknight, S.L. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science*, **1989**, *243*, 1681-1688.
- [73] Ellenberger, T.E.; Brandl, C.J.; Struhl, K.; Harrison, S.C. The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell*, **1992**, *71*, 1223-1237.
- [74] Glover, J.N.; Harrison, S.C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature*, **1995**, *373*, 257-261.
- [75] Fujii, Y.; Shimizu, T.; Toda, T.; Yanagida, M.; Hakoshima, T. Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat. Struct. Biol.*, **2000**, *7*, 889-893.
- [76] Schumacher, M.A.; Goodman, R.H.; Brennan, R.G. The structure of a CREB bZIP-somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J. Biol. Chem.*, **2000**, *275*, 35242-35247.
- [77] Miller, M.; Shuman, J.D.; Sebastian, T.; Dauter, Z.; Johnson, P.F. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J. Biol. Chem.*, **2003**, *278*, 15178-15184.
- [78] Wolberger, C. Transcription factor structure and DNA binding. *Curr. Opin. Struct. Biol.*, **1993**, *3*, 3-10.
- [79] Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*, **1994**, *265*, 346-355.
- [80] Weiss, M.A.; Ellenberger, T.; Wobbe, C.R.; Lee, J.P.; Harrison, S.C.; Struhl, K. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, **1990**, *347*, 575-578.
- [81] Krylov, D.; Olive, M.; Vinson, C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J.*, **1995**, *14*, 5329-5337.
- [82] Patel, L.R.; Curran, T.; Kerppola, T.K. Energy transfer analysis of Fos-Jun dimerization and DNA binding. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 7360-7364.
- [83] Garvie, C.W.; Wolberger, C. Recognition of specific DNA sequences. *Mol. Cell*, **2001**, *8*, 937-946.
- [84] Vinson, C.R.; Hai, T.; Boyd, S.M. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev.*, **1993**, *7*, 1047-1058.
- [85] Vinson, C.; Myakishev, M.; Acharya, A.; Mir, A.A.; Moll, J.R.; Bonovich, M. Classification of human B-ZIP proteins based on dimerization properties. *Mol. Cell Biol.*, **2002**, *22*, 6321-6335.
- [86] Fassler, J.; Landsman, D.; Acharya, A.; Moll, J.R.; Bonovich, M.; Vinson, C. B-ZIP proteins encoded by the *Drosophila* genome: evaluation of potential dimerization partners. *Genome Res.*, **2002**, *12*, 1190-1200.
- [87] Ryu, T.; Jung, J.; Lee, S.; Nam, H.J.; Hong, S.W.; Yoo, J.W.; Lee, D.K.; Lee, D. bZIPDB: a database of regulatory information for human bZIP transcription factors. *BMC. Genomics*, **2007**, *8*, 136
- [88] Blank, V.; Andrews, N.C. The Maf transcription factors: regulators of differentiation. *Trends Biochem. Sci.*, **1997**, *22*, 437-441.
- [89] Motohashi, H.; O'Connor, T.; Katsuoka, F.; Engel, J.D.; Yamamoto, M. Integration and diversity of the regulatory network composed of Maf and CNC families of transcription factors. *Gene*, **2002**, *294*, 1-12.
- [90] Richards, J.P.; Bachinger, H.P.; Goodman, R.H.; Brennan, R.G. Analysis of the structural properties of cAMP-responsive element-binding protein (CREB) and phosphorylated CREB. *J. Biol. Chem.*, **1996**, *271*, 13716-13723.
- [91] Nagadoi, A.; Nakazawa, K.; Uda, H.; Okuno, K.; Maekawa, T.; Ishii, S.; Nishimura, Y. Solution structure of the transactivation domain of ATF-2 comprising a zinc finger-like subdomain and a flexible subdomain. *J. Mol. Biol.*, **1999**, *287*, 593-607.
