CRC REVIVALS

The FOS and JUN Families of Transcription Factors

Edited by

Peter E. Angel, Peter Herrlich

CRC Press
Taylor & Francis Group

The FOS and JUN Families of Transcription Factors

Edited by

Peter E. Angel, Ph.D

Peter A. Herrlich, M.D., Ph.D.

Institute for Genetics and Toxicology

Karlsruhe, Germany

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the

Taylor & Francis Group, an **informa** business

First published 1994 by CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Reissued 2018 by CRC Press

A Library of Congress record exists under LC control number: 94013477

Publisher's Note
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original copies may be apparent.

Disclaimer
The publisher has made every effort to trace copyright holders and welcomes correspondence from those they have been unable to contact.

ISBN 13: 978-1-138-50650-3 (hbk)
ISBN 13: 978-0-203-71003-6 (ebk)

Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com and the
CRC Press Web site at http://www.crcpress.com

PREFACE

This book attempts to summarize our current understanding of a particularly important transcription factor family: AP-1. Is this the right time to give an interim review? In a way, yes, since AP-1 has become one of the prototype factors regulated by physiological and pathological signals from the microenvironment of cells. Much of what we currently explore with other transcription factors has been primed by work on AP-1.

Nevertheless, the AP-1 area still seems to expand exponentially; for example, the leucine zipper dimerization principle and the expanding number of putatively interacting subunits may "generate" the existence of some 100 to 200 different AP-1 factors. It is, therefore, a difficult area to review with any confidence of completeness. We have asked several experts in this field to contribute chapters. In order to keep chapters in an independently readable form, we have not rigorously eliminated repetitions. Rather, we have compromised by inserting connecting paragraphs and allowing for some redundancies, particularly if topics were approached from different angles. Furthermore, we have allowed a variety of interpretations.

We wish to thank all contributors and Ingrid Kammerer for help in preparing this book.

THE EDITORS

Peter E. Angel, Ph.D., is a junior staff member of the Institute for Genetics, Kernforschungszentrum Karlsruhe, Germany. He studied biology at the University of Karlsruhe, obtaining an M.S. degree in 1983. The diploma thesis was completed at the Institute for Genetics in 1983, the Ph.D. thesis (summa cum laude) in 1987. Dr. Angel then worked as a research associate at the University of California, San Diego, for almost three years. In 1990 he obtained a cancer fellowship at the Deutsche Forschungsgemeinschaft (Heisenberg Stipendium). He chose the Institute for Genetics in Karlsruhe as his home institution in 1990 and assumed his present position as a principal investigator in 1993. His research is supported, in addition to Kernforschungszentrum Karlsruhe, by Deutsche Forschungsgemeinschaft and the European Community. Dr. Angel is the author of 50 original papers, some of which became citation classics. His research interests relate to transcription factors, protein-protein interactions, and the function and regulation of matrix metalloproteinases.

Peter A. Herrlich, M.D., Ph.D., is chairman of the Institute for Genetics, Kernforschungszentrum Karlsruhe, and professor of genetics at the University of Karlsruhe. Dr. Herrlich obtained his M.D. degree at the University of Munich in 1964, and, after an internship at Cook County Hospital in Chicago, a medical license. After his postdoctoral years at the Max Planck Institute for Biochemistry, Munich, and at Rockefeller University, New York, he obtained a "Habilitation" (Ph.D.) degree at the Free University of Berlin in 1972. Dr. Herrlich was a senior staff member (associate professor) at the Max Planck Institute for Molecular Genetics, Berlin (1973-1977) and assumed his present position in 1977. In 1986-1987 he was a visiting professor at the University of California, San Diego. Dr. Herrlich is a member of the European Molecular Biology Organization, the European Science Foundation, and several other scientific associations. He is a co-editor of *Carcinogenesis* and a member of the editorial boards of *New Biologist*, *Molecular Carcinogenesis*, and the *International Journal of Cancer*. He has received the Meyenburg and Acker Prizes for Cancer Research. Dr. Herrlich serves on various agencies and study sections, e.g., as a member of the Kuratorium of Boehringer Ingelheim Fonds. His research has been supported, in addition to Kernforschungszentrum Karlsruhe, by Deutsche Forschungsgemeinschaft, the European Community, the Fonds der Chemischen Industrie, and Boehringer Ingelheim. Dr. Herrlich is an author on more than 300 papers and two books, and has co-edited one book. Examples of his current research interests are: genes involved in metastatic properties of cancer, DNA repair, signal transduction, and transcription factors.

INTRODUCTION

The vast array of behaviors and functions exhibited by a living organism are the result of induced changes of gene expression, especially at the level of transcription. Genes from the large complement of inherited genetic information are selected and their transcription turned on or off according to the needs that a given new condition imposes. Changes in transcription occur in response to a great variety of microenvironmental cues, including the supply of nutrients, stress agents, hormones, growth factors, position of cells (or nuclei) within a morphogen gradient, extracellular matrix, and components on neighboring cells. Indeed, understanding the control of gene expression is equivalent to understanding many of the fundamental properties that define life. If we set our goals modestly, we can say that there has been enormous progress in the last 10 years with respect to understanding the molecules participating in gene regulation. We can, for instance, answer questions such as: What are the biological structures through which a cell measures the conditions outside? How is this "knowledge" transformed into short- and long-term changes in transcription? And, finally, how are such responses integrated into crucial internal cellular programs, such as the cell cycle?

SOME BASICS OF THE TRANSCRIPTIONAL PROCESS IN EUCARYOTES

Transcription is the function of a multisubunit enzyme, RNA polymerase, which finds, with the help of auxiliary factors, the beginning of the gene and the noncoding strand. All of the proteins involved at the stage of transcription form the initiation complex at the promoter. In most of the genes transcribed by RNA polymerase II, a sequence conserved among most genes, the TATA box, determines where this complex assembles. In TATA-less promoters, an initiator sequence substitutes. Figure I-1 shows the multiprotein complex in position. Evidence for the constitution of the basic transcriptional initiation complex stems from *in vitro* transcription studies and purification from nuclear extracts of the minimal set of components required for the start reaction. Work on polymerase II was guided by previous studies on bacterial transcription and by sequence comparisons of promoter regions revealing the conserved features of the TATA-like elements.

When such defined start sequences in conjunction with a measurable reporter gene are incorporated into cells, only extremely low rates of expression will occur, regardless of how the cells are treated. None of the environmental cues listed above can turn on transcription. Promoter activity is only established by linking additional sequences to the TATA–reporter construct. In general terms, such sequences could be named enhancers since they "enhance" the activity of the basic initiation complex. The TATA–reporter construct would serve as an "enhancer trap," since it indicates the presence of sequences from anywhere in the genome that are able to enhance transcription. In a given gene such sequences are most often found 5′ of the TATA box (with respect to the direction of transcription; the 5′ is determined by the first nucleotide of the transcript). In fact, one or the other sequence needs to be in close proximity (e.g., within 100 bp) to make a promoter functional. However, many enhancing sequences have been found at a considerable distance (e.g., 3.6 kb) and also within the gene (in introns and exons) and 3′ of the gene. To assemble all of the sequence elements necessary for completely "normal" behavior of a gene, rather large segments of the genome (possibly 300 kb or more) are needed. Evidence of this latter conviction comes from transgene technology. It appears that only the large yeast artificial chromosome (YAC) clones ensure that a gene carries all the intrinsic information needed for transcriptional regulation through ontogeny and adult life.

Enhancer sequences (and their opposite: silencers) are recognized by specialized proteins: transcription factors. One common property of these factors is their ability to bind DNA in a highly specific manner. Each factor binds with preference to the specific

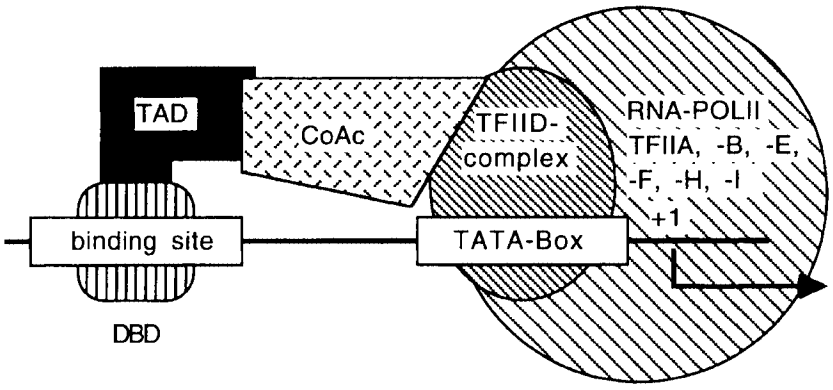

Figure I-1 The RNA polymerase II preinitiation complex. The transcription factor, binding to specific upstream enhancer elements, is composed of a DNA binding domain (DBD) and a transactivation domain (TAD). The TAD interacts with the components of the basal transcription machinery (RNA polymerase II plus basal factors such as TFIIA, -B, -D, etc). TFIID stands for a protein complex that consists of the TATA-binding protein (TBP) binding to the TATA sequence and associated factors (TAFs) that do not bind to DNA but interact directly or indirectly with TBP. The individual subunits of TFIID as well as their putative spatial arrangement are described in: Drapkin et al., *Curr. Opinion Cell Biol.*, 5, 469, 1993; and Weinzierl et al., *Nature*, 362, 511, 1993. The TAD of the transcription factor can interact either directly with TBP or TAFs or through the action of a bridging protein (co-activator) that mediates the link between the TAD of the transcription factor and TBP or TAFs.

enhancer sequence it is "built" for. The affinity calculated is in the range of 10^{-10} *M*. The transcription factors this book is devoted to are among the best characterized and have contributed considerably to the discovery of transcription factor properties.