- [92] Kusunoki, H.; Motohashi, H.; Katsuoka, F.; Motohashi, A.; Yamamoto, M.; Tanaka, T. Solution structure of the DNA-binding domain of MafG. *Nat. Struct. Biol.*, **2002**, *9*, 252-256.
- [93] Yoshida, C.; Tokumasu, F.; Hohmura, K.I.; Bungert, J.; Hayashi, N.; Nagasawa, T.; Engel, J.D.; Yamamoto, M.; Takeyasu, K.; Igarashi, K. Long range interaction of cis-DNA elements mediated by architectural transcription factor Bach1. *Genes Cells*, **1999**, *4*, 643-655.
- [94] Williamson, E.A.; Xu, H.N.; Gombart, A.F.; Verbeek, W.; Chumakov, A.M.; Friedman, A.D.; Koeffler, H.P. Identification of transcriptional activation and repression domains in human CCAAT/enhancer-binding protein epsilon. *J. Biol. Chem.*, **1998**, *273*, 14796-14804.
- [95] Smith, J.L.; Freebern, W.J.; Collins, I.; De Siervi, A.; Montano, I.; Haggerty, C.M.; McNutt, M.C.; Butscher, W.G.; Dzekunova, I.; Petersen, D.W.; Kawasaki, E.; Merchant, J.L.; Gardner, K. Kinetic profiles of p300 occupancy in vivo predict common features of promoter structure and coactivator recruitment. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 11554-11559.
- [96] Wagner, E.F. AP-1--Introductory remarks. *Oncogene*, **2001**, *20*, 2334-2335.
- [97] Yoon, M.K.; Shin, J.; Choi, G.; Choi, B.S. Intrinsically unstructured N-terminal domain of bZIP transcription factor HY5. *Proteins*, **2006**, *65*, 856-866.
- [98] Sharma, N.; Lopez, D.I.; Nyborg, J.K. DNA binding and phosphorylation induce conformational alterations in the kinase-inducible domain of CREB. Implications for the mechanism of transcription function. *J. Biol. Chem.*, **2007**, *282*, 19872-19883.
- [99] Romero, P.; Obradovic, Z.; Li, X.; Garner, E.C.; Brown, C.J.; Dunker, A.K. Sequence complexity of disordered protein. *Proteins*, **2001**, *42*, 38-48.
- [100] Miller, M. Phospho-Dependent Protein Recognition Motifs Contained in C/EBP Family of Transcription Factors: in Silico Studies. *Cell Cycle*, **2006**, *5*, 2501-2508.
- [101] Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L.M.; Cortese, M.S.; Lawson, J.D.; Brown, C.J.; Sikes, J.G.; Newton, C.D.; Dunker, A.K. DisProt: a database of protein disorder. *Bioinformatics*, **2005**, *21*, 137-140.
- [102] Letunic, I.; Goodstadt, L.; Dickens, N.J.; Doerks, T.; Schultz, J.; Mott, R.; Ciccarelli, F.; Copley, R.R.; Ponting, C.P.; Bork, P. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.*, **2002**, *30*, 242-244.
- [103] Podust, L.M.; Krezel, A.M.; Kim, Y. Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J. Biol. Chem.*, **2001**, *276*, 505-513.
- [104] Thompson, K.S.; Vinson, C.R.; Freire, E. Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. *Biochemistry*, **1993**, *32*, 5491-5496.
- [105] Krylov, D.; Mikhailenko, I.; Vinson, C. A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J.*, **1994**, *13*, 2849-2861.
- [106] Sosnick, T.R.; Jackson, S.; Wilk, R.R.; Englander, S.W.; DeGrado, W.F. The role of helix formation in the folding of a fully alpha-helical coiled coil. *Proteins*, **1996**, *24*, 427-432.
- [107] d'Avignon, D.A.; Bretthorst, G.L.; Holtzer, M.E.; Holtzer, A. Site-specific thermodynamics and kinetics of a coiled-coil transition by spin inversion transfer NMR. *Biophys. J.*, **1998**, *74*, 3190-3197.

- [108] Dragan, A.I.; Privalov, P.L. Unfolding of a leucine zipper is not a simple two-state transition. *J. Mol. Biol.*, **2002**, *321*, 891-908.
- [109] Zitzewitz, J.A.; Ibarra-Molero, B.; Fishel, D.R.; Terry, K.L.; Matthews, C.R. Prefolded secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.*, **2000**, *296*, 1105-1116.
- [110] Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. Prefolded structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.*, **2004**, *338*, 1015-1026.
- [111] Lamb, P.; Mcknight, S.L. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.*, **1991**, *16*, 417-422.
- [112] Vinson, C.; Acharya, A.; Taparowsky, E.J. Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochim. Biophys. Acta.*, **2006**, *1759*, 4-12.
- [113] Alber, T. Structure of the leucine zipper. *Curr. Opin. Genet. Dev.*, **1992**, *2*, 205-210.
- [114] Grigoryan, G.; Keating, A.E. Structure-based prediction of bZIP partnering specificity. *J. Mol. Biol.*, **2006**, *355*, 1125-1142.
- [115] Fong, J.H.; Keating, A.E.; Singh, M. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol.*, **2004**, *5*, R11
- [116] Harbury, P.B.; Zhang, T.; Kim, P.S.; Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, **1993**, *262*, 1401-1407.
- [117] Newman, J.R.; Keating, A.E. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*, **2003**, *300*, 2097-2101.
- [118] Deppmann, C.D.; Acharya, A.; Rishi, V.; Wobbes, B.; Smeekens, S.; Taparowsky, E.J.; Vinson, C. Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to Homo sapiens B-ZIP motifs. *Nucleic Acids Res.*, **2004**, *32*, 3435-3445.
- [119] Parkin, S.E.; Baer, M.; Copeland, T.D.; Schwartz, R.C.; Johnson, P.F. Regulation of CCAAT/enhancer-binding protein (C/EBP) activator proteins by heterodimerization with C/EBPgamma (I γ /EBP). *J. Biol. Chem.*, **2002**, *277*, 23563-23572.
- [120] O'Neil, K.T.; Shuman, J.D.; Ampe, C.; DeGrado, W.F. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. *Biochemistry*, **1991**, *30*, 9030-9034.
- [121] Bracken, C.; Carr, P.A.; Cavanagh, J.; Palmer, A.G., III Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.*, **1999**, *285*, 2133-2146.
- [122] Hollenbeck, J.J.; McClain, D.L.; Oakley, M.G. The role of helix stabilizing residues in GCN4 basic region folding and DNA binding. *Protein Sci.*, **2002**, *11*, 2740-2747.
- [123] Kabe, Y.; Goto, M.; Shima, D.; Imai, T.; Wada, T.; Morohashi, K.; Shirakawa, M.; Hirose, S.; Handa, H. The role of human MBF1 as a transcriptional coactivator. *J. Biol. Chem.*, **1999**, *274*, 34196-34202.
- [124] Lively, T.N.; Nguyen, T.N.; Galasinski, S.K.; Goodrich, J.A. The basic leucine zipper domain of c-Jun functions in transcriptional activation through interaction with the N terminus of human TATA-binding protein-associated factor-1 (human TAF(II)250). *J. Biol. Chem.*, **2004**, *279*, 26257-26265.
- [125] Konkright, M.D.; Canettieri, G.; Sreanor, R.; Guzman, E.; Miraglia, L.; Hogenesch, J.B.; Montminy, M. TORCs: transducers of regulated CREB activity. *Mol. Cell*, **2003**, *12*, 413-423.
- [126] Baranger, A.M. Accessory factor-bZIP-DNA interactions. *Curr. Opin. Chem. Biol.*, **1998**, *2*, 18-23.
- [127] Miotto, B.; Struhl, K. Differential gene regulation by selective association of transcriptional coactivators and bZIP DNA-binding domains. *Mol. Cell Biol.*, **2006**, *26*, 5969-5982.