Transcription factors carry several functional domains in addition to DNA binding. One of these, transactivation domain (TAD), is responsible for making contact with the basic transcriptional machinery. For an enhancer factor, this contact results in elevated transcriptional initiation. Transcription factors seem to be able to interact directly with the TATA-binding protein (TBP), at least *in vitro*, or indirectly with the TBP via interaction with TBP-associated factors (TAFs). Thus, the "active" initiation complex possesses at least two anchors on the DNA: one at the enhancer binding site, the other through TBP binding to the TATA box. To permit formation of such contacts over considerable distances, enhancer sequences may be located away from the initiation complex. It is thought that the DNA stretch between the initiation complex and the enhancer loops to bring the enhancer sequence and its transcription factor in close proximity to the preinitiation complex (Figure I-1).

The most interesting problem of transcription factor function concerns the modulation of the transactivating function. Transactivation is the interaction between the components of the basic preinitiation complex, and the interacting portion of the transcription factor is referred to as the transactivation domain (Figure I-1). The activity of many, if not all, transcription factors is influenced in response to the conditions of the microenvironment. They are responsible for changing transcription when appropriate stimuli act on cells. These activations are rather specific for given agents, ensuring that changes in the genetic program occur only in response to the specific stimulus. The molecular mechanism of such activation has been identified in only a few examples. Steroid hormone receptors,

for example, are converted to active transcription factors by the binding of the specific hormone ligand. Stimulation at the outer surface of the plasma membrane often results in a process of signal transduction via protein kinase–phosphatase cascades. Transcription factors, as the ultimate targets of these cascades, are posttranslationally modified (e.g., directly by phosphorylation at serines/threonines). Probably only very few of the transactivation-modulating mechanisms have yet been identified. It is conceivable, furthermore, that transcription factors are responsive to more than one modifying reaction and, in particular, to other proteins by protein–protein interaction.

THE AP-1 FAMILY

Much of our present knowledge about transcription factors comes from the discovery and study of the AP-1 factor family. One of its members, the heterodimer Fos–Jun, was found as early as 1982, as a protein complex containing the viral oncogene product Fos, without a clue of its function. The term AP-1 (activating protein-1) was coined for an activity that supports basal level transcription *in vitro* at the metallothionein gene promoter (1987). AP-1 was also immediately recognized as the decisive control element of the collagenase promoter *in vivo*, and it was demonstrated that it could be activated by external stimulation with the tumor promoter TPA.

The classical control site where AP-1 binds in the collagenase promoter reads 5′–TGAGTCA–3′. Even in the first purification, AP-1 looked like a family of related proteins. This interpretation was later proven to be true by the identification of multiple cJun- and cFos-related proteins encoded by distinct genes. This family has relative affinity to the classical AP-1 binding sequence. It appears now, however, that each member of the family has a strong preference for one very specific DNA sequence that is only related to the classical one. For instance, one portion of the family prefers the consensus 5′–TGANNTCA–3′, with an additional nucleotide in the center of the recognition sequence. This sequence is preferentially bound by members of the cAMP responsive element binding protein (CREB/ATF) family. In fact, distinct members of the CREB/ATF family dimerize with the various Jun proteins, resulting in a change of sequence specificity. Moreover, in addition to the core element 5′–TGA(N)$_1$or$_2$TCA–3′, adjacent nucleotides can influence specificity.

Each AP-1 factor is composed of two subunits. It is not known, however, how many AP-1 factors exist. If each subunit could combine with each other subunit (which is not strictly possible), we could predict some 100 members of the family. Obviously, an understanding of the functional significance of such a large family lags behind our ability to characterize each member.

Most data on promoters cannot distinguish which AP-1 factor is active under a given *in vivo* condition. Many assignments are, therefore, arbitrary unless the need for a particular factor has been proven by antisense or knockout techniques. An added level of complexity is rendered to this system by the observation that different AP-1 dimers seem to affect DNA structure differently (DNA bending). DNA may also affect transcription factor conformation, as has been shown for steroid receptors. Furthermore, individual dimeric complexes can function in opposite directions. Fos, for example, is active in proliferation, differentiation, and apoptosis. Clearly, the context of activation as well as the variety of dimerizing partners present determine the ultimate response of the cell.

Why a book on the AP-1 family and not on other families, such as SP-1 or the octamer binding factors? We believe that a unique treatment is justified by the central role AP-1 plays both in mammalian organisms and in the investigation of transcription factors in general. AP-1 has served to detect one of the decisive DNA binding motifs and the bZip interaction. It is the AP-1 family (and NFκB) for which inducible activation was first detected and carried into molecular exploration. Its members are engaged in the control

of cell proliferation as well as various types of differentiation, and, further, in neural function and stress responses. AP-1 is one of the key factors that translates external stimuli into both short- and long-term changes of gene expression. Several subunits of AP-1 are transforming proteins and required for transformation by other oncogenes. Work on AP-1 has also led to the discovery of cross-talk and cross-coupling between transcription factors.

This monograph attempts, therefore, to describe the "state of the art" with respect to Fos and Jun in detail. Topics will range from the structure of AP-1 and the principles of its DNA binding and function, to the exploration of its functions in physiology and pathology. The reader will realize that the story of AP-1 focuses on most of the current fronts of transcription research, including coordination with the cell cycle and the role of modifying enzymes in transcriptional regulation.

CONTENTS

III. Functions in Physiological Processes

A. AP-1-Regulated Target Genes

B. Neural Cell Function

C. Differentiation

D. Stress Response

IV. Functions in Pathology

A. Fos and Jun Oncogenes

B. Oncogenes Acting Through Fos and Jun

Section I

CHARACTERISTICS OF AP-1 SUBUNITS

Chapter 1

General Structure of AP-1 Subunits and Characteristics of the Jun Proteins

Peter Angel and Peter Herrlich

CONTENTS

GENERAL STRUCTURE OF THE AP-1 SUBUNITS

According to its function in controlling gene expression, the prototype of a transcription factor has to comprise two, or possibly three, properties: a region of the protein responsible for binding to a specific DNA recognition sequence (DNA binding domain); a region required for transcriptional activation (transactivation domain) once the protein is bound to DNA; and, possibly, a modulation domain. These functions are encoded by separate regions (modules) of the protein that generally function independently; however, domain swapping experiments have shown that the transactivation domain can be fused to a heterologous DNA binding domain to form a potent transcription factor of new promoter specificity.

In Jun proteins, the transactivation domain is located within the *N*-terminal half of the protein while the DNA binding domains of Jun and CREB/ATF are located at the *C*-terminus (Figure 1-1A). The transactivation domains of Fos and CREB/ATF have not been determined precisely, but, in the case of Fos, transactivation seems to be influenced by amino acids at both the *N*-terminus and *C*-terminus of the protein (see Chapters 2 and 8). The DNA binding domain of Fos is located near the center of the protein.

In vivo mutation analysis of the Jun proteins has identified three subdomains that together form the transactivation domain (Angel et al., 1989; Hirai et al., 1990). The subdomains are characterized by an abundance of acidic amino acids that are essential for activity (Angel et al., 1989). In addition, *in vitro*, a fourth region near the DNA binding domain has been identified (Bohmann and Tjian, 1989). These regions are thought to be responsible for the link between the transcription factor bound to DNA and the RNA polymerase II preinitiation complex. The link is established by either direct or indirect interaction with components of the basal transcription machinery. Based on these amino acid sequences of the transactivation domain, the Jun proteins belong to the family of "acidic-blob"-type transcription factors (Angel et al., 1989; Oehler and Angel, 1992). However, in contrast to other members of the acidic-blob group, for example, the yeast transcription factor GAL4 (Ma and Ptashne, 1987) or the VP16 protein of herpes simplex virus (Trietzenberg et al., 1988), the transactivation domains of the Jun proteins do not seem to form an extensive α-helical structure, since these regions also contain single or multiple proline residues known to disrupt α-helices.