- [128] Williams, S.C.; Angerer, N.D.; Johnson, P.F. C/EBP proteins contain nuclear localization signals imbedded in their basic regions. *Gene Expr.*, **1997**, *6*, 371-385.
- [129] Weis, K. Importins and exportins: how to get in and out of the nucleus. *Trends Biochem. Sci.*, **1998**, *23*, 185-189.
- [130] Conti, E.; Izaurralde, E. Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.*, **2001**, *13*, 310-319.
- [131] Fontes, M.R.; Teh, T.; Kobe, B. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J. Mol. Biol.*, **2000**, *297*, 1183-1194.
- [132] Harreman, M.T.; Kline, T.M.; Milford, H.G.; Harben, M.B.; Hodel, A.E.; Corbett, A.H. Regulation of nuclear import by phosphorylation adjacent to nuclear localization signals. *J. Biol. Chem.*, **2004**, *279*, 20613-20621.
- [133] Buck, M.; Zhang, L.; Halasz, N.A.; Hunter, T.; Chojkier, M. Nuclear export of phosphorylated C/EBPbeta mediates the inhibition of albumin expression by TNF-alpha. *EMBO J.*, **2001**, *20*, 6712-6723.
- [134] Bannister, A.J.; Cook, A.; Kouzarides, T. In vitro DNA binding activity of Fos/Jun and BZLF1 but not C/EBP is affected by redox changes. *Oncogene*, **1991**, *6*, 1243-1250.
- [135] Jindra, M.; Gaziova, I.; Uhlirova, M.; Okabe, M.; Hiromi, Y.; Hirose, S. Coactivator MBF1 preserves the redox-dependent AP-1 activity during oxidative stress in *Drosophila*. *EMBO J.*, **2004**, *23*, 3538-3547.
- [136] Metallo, S.J.; Schepartz, A. Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA. *Nat. Struct. Biol.*, **1997**, *4*, 115-117.
- [137] Wu, X.; Spiro, C.; Owen, W.G.; McMurray, C.T. cAMP response element-binding protein monomers cooperatively assemble to form dimers on DNA. *J. Biol. Chem.*, **1998**, *273*, 20820-20827.
- [138] Pomerantz, J.L.; Wolfe, S.A.; Pabo, C.O. Structure-based design of a dimeric zinc finger protein. *Biochemistry*, **1998**, *37*, 965-970.
- [139] Kohler, J.J.; Schepartz, A. Kinetic studies of Fos.Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry*, **2001**, *40*, 130-142.
- [140] Kohler, J.J.; Metallo, S.J.; Schneider, T.L.; Schepartz, A. DNA specificity enhanced by sequential binding of protein monomers. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 11735-11739.
- [141] Berger, C.; Piubelli, L.; Haditsch, U.; Bosshard, H.R. Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor. *FEBS Lett.*, **1998**, *425*, 14-18.
- [142] Kim, B.; Little, J.W. Dimerization of a specific DNA-binding protein on the DNA. *Science*, **1992**, *255*, 203-206.
- [143] Williams, S.C.; Baer, M.; Dillner, A.J.; Johnson, P.F. CRP2 (C/EBP beta) contains a bipartite regulatory domain that controls transcriptional activation, DNA binding and cell specificity. *EMBO J.*, **1995**, *14*, 3170-3183.
- [144] Hattori, T.; Ohoka, N.; Inoue, Y.; Hayashi, H.; Onozaki, K. C/EBP family transcription factors are degraded by the proteasome but stabilized by forming dimer. *Oncogene*, **2003**, *22*, 1273-1280.
- [145] Lee, S.J.; Sekimoto, T.; Yamashita, E.; Nagoshi, E.; Nakagawa, A.; Imamoto, N.; Yoshimura, M.; Sakai, H.; Chong, K.T.; Tsukihara, T.; Yoneda, Y. The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor. *Science*, **2003**, *302*, 1571-1575.
- [146] Johnson, P.F.; Sterneck, E.; Williams, S.C. Activation domains of transcriptional regulatory proteins. *J. Nutr. Biochem.*, **1993**, *4*, 386-398.
- [147] Triezenberg, S.J. Structure and function of transcriptional activation domains. *Curr. Opin. Genet. Dev.*, **1995**, *5*, 190-196.
- [148] Gill, G.; Ptashne, M. Negative effect of the transcriptional activator GAL4. *Nature*, **1988**, *334*, 721-724.
- [149] O'Hare, P.; Williams, G. Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using 1H NMR. *Biochemistry*, **1992**, *31*, 4150-4156.
- [150] Campbell, K.M.; Terrell, A.R.; Laybourn, P.J.; Lumb, K.J. Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. *Biochemistry*, **2000**, *39*, 2708-2713.
- [151] Fladvad, M.; Zhou, K.; Moshref, A.; Pursglove, S.; Safsten, P.; Sunnerhagen, M. N and C-terminal sub-regions in the c-Myc transactivation region and their joint role in creating versatility in folding and binding. *J. Mol. Biol.*, **2005**, *346*, 175-189.
- [152] Metz, R.; Bannister, A.J.; Sutherland, J.A.; Hagemeyer, C.; O'Rourke, E.C.; Cook, A.; Bravo, R.; Kouzarides, T. c-Fos-induced activation of a TATA-box-containing promoter involves direct contact with TATA-box-binding protein. *Mol. Cell Biol.*, **1994**, *14*, 6021-6029.
- [153] Hermann, S.; Berndt, K.D.; Wright, A.P. How transcriptional activators bind target proteins. *J. Biol. Chem.*, **2001**, *276*, 40127-40132.
- [154] Schmitz, M.L.; dos Santos Silva, M.A.; Altmann, H.; Czisch, M.; Holak, T.A.; Baeuerle, P.A. Structural and functional analysis of the NF-kappa B p65 C terminus. An acidic and modular transactivation domain with the potential to adopt an alpha-helical conformation. *J. Biol. Chem.*, **1994**, *269*, 25613-25620.
- [155] Nerlov, C.; Ziff, E.B. CCAAT/enhancer binding protein-alpha amino acid motifs with dual TBP and TFIIB binding ability cooperate to activate transcription in both yeast and mammalian cells. *EMBO J.*, **1995**, *14*, 4318-4328.

- [156] Kovacs, K.A.; Steinmann, M.; Magistretti, P.J.; Halfon, O.; Cardinaux, J.R. CCAAT/enhancer-binding protein family members recruit the coactivator CREB-binding protein and trigger its phosphorylation. *J. Biol. Chem.*, **2003**, *278*, 36959-36965.
- [157] Kussie, P.H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A.J.; Pavletich, N.P. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, **1996**, *274*, 948-953.
- [158] Choi, Y.; Asada, S.; Uesugi, M. Divergent hTAFII31-binding motifs hidden in activation domains. *J. Biol. Chem.*, **2000**, *275*, 15912-15916.
- [159] Anderson, C.W.; Appella, E. In *The p53 Tumor Suppressor Pathway and Cancer*; Zanetti, G.P., Ed.; Posttranslational modifications of p53: Upstream signaling. Springer Science+Business Media, Inc., New York, **2005**, 95-114.
- [160] Bannister, A.J.; Brown, H.J.; Sutherland, J.A.; Kouzarides, T. Phosphorylation of the c-Fos and c-Jun HOB1 motif stimulates its activation capacity. *Nucleic Acids Res.*, **1994**, *22*, 5173-5176.
- [161] Du, K.; Montminy, M. CREB is a regulatory target for the protein kinase Akt/PKB. *J. Biol. Chem.*, **1998**, *273*, 32377-32379.
- [162] Shanware, N.P.; Trinh, A.T.; Williams, L.M.; Tibbetts, R.S. Coregulated ataxia telangiectasia-mutated and casein kinase sites modulate cAMP-response element-binding protein-coactivator interactions in response to DNA damage. *J. Biol. Chem.*, **2007**, *282*, 6283-6291.