0-8493-4573-1/94/$0.00+$.50

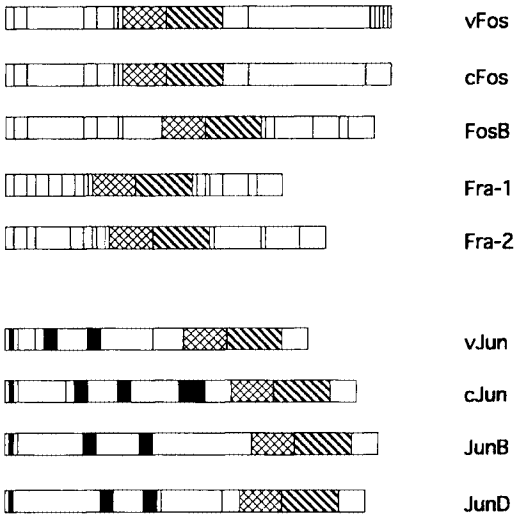

Figure 1-1A Structural organization of the Jun and Fos proteins. The hatched boxes (▨) represent the "leucine zipper" region. The cross-hatched boxes (▩) indicate the "basic domain." The box marked by stripes (▥) indicates the part of vFos generated by a frame shift mutation in cFos. In the Fos proteins the shaded boxes (□) denote highly conserved regions of unknown functions; vertical lines represent deletions (comparing members of either the Fos or Jun families). In the Jun proteins the solid boxes (■) represent the transactivation regions; the shaded boxes (▒) in cJun show the glutamine- and proline-rich region of unknown function (present in human but not in mouse or avian cJun).

In contrast to the transactivation domains, whose structural properties are poorly understood, a large body of information on the DNA binding domains of Jun, Fos, and other AP-1 factors has been collected (Vogt and Bos, 1990; Busch and Sassone-Corsi, 1990; Ransone and Verma, 1990). Mutation analysis has revealed characteristic properties that are evolutionarily conserved between the Fos, Jun, and CREB proteins, thus defining the protein family called "bZip proteins" (Landschulz et al., 1988). bZip (see Figure 1-1B) stands for the amino acid sequences of the two independently acting subregions of the DNA binding domain: the "basic domain," rich in basic amino acids responsible for contacting the DNA, and the "leucine zipper" region, characterized by heptad repeats of leucine, which is responsible for dimerization that in turn is a prerequisite for DNA binding (Kouzarides and Ziff, 1988; Sassone-Corsi et al., 1988; Gentz et al., 1989; Turner and Tjian, 1989; Neuberg et al., 1989b). While amino acid substitutions or deletions in the leucine zipper abolish dimerization, mutations in the basic domain abolish DNA binding without affecting dimerization. Domain swapping experiments have shown that both domains are interchangeable among the different bZip proteins without loss of their physical properties (Neuberg et al., 1989a; Sellers and Struhl, 1989; Kouzarides and Ziff, 1989; Cohen and Curran, 1990).

Based on the potential formation of an α-helical structure of the leucine zipper region and the need for dimerization, dimer formation was proposed to be mediated by interdigitation (zipper formation) of the leucines, which are located in a linear fashion at the inner phase of the helix (Landschulz et al., 1988). Since proteins exist that contain a

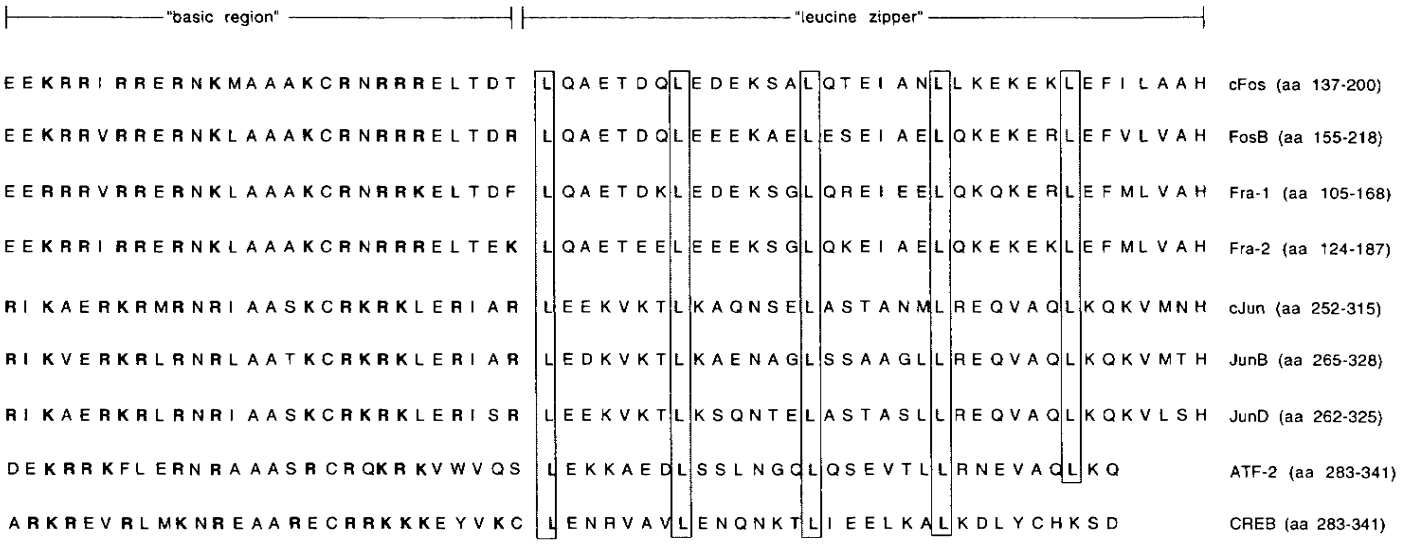

Figure 1-1B Amino acid sequences of the "bZip" region of the Fos, Jun, and CREB/ATF proteins. The positions of the leucines in the "leucine zipper" are designated by shaded boxes; basic amino acids in the "basic region" are indicated by bold letters.

heptad repeat of leucine residues but do not dimerize with Fos, CREB/ATF, or Jun (e.g., the Myc protein), and because other amino acids within the leucine zipper region were also found to be critical for dimer formation, the "zipper model" was revised (O'Shea et al., 1989a, 1989b), proposing that the leucine zipper region of both Fos and Jun forms an α-helical structure of 4-3 repeats (see Figure 5 in Chapter 2). The location of a hydrophobic amino acid at position *a* and the leucines at position *d* favor the formation of dimers in parallel orientation by a so-called "coiled coil" interaction, with the leucines and hydrophobic amino acids positioned toward the contact interface (see Figure 2-5). It has been proposed that the stability of dimers is determined by charged amino acids at positions *e* and *g* of the heptad repeat, forcing either electrostatic repulsion or stabilization of the dimer. Amino acid residues at positions *b*, *c*, and *f* are not involved in dimer formation (Schuermann et al., 1991; O'Shea et al., 1992). On alignment of two Fos monomers in a coiled coil structure, the *e* and *g* positions are both covered by acidic amino acids that are likely to cause electrostatic repulsion, which may explain why Fos is not able to form stable homodimers. In contrast, efficient dimerization between the Jun and Fos proteins is likely to be promoted by ionic interactions between the positively charged residues at the *e* and *g* positions of Jun with the negatively charged residues in Fos (see also Chapter 2). The number of such putative interactions is lower in Jun/Jun homodimers, resulting in reduced stability (O'Shea et al., 1989). The same type of interaction rules are probably responsible for complex formation of cJun with specific members of the CREB/ATF protein family (ATF-2), whereas Fos does not efficiently form Fos/CREB heterodimers (Benbrook and Jones, 1990; McGregor et al., 1990; Hai and Curran, 1991; Müller et al., unpublished results).

We have tested the validity of these rules by individually modifying charged amino acids at positions *e* and *g* of the leucine zippers of cJun and ATF-2, and have obtained contradictory results: cJun homodimer formation in solution was not significantly affected after replacing two lysines with two glutamines, although this change was expected to result in strong repulsion. In contrast, however, the replacement of two glutamines in the ATF-2 leucine zipper with two lysines did have a negative effect on dimer formation (van Dam et al., unpublished results). Thus, although electrostatic interactions between amino acids at positions *e* and *g* may contribute to dimer stability, additional interactions, including association with DNA, appear to be important as well. Another exception to the *e–g* rule has been found for homodimerization of JunB. Two amino acids at positions *b* and *c* strongly affect homodimer formation by an unknown mechanism (Deng and Karin, 1993), most likely a general destabilization of the α-helical structure of the leucine zipper of JunB (Deng and Karin, 1993).

Most studies have been performed in solution *in vitro* in the absence of DNA. It is conceivable that DNA exerts influence on dimer formation by preference for specific dimers or by participating in the complex formation. In turn, as will be described below, Jun/Jun and Jun/Fos each cause different bending of DNA (opposite bending; see Kerppola and Curran, 1991a, 1991b).

Coiled coil formation as a major mechanism of protein–protein interaction is not without precedence; this type of interaction was first described for the structure of the α-class of fibrous proteins, such as keratin, myosin, and fibrinogen (Cohen and Parry, 1986). A special feature of transcription factors, the ability to form dimers, is combined with the ability to bind to DNA in a sequence-specific manner. Sequence specificity in the basic domain is encoded by the amino acid sequence that interacts with the major groove of the DNA (Vinson et al., 1989; O'Neil et al., 1990). Based on studies on C/EBP, sequence specificity is established in such a way that both α-helices of the dimerized proteins bifurcate beyond the leucine zipper as a consequence of the positive charges of the juxtaposed basic regions, forcing the two basic regions to track along the major groove of the DNA (scissor grip model; Vinson et al., 1989). Because the Fos, Jun, and CREB/

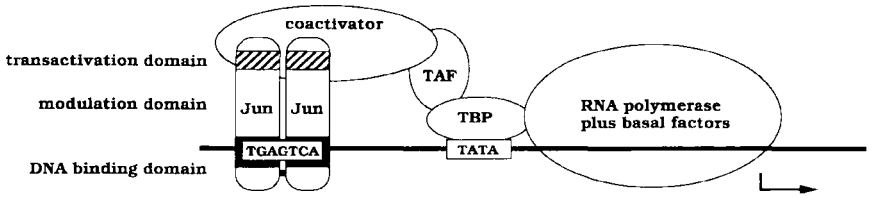

Figure 1-2 Transactivation of AP-1-dependent genes. The hatched (▨) and solid boxes (■) signify the transactivation and DNA binding domains, respectively, of Jun. The existence of putative modulation domain(s) of Jun interacting with other cellular proteins still must be confirmed. The TATA binding protein (TBP), TBP-associated factors (TAFs), RNA polymerase II, and basal transcription factors (TFIIA, -B, -E, -F, -H, and -I) make up the RNA polymerase II preinitiation complex (see also Figure 1-3 and Figure I-1). In Jun homodimers, the transactivation domain of Jun interacts with a specific coactivator (p52/54) to make contact with the preinitiation complex. The receiving target of coactivator interaction (which type of TAF?) remains to be determined. The mechanism by which the members of the Fos and CREB/ATF protein families interact with the components of the basal transcription machinery is still largely unknown.

ATF proteins exhibit some differences in sequence specificity and structure of the DNA–protein complexes (Hai and Curran, 1991; Ryseck and Bravo, 1991), the combined ability of dimerization and DNA binding allows for multiple combinations of dimers that differ in their biological properties (see also Chapter 2).

In addition to the DNA binding and transactivation domains, transcription factors may also contain specific regions responsible for: (1) interaction with other cellular proteins distinct from components of the RNA polymerase initiation complex (modulation domain, Figure 1-2), (2) nuclear translocation, and (3) regulation of protein stability. In both the Fos and Jun proteins, regions have been identified that interact with other cellular proteins; for example, steroid hormone receptors (see Chapters 3 and 4). While the nuclear localization signal in both Fos and Jun has been identified within the basic domain (Tratner and Verma, 1991; Chida and Vogt, 1992), the specific regions involved in the rapid degradation of the Fos and Jun proteins, possibly regulated by phosphorylation and ubiquitination (Papavasiliou et al., 1992; D. Bohmann, personal communication), have not yet been found.

CHARACTERISTICS OF THE JUN PROTEINS

Originally thought to be encoded by a single type of protein, AP-1 is generated by a series of dimers of products of the Fos, Jun, and CREB/ATF protein families (Bohmann et al., 1987; Angel et al., 1988a; Distel et al., 1987; Rauscher et al., 1988a, 1988b; Hai et al., 1988), as well as other bZip proteins (e.g., LRF-1), which have not yet been characterized (Hsu et al., 1991). In addition, associations have been observed between Fos or Jun and the p65 subunit of NFκB (Stein et al., 1993), and ATF-2 and p50-NFκB (Du et al., 1993). Combinatorial association can draw on three Jun genes (c-jun, junB, junD), four Fos genes (c-fos, fosB, fra-1, fra-2) and several CREB/ATF genes (Benbrook and Jones, 1990; Hai and Curran, 1991; Ryseck and Bravo, 1991; Vogt and Bos, 1990). Despite the high degree of homology in the overall structural features described in the previous section, the different members of the Fos, Jun, and CREB families exhibit significant differences, which lead to subtle differences in DNA binding and transcriptional activation (Ryseck and Bravo, 1991; Hai and Curran, 1991; Chiu et al., 1989; Hirai et al., 1989;

Deng and Karin, 1993) and which suggest specific functions in gene regulation for individual dimers. This chapter focuses primarily on the Jun proteins. Characteristics of the Fos proteins will be described in Chapter 2.

THE DNA BINDING DOMAIN

Considerable differences in selectivity and stability among the bZip proteins have been found with regard to dimerization. As described above, the efficiency of stable dimer formation depends primarily on the charged amino acid residues at positions *e* and *g* of the α-helical structure of the leucine zipper region (Schuermann et al., 1991; O'Shea et al., 1992). With respect to these sites, the amino acid sequences between individual members of each family (e.g., cJun, JunB, and JunD) are fairly conserved. They differ greatly, however, when comparing the subfamilies Jun, Fos, and CREB/ATF (Figure 1B). These differences in amino acid composition of the contact interface of the dimer probably determine the specificity rules of dimerization. Such specificity of coiled coil interaction has been proved by domain-swapping experiments (Sellers and Struhl, 1989; Neuberg et al., 1989a) and has served to screen cDNA libraries for interacting subunits (Benbrook and Jones, 1990; McGregor et al., 1990; Ivashkiv et al., 1990). Most importantly, Fos cannot bind to DNA and regulate gene expression on its own, but depends on the presence of Jun or other proteins to form heterodimers. In contrast, both CREB/ATF and Jun proteins can bind to DNA on their own by forming homodimers (Bohmann and Montminy, 1987; Angel et al., 1988a; Hoeffler et al., 1988; Mackawa et al., 1989; Gonzalez and Montminy, 1989), although with lower stability (Jun/Jun as compared to Jun/Fos heterodimers).

The existence of a large number of possible combinations of partners would only by relevant physiologically if the individual dimers differed in DNA affinity or sequence specificity of the combined DNA binding domains of the subunits. Putative core sequence elements recognized by AP-1 factors are quite variable and fall into two classes, 7 bp (TPA responsive element, TRE) and 8 bp (cAMP responsive element, CRE) long. Site-directed mutagenesis experiments have shown that the amino acids in the basic regions of Fos and Jun required for interaction with the DNA are highly conserved between the two proteins (see also Chapter 2). *In vitro* binding studies of chimeric bZip proteins (e.g., of Jun background in which the basic domain of Jun was replaced by the equivalent sequence of Fos) revealed differences in sequence specificity (Nakabeppu and Nathans, 1989): dimers with two basic regions of Jun (or Fos) bind to the CRE with somewhat greater affinity than to the "classical" AP-1 site (TRE). In contrast, heterodimers of Fos and Jun bind preferentially to the TRE (Nakabeppu and Nathans, 1989). CREB homodimers prefer the CRE but also exhibit significant affinity for the TRE (Masquilier and Sassone-Corsi, 1992). There is little difference in the binding site specificity between the Fos and Jun basic regions since some base pair substitutions of the binding site resulted in a marked decrease in binding whereas others had a lesser effect or resulted in enhanced binding (Risse et al., 1989). Nevertheless, UV cross-linking experiments have suggested that both Jun and Fos make direct contact with DNA in a favored orientation of the complex on the TRE (Nakabeppu and Nathans, 1989). This interpretation is in line with the finding that Jun/Fos heterodimers prefer binding to asymmetrical, 7-mer TREs while Jun homodimers prefer binding to symmetrical, 8-mer CRE sequences (Nakabeppu and Nathans, 1989; Ryseck and Bravo, 1991). Exchanging Fos in a Fos/Jun heterodimer against ATF-2 abolishes binding to the TRE but allows efficient interaction with the CRE (Benbrook and Jones, 1990; McGregor et al., 1990; Ivashkiv et al., 1990; Hai and Curran, 1991; van Dam et al., 1993). In addition to the core nucleotide sequence of the TRE or CRE, sequence specificity is also affected by the flanking nucleotides (Deutsch et al., 1988; Ryseck and Bravo, 1991).