- [163] Shuman, J.D.; Sebastian, T.; Kaldis, P.; Copeland, T.D.; Zhu, S.; Smart, R.C.; Johnson, P.F. Cell cycle-dependent phosphorylation of C/EBP β mediates oncogenic cooperativity between C/EBP β and H-RasV12. *Mol. Cell Biol.*, **2004**, *24*, 7380-7391.
- [164] Monje, P.; Hernandez-Losa, J.; Lyons, R.J.; Castellone, M.D.; Gutkind, J.S. Regulation of the transcriptional activity of c-Fos by ERK. A novel role for the prolyl isomerase PIN1. *J. Biol. Chem.*, **2005**, *280*, 35081-35084.
- [165] Manke, I.A.; Lowery, D.M.; Nguyen, A.; Yaffe, M.B. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*, **2003**, *302*, 636-639.
- [166] Gardner, K.H.; Montminy, M. Can you hear me now? Regulating transcriptional activators by phosphorylation. *Sci. STKE*, **2005**, *e44*
- [167] Kwong, P.D.; Wyatt, R.; Majeed, S.; Robinson, J.; Sweet, R.W.; Sodroski, J.; Hendrickson, W.A. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure Fold Des.*, **2000**, *8*, 1329-1339.
- [168] Freedman, S.J.; Sun, Z.Y.; Poy, F.; Kung, A.L.; Livingston, D.M.; Wagner, G.; Eck, M.J. Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 5367-5372.
- [169] Campbell, K.M.; Lumb, K.J. Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry*, **2002**, *41*, 13956-13964.
- [170] De Guzman, R.N.; Goto, N.K.; Dyson, H.J.; Wright, P.E. Structural basis for cooperative transcription factor binding to the CBP coactivator. *J. Mol. Biol.*, **2006**, *355*, 1005-1013.
- [171] Radhakrishnan, I.; Perez-Alvarado, G.C.; Parker, D.; Dyson, H.J.; Montminy, M.R.; Wright, P.E. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell*, **1997**, *91*, 741-752.
- [172] Zor, T.; De Guzman, R.N.; Dyson, H.J.; Wright, P.E. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.*, **2004**, *337*, 521-534.
- [173] Roach, P.J. Multisite and hierarchical protein phosphorylation. *J. Biol. Chem.*, **1991**, *266*, 14139-14142.
- [174] Ilouz, R.; Kowalsman, N.; Eisenstein, M.; Eldar-Finkelman, H. Identification of novel glycogen synthase kinase-3 β substrate-interacting residues suggests a common mechanism for substrate recognition. *J. Biol. Chem.*, **2006**, *281*, 30621-30630.
- [175] Kim, J.; Sharma, S.; Li, Y.; Cobos, E.; Palvimo, J.J.; Williams, S.C. Repression and coactivation of CCAAT/enhancer-binding protein epsilon by sumoylation and protein inhibitor of activated STATx proteins. *J. Biol. Chem.*, **2005**, *280*, 12246-12254.
- [176] Lise, S.; Jones, D.T. Sequence patterns associated with disordered regions in proteins. *Proteins*, **2005**, *58*, 144-150.
- [177] Fontana, A.; Polverino, d.L.; De, F.; V; Scaramella, E.; Zambonin, M. Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.*, **1997**, *2*, R17-R26
- [178] Singh, G.P.; Ganapathi, M.; Sandhu, K.S.; Dash, D. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins*, **2006**, *62*, 309-315.
- [179] Hill, C.S.; Treisman, R. Transcriptional regulation by extracellular signals: mechanisms and specificity. *Cell*, **1995**, *80*, 199-211.
- [180] Ciechanover, A.; DiGiuseppe, J.A.; Bercovich, B.; Orian, A.; Richter, J.D.; Schwartz, A.L.; Brodeur, G.M. Degradation of nuclear oncoproteins by the ubiquitin system in vitro. *Proc. Natl. Acad. Sci. USA*, **1991**, *88*, 139-143.
- [181] Ramji, D.P.; Foka, P. CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochem. J.*, **2002**, *365*, 561-575.
- [182] Sharrocks, A.D.; Yang, S.H.; Galanis, A. Docking domains and substrate-specificity determination for MAP kinases. *Trends Biochem. Sci.*, **2000**, *25*, 448-453.
- [183] Narlikar, G.J.; Fan, H.Y.; Kingston, R.E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell*, **2002**, *108*, 475-487.
- [184] Orphanides, G.; Reinberg, D. A unified theory of gene expression. *Cell*, **2002**, *108*, 439-451.
- [185] Merika, M.; Thanos, D. Enhanceosomes. *Curr. Opin. Genet. Dev.*, **2001**, *11*, 205-208.
- [186] Chen, L.; Glover, J.N.; Hogan, P.G.; Rao, A.; Harrison, S.C. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature*, **1998**, *392*, 42-48.
- [187] Panne, D.; Maniatis, T.; Harrison, S.C. Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J.*, **2004**, *23*, 4384-4393.
- [188] Tahirov, T.H.; Sato, K.; Ichikawa-Iwata, E.; Sasaki, M.; Inoue-Bungo, T.; Shiina, M.; Kimura, K.; Takata, S.; Fujikawa, A.; Morii, H.; Kumasaka, T.; Yamamoto, M.; Ishii, S.; Ogata, K. Mechanism of c-Myb-C/EBP β cooperation from separated sites on a promoter. *Cell*, **2002**, *108*, 57-70.
- [189] Cha-Molstad, H.; Young, D.P.; Kushner, I.; Samols, D. The interaction of C-Rel with C/EBP β enhances C/EBP β binding to the C-reactive protein gene promoter. *Mol. Immunol.*, **2007**, *44*, 2933-2942.
- [190] Wolberger, C. Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.*, **1999**, *28*, 29-56.
- [191] Ogata, K.; Sato, K.; Tahirov, T.H. Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. *Curr. Opin. Struct. Biol.*, **2003**, *13*, 40-48.
- [192] Reeves, R.; Beckerbauer, L. HMG1/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim. Biophys. Acta.*, **2001**, *1519*, 13-29.
- [193] Sgarra, R.; Tessari, M.A.; Di Bernardo, J.; Rustighi, A.; Zago, P.; Liberatori, S.; Armini, A.; Bini, L.; Giancotti, V.; Manfioletti, G. Discovering high mobility group A molecular partners in tumour cells. *Proteomics*, **2005**, *5*, 1494-1506.
- [194] Naar, A.M.; Lemon, B.D.; Tjian, R. Transcriptional coactivator complexes. *Annu. Rev. Biochem.*, **2001**, *70*, 475-501.
- [195] Di Poi, N.; Desvergne, B.; Michalik, L.; Wahli, W. Transcriptional repression of peroxisome proliferator-activated receptor beta/delta in murine keratinocytes by CCAAT/enhancer-binding proteins. *J. Biol. Chem.*, **2005**, *280*, 38700-38710.
- [196] Conaway, R.C.; Sato, S.; Tomomori-Sato, C.; Yao, T.; Conaway, J.W. The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.*, **2005**, *30*, 250-255.
- [197] Starita, L.M.; Parvin, J.D. The multiple nuclear functions of BRCA1: transcription, ubiquitination and DNA repair. *Curr. Opin. Cell Biol.*, **2003**, *15*, 345-350.
- [198] Edmunds, J.W.; Mahadevan, L.C. Cell signaling. Protein kinases seek close encounters with active genes. *Science*, **2006**, *313*, 449-451.
- [199] Pokholok, D.K.; Zeitlinger, J.; Hannett, N.M.; Reynolds, D.B.; Young, R.A. Activated signal transduction kinases frequently occupy target genes. *Science*, **2006**, *313*, 533-536.