Models of bZip protein/DNA interaction propose that contacts between the straight α-helix of the basic domain and the major groove of straight B-DNA are limited to a maximum of 12 contiguous amino acids, which can contact a maximum of 5 bp of DNA (Vinson et al., 1989; O'Neil et al., 1990). In contrast, the basic region extends over 20 amino acid residues (Nakabeppu and Nathans, 1989), and the DNA binding site consists of at least 7 nucleotides (Angel et al., 1987; Risse et al., 1989). Thus, the basic region of the protein and/or the DNA must be bent or distorted to allow for the identified regions of contact between the molecules. Circular dichroism spectroscopy has demonstrated that the basic domain undergoes a conformational transition to a structure of high α-helix content in the presence of the cognate DNA binding site (O'Neil et al., 1990; Patel et al., 1990). This conformational change of the protein induced by DNA binding enforces DNA bending, as measured by the anomalous electrophoretic mobility of bent DNA fragments (Kerppola and Curran, 1991a, 1991b). Based on circular permutation and phasing analysis, Jun/Jun homodimers and Jun/Fos heterodimers induce bends in opposite directions: Fos bends DNA away from the dimer interface, causing the major groove to extend over the recognition site, and Jun/Jun homodimers bend DNA toward the dimer interface (Kerppola and Curran, 1991a, 1991b). While these data have been obtained by analyzing Jun/Jun homodimers and Jun/Fos heterodimers, it is not known whether this is true for other dimeric complexes, such as Fra-1/Jun or Jun/ATF-2. In addition, posttranslational modifications (e.g., phosphorylation) within the DNA binding domain of Jun (see Chapter 5) may affect the ability of Jun to bend DNA. Regardless of dimer specificity, differences in DNA bending might have important consequences on the transactivation function of Jun/Jun homodimers or Jun/Fos heterodimers: protein-induced bending has been proposed to be a mechanism that allows the interaction of factors bound to separate sites on the DNA and selection among different potential protein–protein interactions in a region containing multiple factors bound to DNA (Moitoso de Vargas et al., 1989).

THE TRANSACTIVATION DOMAIN

In vivo competition (squelching) experiments (Figure 1-3) have suggested the necessary participation of an intermediary protein (coactivator) interacting with the transactivation domain of cJun, in order to link Jun to TBP or TAFs (Angel et al., 1989; Oehler and Angel, 1992). In this type of experiment, transcriptional activation of an AP-1-dependent promoter by overexpression of cJun is strongly reduced by coexpression of a Jun mutant that contains the transactivation domain but lacks the DNA binding domain (schematically illustrated in Figure 1-3). Overexpression of the Jun mutant represses the Jun-induced transcription but does not affect basal activity of the reporter plasmid or the activity of other promoters whose expression is regulated by DNA binding proteins distinct from AP-1. These results suggest that the Jun mutant competes for binding of an intermediary factor to the transactivation domain of Jun required for transcriptional activation. *In vivo* competition correlates with the binding of a cluster of proteins with a molecular weight of 52, 53, and 54 kDa (p52/54; Oehler and Angel, 1992) in that transactivation domain-negative Jun mutants neither squelch nor bind the 52- to 54-kDa proteins. Transcriptional interference of cJun-specific transactivation as well as reduced physical interaction between Jun and p52/54 is also observed in the presence of other "acidic-blob"-type transactivators, including GAL4 and VP16, suggesting that the intermediary factor(s), including p52/54, are shared by transcription factors whose transactivation domains are characterized by an abundance of acidic amino acids. In contrast, transcription factors that are characterized by a different type of transactivation domain (e.g., Sp1, CTF, GHF1/Pit-1, or the estrogen receptor) do not interfere with Jun-specific transactivation, showing that these proteins either use a different type of bridging protein or interact directly with TBP or TAFs (Oehler and Angel, 1992).

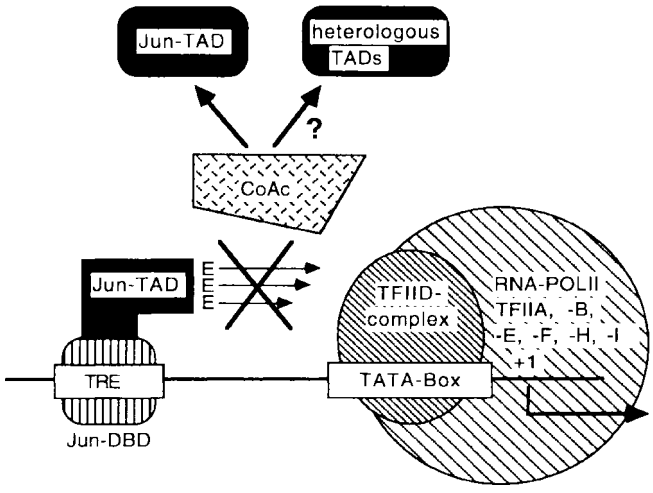

Figure 1-3 *In vivo* competition of Jun-dependent transactivation. To demonstrate the requirement and specificity of a bridging protein (coactivator, CoAc), transactivation of a TRE-dependent promoter is measured in the presence of an excess of truncated forms of Jun (Jun-TAD) or other transcription factors (heterologous TADs) that contain the transactivation domain (TAD) but lack the DNA binding domain (DBD). Diminution of promoter activity is explained by titrating out the coactivator through competitive binding of the mutants. "E" stands for the presence of glutamate residues required for TAD activity (and, most likely, interaction with the coactivator protein).

Diminution of cJun-specific transactivation is also observed by coexpression of the transactivation domain of JunB, suggesting that cJun and JunB make use of the same bridging protein (Oehler and Angel, 1992). In fact, the amino acid sequences of the three subdomains of the cJun and JunB transactivation domains share high degrees of homology (Vogt and Bos, 1990), and both proteins are efficient activators of promoters containing multiple binding sites (Chiu et al., 1989). Nevertheless, analysis of transactivation of AP-1-dependent promoters containing a single binding site has clearly shown that nonconserved amino acids outside these three subdomains have an important function in establishing a functional transactivation domain: while cJun homodimers strongly activate such promoters, neither JunB nor JunD homodimers efficiently stimulate transcription but, rather, suppress stimulation of these promoters by cJun, possibly by forming inactive cJun/JunB heterodimers (Deng and Karin, 1993). These distinct behaviors of cJun and JunB with respect to their transactivation potential are probably due to differences in their transactivation domains. Studies with chimeric proteins containing the DNA binding domain of GAL4 and a minimal transactivation domain, cJun or JunB (amino acids 5–89), revealed that GAL4–cJun is a potent transactivator while GAL4–JunB fails to activate transcription (Franklin et al., 1992). In the case of JunB, a switch from the inactive to the active state of the transactivation domain might occur because of an unknown mechanism of cooperativity established by binding of JunB to multiple binding sites in a given promoter. This switch might also occur by heterodimerization of JunB or JunD with Fos, resulting in potent transactivators of promoters containing either single or multiple AP-1 binding sites. Interestingly, two of the three regions that constitute the transactivation domain of the Jun proteins (domains II and III) are also found in Fos and are required for transactivation by Fos (see also Chapter 2). These motifs, homology boxes 1 and 2 (HOB1, HOB2), can be exchanged; e.g., HOB1 of Fos can be replaced by

HOB1 of Jun (Sutherland et al., 1992). Since interaction of Jun and p52/54 depends on the presence of Jun's transactivation domain, it is possible that p52/54, or a closely related protein, also interacts with Fos. So far, however, we have not detected stable physical interactions between Fos and p52/54. As a tentative interpretation, Fos does not carry domain I (amino acids 6–12 of Jun), which was found to be essential for the interaction between Jun and p52/54 (Oehler and Angel, 1992).

In summary, the formation of specific homo- and heterodimers in a given cell depends on the relative abundance of individual members of the bZip protein family. The mix of dimers will change with any change of a given subunit's abundance. Due to the unique properties of the specific domains of individual bZip proteins (for the Fos proteins, see Chapter 2) genes will be activated according to the subset of single or multiple TRE or CRE sequences they carry in their promoters, and according to the prevalent AP-1 dimers present.