- [200] Zhang, X.; Odom, D.T.; Koo, S.H.; Conkright, M.D.; Canetti, G.; Best, J.; Chen, H.; Jenner, R.; Herbolsheimer, E.; Jacobsen, E.; Kadam, S.; Ecker, J.R.; Emerson, B.; Hogenesch, J.B.; Unterman, T.; Young, R.A.; Montminy, M. Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 4459-4464.
- [201] Vries, R.G.; Prudenziati, M.; Zwartjes, C.; Verlaan, M.; Kalkhoven, E.; Zantema, A. A specific lysine in c-Jun is required for tran-

- scriptional repression by E1A and is acetylated by p300. *EMBO J.*, **2001**, *20*, 6095-6103.
- [202] Cesena, T.I.; Cardinaux, J.R.; Kwok, R.; Schwartz, J. CCAAT/enhancer-binding protein (C/EBP) beta is acetylated at multiple lysines: acetylation of C/EBPbeta at lysine 39 modulates its ability to activate transcription. *J. Biol. Chem.*, **2007**, *282*, 956-967.
- [203] Schwartz, C.; Beck, K.; Mink, S.; Schmolke, M.; Budde, B.; Wenning, D.; Klempnauer, K.H. Recruitment of p300 by C/EBPbeta triggers phosphorylation of p300 and modulates coactivator activity. *EMBO J.*, **2003**, *22*, 882-892.
- [204] Boulton, S.J. BRCA1-Mediated Ubiquitylation. *Cell Cycle*, **2006**, *5*, 1481-1486.
- [205] Xu, W.; Chen, H.; Du, K.; Asahara, H.; Tini, M.; Emerson, B.M.; Montminy, M.; Evans, R.M. A transcriptional switch mediated by cofactor methylation. *Science*, **2001**, *294*, 2507-2511.
- [206] Mark, W.Y.; Liao, J.C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabartty, A.; Arrowsmith, C.H. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? *J. Mol. Biol.*, **2005**, *345*, 275-287.
- [207] Lo, C.L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.*, **1999**, *285*, 2177-2198.
- [208] Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S.R.; Griffiths-Jones, S.; Howe, K.L.; Marshall, M.; Sonnhammer, E.L. The Pfam protein families database. *Nucleic Acids Res.*, **2002**, *30*, 276-280.
- [209] Hai, T.; Hartman, M.G. The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis. *Gene*, **2001**, *273*, 1-11.
- [210] Bailey, D.; O'Hare, P. Transmembrane bZIP Transcription Factors in ER Stress Signaling and the Unfolded Protein Response. *Antioxid. Redox. Signal.*, **2007**, *9*.
- [211] Hunger, S.P.; Li, S.; Fall, M.Z.; Naumovski, L.; Cleary, M.L. The proto-oncogene HLF and the related basic leucine zipper protein TEF display highly similar DNA-binding and transcriptional regulatory properties. *Blood*, **1996**, *87*, 4607-4617.
- [212] Goodman, R.H.; Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.*, **2000**, *14*, 1553-1577.
- [213] McManus, K.J.; Hendzel, M.J. CBP, a transcriptional coactivator and acetyltransferase. *Biochem. Cell Biol.*, **2001**, *79*, 253-266.
- [214] Lamprecht, C.; Mueller, C.R. D-site binding protein transactivation requires the proline- and acid-rich domain and involves the coactivator p300. *J. Biol. Chem.*, **1999**, *274*, 17643-17648.
- [215] Nerlov, C. The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol.*, **2007**, *17*, 318-324.
- [216] Yu, X.; Chini, C.C.; He, M.; Mer, G.; Chen, J. The BRCT domain is a phospho-protein binding domain. *Science*, **2003**, *302*, 639-642.
- [217] Hu, Y.F.; Li, R. JunB potentiates function of BRCA1 activation domain 1 (AD1) through a coiled-coil-mediated interaction. *Genes Dev.*, **2002**, *16*, 1509-1517.