ACKNOWLEDGMENTS

We would like to thank Thomas Oehler for preparing some of the figures and Dr. Larry Sherman for critical reading of the manuscript.

REFERENCES

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R. J., Rahmsdorf, H. J., Jonat, C., Herrlich, P., and Karin, M., Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor, *Cell*, 49, 729, 1987.

Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., Hunter, T., and Karin, M., Oncogene jun encodes a sequence-specific trans-activator similar to AP-1, *Nature*, 332, 166, 1988a.

Angel, P., Hattori, K., Smeal, T., and Karin, M., The jun proto-oncogene is positively autoregulated by its product, Jun/AP-1, *Cell*, 55, 875, 1988b.

Angel, P., Smeal, T., Meek, J., and Karin, M., Jun and v-Jun contain multiple regions that participate in transcriptional activation in an interdependent manner, *New Biol.*, 1, 35, 1989.

Benbrook, D. M. and Jones, N. C., Heterodimer formation between CREB and JUN proteins, *Oncogene*, 5, 295, 1990.

Bohmann, D. and Tjian, R., Biochemical analysis of transcriptional activation by Jun: Differential activity of c- and v-Jun, *Cell*, 59, 709, 1989.

Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K., and Tjian, R., Human proto-oncogene c-jun encodes a DNA binding protein with structural and functional properties of transcription factor AP-1, *Science*, 238, 1386, 1987.

Busch, S. J. and Sassone-Corsi, P., Dimers, leucine-zippers and DNA binding domains, *Trends Genet.*, 6, 36, 1990.

Chida, K. and Vogt, P. K., Nuclear translocation of viral Jun but not of cellular Jun is cell cycle dependent, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 4290, 1992.

Chiu, R., Angel, P., and Karin, M., Jun-B differs in its biological properties from, and is a negative regulator of, cJun, *Cell*, 59, 979, 1989.

Cohen, C. and Parry, D. A. D., α-Helical coiled coils — a widespread motif in proteins, *Trends Biochem. Sci.*, 11, 245, 1986.

Cohen, D. R. and Curran, T., Analysis of dimerization and DNA binding functions in Fos and Jun by domain-swapping: involvement of residues outside the leucine zipper/basic region, *Oncogene*, 5, 929, 1990.

Deng, T. and Karin, M., JunB differs from cJun in its DNA-binding and dimerization domains, and represses c-Jun by formation of inactive heterodimers, *Genes Dev.*, 7, 479, 1993.

Deutsch, P. J., Hoeffler, J. P., Jameson, J. L., and Habener, J. F., Cyclic AMP and phorbol ester-stimulated transcription mediated by similar DNA elements that bind distinct proteins, *Proc. Natl. Acad. Sci. U.S.A.*, 85, 7922, 1988.

Distel, R. J., Ro, H.-S., Rosen, B. S., Groves, D. L., and Spiegelman, B. M., Nucleoprotein complexes that regulate gene expression in adipocyte differentiation: Direct participation of c-fos, *Cell*, 49, 835, 1987.

Du, W., Thanos, D., and Maniatis, T., Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements, *Cell*, 74, 887, 1993.

Franklin, C. C., Sanchez, V., Wagner, F., Woodgett, J. R., and Kraft, A. S., Phorbol ester-induced amino-terminal phosphorylation of human JUN but not JUNB regulates transcriptional activation, *Proc. Natl. Acad. Sci. U.S.A.*, 89, 7247, 1992.

Gentz, R., Rauscher, F. J., III, Abate, C., and Curran, T., Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains, *Science*, 243, 1695, 1989.

Gonzalez, G. A. and Montminy, M. R., A cluster of phosphorylation sites on the cyclic AMP-regulated nuclear factor CREB predicted by its sequence.,*Nature*, 337, 749, 1989.

Hai, T. and Curran, T., Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 3720, 1991.

Hai, T., Liu, F., Allegretto, E. A., Karin, M., and Green, M. R., A family of immunologically related transcription factors that includes multiple forms of ATF and AP-1, *Genes Dev.*, 2, 1216, 1988.

Hirai, S.-I., Ryseck, R.-P., Mechta, F., Bravo, R., and Yaniv, M., Characterization of junD: A new member of the jun proto-onocogene family, *EMBO J.*, 8, 1433, 1989.

Hirai, S.-I., Bourachot, B., and Yaniv, M., Both Jun and Fos contribute to transcription activation by the heterodimer, *Oncogene*, 5, 39, 1990.

Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L., and Habener, J. F., Cyclic AMP-responsive DNA-binding protein: Structure based on a cloned placental cDNA, *Science*, 242, 1430, 1988.

Hsu, J.-C., Laz, T., Mohn, K. L., and Taub, R., Identification of LRF-1, a leucine-zipper protein that is rapidly and highly induced in regenerating liver, *Proc. Natl. Acad. Sci. U.S.A.*, 88, 3511, 1991.

Ivashkiv, L. B., Liou, H.-C., Kara, C. J., Lamph, W. W., and Verma, I. M., mXBP/CRE-BP2 and c-Jun form a complex which binds to the cyclic AMP, but not to the 12-O-tetradecanoylphorbol-13-acetate, response element, *Mol. Cell. Biol.*, 10, 1609, 1990.

Kerppola, T. K. and Curran, T., Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: Implications for transcription factor cooperativity, *Cell*, 66, 317, 1991a.

Kerppola, T. K. and Curran, T., DNA bending by Fos and Jun: The flexible hinge model, *Science*, 254, 1210, 1991b.

Kouzarides, T. and Ziff, E., The role of the leucine zipper in the fos–jun interaction, *Nature*, 336, 646, 1988.

Kouzarides, T. and Ziff, E., Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding, *Nature*, 340, 568, 1989.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L., The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins, *Science*, 240, 1759, 1988.

Ma, J. and Ptashne, M., Deletion analysis of GAL4 defines two transcriptional activating segments, *Cell*, 48, 847, 1987.

McGregor, P. F., Abate, C., and Curran, T., Direct cloning of leucine zipper proteins: Jun binds cooperatively to the CRE with CRE-BP1, *Oncogene*, 5, 451, 1990.

Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J.-I., Yoshida, M., and Ishii, S., Leucine zipper structure of the protein CRE-BP1 binding to the cyclic AMP response element in brain, *EMBO J.*, 8, 2020, 1989.

Masquilier, D. and Sassone-Corsi, P., Transcriptional cross-talk: Nuclear factors CREM and CREB bind to AP-1 sites and inhibit activation by Jun, *J. Biol. Chem.* 267, 22460, 1992.

Moitoso de Vargas, L., Kim, S., and Landy, A., DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase, *Science*, 244, 1457, 1989.

Nakabeppu, Y. and Nathans, D., The basic region of Fos mediates specific DNA binding, *EMBO J.*, 8, 3833, 1989.

Neuberg, M., Adamkiewicz, J., Hunter, J. B., and Müller, R., A Fos protein containing the Jun leucine zipper forms a homodimer which binds to the AP1 binding site, *Nature*, 341, 243, 1989a.

Neuberg, M., Schuermann, M., Hunter, J. B., and Müller, R., Two functionally different regions in Fos are required for the sequence-specific DNA interaction of the Fos/Jun protein complex, *Nature*, 338, 589, 1989b.

Oehler, T. and Angel, P., A common intermediary factor (p52/54) recognizing "acidic-blob"-type domains is required for transcriptional activation by the Jun proteins, *Mol. Cell. Biol.*, 12, 5508, 1992.

O'Neil, K. T., Hoess, R. H., and DeGrado, W. F., Design of DNA-binding peptides based on the leucine zipper motif, *Science*, 249, 774, 1990.

O'Shea, E. K., Rutkowski, R., and Kim, P. S., Evidence that the leucine zipper is a coiled coil, *Science*, 243, 538, 1989a.

O'Shea, E. K., Rutkowski, R., Stafford, W. F., III, and Kim, P. S., Preferential heterodimer formation by isolated leucine zippers from Fos and Jun, *Science*, 245, 646, 1989b.

O'Shea, E., Rutkowski, R., and Kim, P. S., Mechanism of specificity in the Fos-Jun oncoprotein heterodimer, *Cell*, 68, 699, 1992.

Papavasiliou, A. G., Treier, M., Chavrier, C., and Bohmann, D., Targeted degradation of c-Fos, but not v-Fos, by a phosphorylation-dependent signal on c-Jun, *Science*, 258, 1941, 1992.

Patel, L., Abate, C., and Curran, T., DNA binding by Fos Jun results in altered protein conformation, *Nature*, 347, 572, 1990.

Ransone, L. J. and Verma, I. M., Nuclear proto-oncogenes Fos and Jun, *Annu. Rev. Cell. Biol.*, 6, 539, 1990.

Rauscher, F. J., III, Cohen, D. R., Curran, T., Bos, T. J., Vogt, P. K., Bohmann, D., Tjian, R., and Franza, B. R., Jr., Fos-associated protein p39 is the product of the jun proto-oncogene, *Science*, 240, 1010, 1988a.

Rauscher, F. J., III, Sambucetti, L. C., Curran, T., Distel, R. J., and Spiegelman, B. M., Common DNA binding site for Fos protein complexes and transcription factor AP-1, *Cell*, 52, 471, 1988b.

Risse, G., Jooss, K., Neuberg, M., Brüller, H.-J., and Müller, R., Asymmetrical recognition of the palindromic AP-1 binding site (TRE) by Fos protein complexes, *EMBO J.*, 8, 3825, 1989.

Ryseck, R.-P. and Bravo, R., c-Jun, JunB and JunD differ in their binding affinities to AP-1 and CRE consensus sequences: Effect of Fos proteins, *Oncogene*, 6, 533, 1991.

Sassone-Corsi, P., Ransone, L. J., Lamph, W. W., and Verma, I. M., Direct interaction between fos and jun nuclear oncoproteins: Role of the "leucine zipper" domain, *Nature*, 336, 692, 1988.

Schuermann, M., Hunter, J. B., Hennig, G., and Müller, R., Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization, *Nucleic Acids Res.*, 19, 739, 1991.

Sellers, J. W. and Struhl, K., Changing Fos oncoprotein to a Jun-independent DNA-binding protein with GCN4 dimerization specificity by swapping 'leucine zipper', *Nature*, 341, 74, 1989.

Stein, B., Baldwin, A., Ballard, D., Greene, W., Angel, P., and Herrlich, P., Cross-coupling of the NFκBp65 and Fos/Jun transcription factors produces potentiated biological functions, *EMBO J.*, 12, 3879, 1993.

Sutherland, J. A., Cook, A., Bannister, A. J., and Kouzarides, T., Conserved motifs in Fos and Jun define a new class of activation domain, *Genes Dev.*, 6, 1810, 1992.

Tratner, I. and Verma, I. M., Identification of a nuclear targeting sequence in the Fos protein, *Oncogene*, 6, 2049, 1991.

Trietzenberg, S. J., Kingsbury, R. C., and McKnight, S. L., Functional dissection of VP16, the transactivator of herpes simplex virus immediate early gene expression, *Genes Dev.*, 2, 718, 1988.

Turner, R. and Tjian, R., Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers, *Science*, 243, 1689, 1989.

van Dam, H., Duyndam, M., Rottier, R., Bosch, A., de Vries-Smits, L., Herrlich, P., Zantema, A., Angel, P., and van der Eb, A., Heterodimer formation of cJun and ATF-2 is responsible for induction of c-jun by the 243 amino acid adenovirus E1A protein, *EMBO J.*, 12, 479, 1993.

Vinson, C. R., Sigler, P. B. and McKnight, S. L. Scissor-grip model for DNA recognition by a family of leucine zipper proteins, *Science*, 246, 911, 1989.

Vogt, P. K. and Bos, T. J., Jun: Oncogene and transcription factor, *Adv. Cancer Res.*, 55, 1, 1990.

Chapter 2

The Fos Family: Gene and Protein Structure, Homologies, and Differences

Marcus Schuermann

CONTENTS

INTRODUCTION

The c-*fos* proto-oncogene was first isolated as the cellular homologue of two viral *fos* oncogenes encoded by the Finkel-Biskis-Jenkins (FBJ) and Finkel-Biskis-Reilly (FBR) murine sarcoma viruses (MuSV) both of which induce osteosarcomas in rats and mice (Van Beveren et al., 1983; Finkel et al., 1966; Finkel et al., 1975). It has since been learned that expression of c-*fos* is rapidly and transiently inducible by the addition of growth factors, such as PDGF and EGF (see Chapter 8, this volume; Greenberg and Ziff, 1984; Kruijer et al., 1984; Müller et al., 1984). This stimulation involves direct transcriptional activation at the promoter level and places the c-*fos* gene among the first cellular "immediate early genes" described. Subsequent work has shown that the encoded product, the c-Fos protein, is a nuclear phosphoprotein, associated with chromatin (Renz et al., 1987; Sambucetti and Curran, 1986) and complexed to a second protein in the range of 39 kDa (hence, termed p39) (Curran and Teich, 1982a). Several important discoveries followed, which have laid the foundation for our current view of Fos as a transcription factor subunit (Müller, 1986; Curran and Franza, 1988; Herrlich and Ponta, 1989; Morgan and Curran, 1989; Lucibello and Müller, 1991; Angel and Karin, 1991). Among these discoveries were the following: (1) p39 is identical to the product of the proto-oncogene *jun* (Rauscher et al., 1988a; Angel et al., 1988; Sassone-Corsi et al., 1988a), (2) both proteins are major components of the transcription factor AP-1 (Bohmann et al., 1987; Lee et al., 1987; Chiu et al., 1988), and (3) a specific palindromic recognition element, $TGA^{G}/_{C}TCA$, described as AP-1 binding site or TRE (derived from *T*PA-*r*esponsive *e*lement), is the major DNA target for both proteins (Angel et al., 1987; Distel et al., 1987).

0-8493-4573-1/94/$0.00+$.50

Similarly, three additional *fos*-related genes were identified, either by cross-hybridization of conserved DNA fragments (leading to isolation of the *fos*B gene) (Zerial et al., 1989) or through immunological cross-reactions of their encoded gene products, so-called *F*os *r*elated *a*ntigens (Franza et al., 1987), giving rise to the isolation of the *fra*-1 (Cohen and Curran, 1988) and *fra*-2 genes (Matsui et al., 1990; Nishina et al., 1990).

All *fos* members exhibit a number of similarities, with respect to the regulation of their expression and the level of protein function. In most differentiated tissues, *fos* genes are expressed at low levels but are inducible by a variety of extracellular stimuli working through common signal transduction pathways (Franza et al., 1987; Cohen and Curran, 1988; Matsui et al., 1990; Nishina et al., 1990, see also Chapter 8, this volume). This rapid increase in the rate of transcription is counteracted by a number of mechanisms acting at the different transcriptional and posttranscriptional levels that help in the negative regulation of gene expression, which also implies cross-regulation among the Fos members (Lucibello and Müller, 1991).

At the protein level, all products encoded by *fos* genes show a number of similar properties, such as sequence-specific binding to the TRE, complex formation with Jun proteins, and transactivation from TRE-containing promoter sequences, which can be explained by the considerable degree of structural conservation within this protein family (Franza et al., 1987; Cohen and Curran, 1988; Matsui et al., 1990; Nishina et al., 1990; Schuermann et al., 1991b). These common aspects make it difficult to assess a particular role for the Fos protein in the process of cell proliferation or differentiation. Only recently has there been evidence that shows that individual members may also diverge in their presumptive biological functions. This evidence is based, on the one hand, on observed differences in protein function, such as the transregulatory influence on distinct promoter segments (Gizang-Ginsberg and Ziff, 1990; Suzuki et al., 1991; Wisdom et al., 1992), the transrepression of transcription from promoters containing serum-responsive elements (SREs) (Nakabeppu and Nathans, 1991; Mumberg et al., 1991; Yen et al., 1991), or the interference with members of the steroid receptor family (Lucibello et al., 1990). On the other hand, this assumption is also supported by the observation of differential expression of the individual *fos* genes in certain tissues (Morgan and Curran, 1989) and during certain stages of development (Dony and Gruss, 1987; Redemann-Fibi et al., 1991). Since many of these features are discussed elsewhere in this volume, it is the purpose of this chapter to summarize primarily the basic structural features of the *fos* genes and proteins as well as the influence of structure on protein function.

ORGANIZATION OF THE *fos* GENES AND GENE EXPRESSION

While the cDNA sequence of all *fos* genes has been determined, less is known about the genomic structure of the respective genes. To date, aside from the organization of the viral genes, only the structure of the c-*fos* and *fos*B genes has been sufficiently documented (Van Beveren et al., 1983; Van Straaten et al., 1983), and the structure of the *fra*-2 gene only in part (Nishina et al., 1990).

C-*fos*

The c-*fos* gene is 4 kbp long, including the proximal promoter sequences, and is interrupted by three introns (see Figure 2-1). From this gene, a single 2.2-kbp mRNA is transcribed, which is in agreement with the position mapped for the transcriptional start and the polyadenylation site. The mouse c-*fos* gene has been assigned to the [E-D] region of chromosome 12 (D'Eustachio, 1984), and the human gene assigned to chromosome 14q21–q31 (Barker et al., 1984). Sequence analysis of c-*fos* genes from mouse, man, and chicken reveals that the gene is highly conserved among vertebrate species; cDNAs

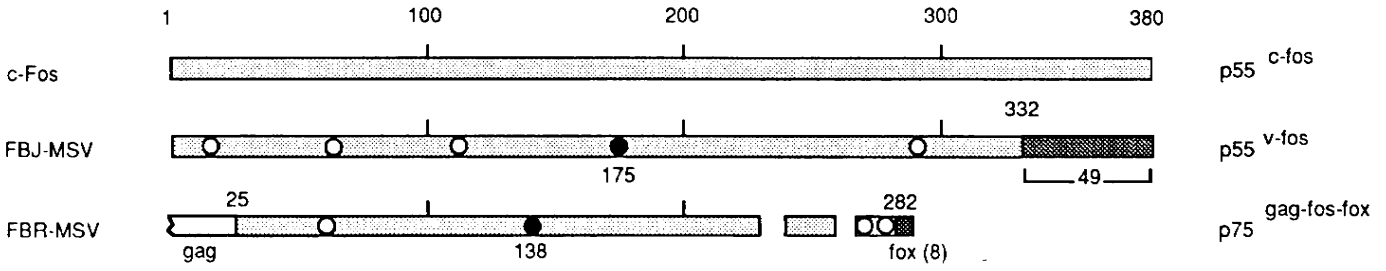

Figure 2-1 Structure of the viral and cellular c-Fos proteins (Curran and Teich, 1982b; Van Beveren et al., 1983, 1984). Hatched bars show areas of almost conserved structural identity, stippled and open bars in FBJ–Fos and FBR–Fos denote sequences not homologous to c-Fos. Amino acid positions are indicated relative to the c-Fos protein. Circles indicate point mutations leading to the substitution of individual amino acids in both proteins (positions 15, 67 110, 175, and 291 in FBJ–Fos and positions 64, 138, 268, 279, and 280 in FBR–Fos, relative to the c-Fos protein). The mutations affecting functional properties are indicated by a filled circle.

derived from mouse and chicken show 79% identity at the protein level (Mölders et al., 1987; Fujiwara et al., 1987). Underlining the high degree of evolutionary conservation, a gene related to c-*fos* (d*FRA*) exists in *Drosophila*. The encoded protein, harbors in its central region a domain, the "bZip region", that is conserved in all Fos proteins and is needed to bind to a palindromic TRE sequence and allow formation of dimeric complexes with a Jun-like protein (dJRA). Nevertheless, with respect to the rest of the protein, there are considerable deviations indicating that, in *Drosophila*, dFRA might also serve other functions (Perkins et al., 1990).

The c-*fos* promoter has been analyzed extensively, and a number of regulatory elements mediating the basal activity, response to growth factors, oncogenes, cAMP pathways and G_1-specific transcription factors have been identified (see Chapter 8, this volume). There are also indications that some of these relevant promoter elements as well as the exon–intron boundaries may be conserved in evolution, since these features are completely conserved between mouse and man (Van Straaten et al., 1983).

In most tissues, c-*fos* expression is tightly controlled. mRNA is expressed at only relatively low levels, but can be rapidly and transiently induced as early as 10 to 15 min after the addition of growth factors, phorbol esters, cytokines, or a number of compounds activating different intracellular pathways. Following a brief peak of expression, mRNA levels are reduced efficiently and kept low in the absence of external stimuli. At least three mechanisms have been shown to cooperate in this downregulation of c-*fos* expression: transcriptional shutoff through negative autoregulation of the c-*fos* promoter (König et al., 1989; Lucibello et al., 1989; Sassone-Corsi et al., 1988b; see also Chapter 8), premature termination of nascent RNA transcripts (Lamb et al., 1990), and, posttranscriptionally, rapid turnover of RNA because of destabilizing sequences located in both the 3′-untranslated and protein-coding regions (Meijlink et al., 1985; Shyu et al., 1989).

V-*fos*

It is generally believed that many proto-oncogenes were acquired by retroviruses during evolution and, hence, have been structurally modified to facilitate viral survival (Varmus, 1982). Underscoring this concept, two viral *fos* genes have been determined from the genome of two viral strains, both of which induce osteosarcomas in mice: the FBJ-MuSV and the FBR-MuSV (Curran and Teich, 1982b; Van Beveren et al., 1984; see also Chapter 14). The discovery of a virus-borne *fos* oncogene led to the identification of the corresponding cellular homologue, the c-*fos* proto-oncogene (Van Beveren et al., 1983).

Both viral *fos* genes show distinct structural deviations from the c-*fos* gene. FBJ-*fos* diverges at the 3′ end of the coding region resulting from a frame-shift mutation caused by a deletion of 104 nucleotides. This affects the translation of the last 48 amino acids following Pro332, which are replaced by 49 residues provided by the different reading frame (see Figure 2-2). Thus, the FBJ-MuSV derived *fos* gene codes for a protein of 381 amino acids, approximately 55 kDa in size, which is comparable to the molecular weight of the c-Fos protein. Because of the *C*-terminal alterations, however, important residues including regulatory phosphorylation sites are missing in FBJ-Fos, leading to increased protein stability, impaired negative autoregulation potential and, probably, release from nuclear export control as described by Roux et al. (1990; see also Chapter 7). Apart from the *C*-terminal alterations, the rest of FBJ-Fos is nearly identical to c-Fos, with the exception of five amino acid substitutions that are of no or only marginal relevance for protein function. To date, only the Glu–Lys substitution at position 175 has been shown to have a detectable influence on the interaction with Jun proteins (Schuermann et al., 1991a).

Contrary to FBJ, the FBR-MuSV derived *fos* gene diverges considerably from c-*fos*, owing to large structural alterations in the FBR-MuSV genome, including truncation at both termini: 25 amino acids of c-Fos at the amino-terminus and 98 amino acids at the

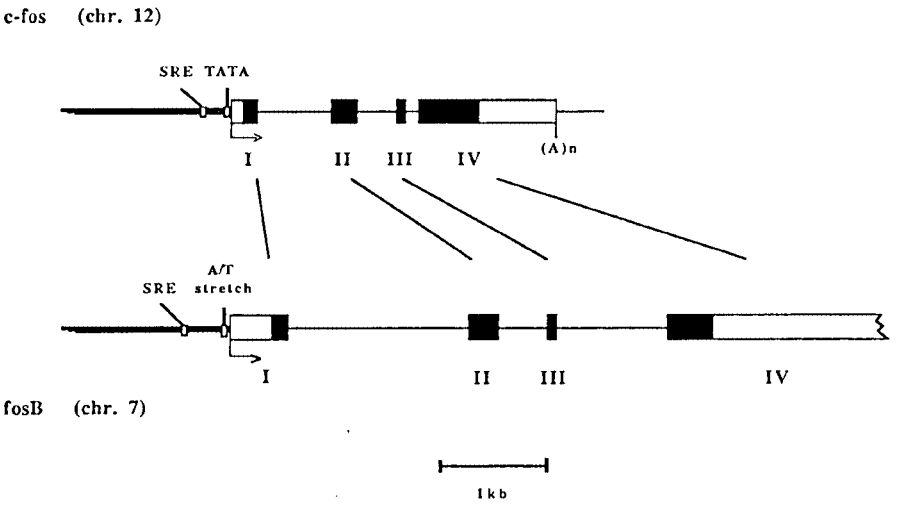

Figure 2-2 Genomic organization of the mouse c-*fos* and *fos*B genes (Lazo et al., 1992; Mumberg and Schuermann, unpublished results). Shown are the conserved elements in the promoter of both genes and the common exon–intron structure. SRE, serum response element.

C-terminus have been truncated. The *C*-terminus has been replaced by eight residues derived from cellular sequences (termed *fox)* (Müller, 1986). Moreover, due to the lack of premature splice and termination signals, the FBR-*fos* gene is expressed only as a Gag–Fos–Fox fusion protein with a molecular weight of 75 kDa. In addition to these gross alterations, two small internal deletions, spanning 13 and 10 amino acids, respectively, and five point mutations leading to amino acid substitutions have been found (see Figure 2-2). At present, it is not clear to what extent these individual deletions affect the functional properties of the FBR-Fos protein. Most noticeable, however, is the increased potential of FBR-Fos to morphologically transform (see Chapter 14) and to promote the establishment of mouse connective tissue cells. The latter was shown to be due to a single point mutation in the FBR-*fos* gene, leading to a change of Glu138 to Val (Jenuwein and Müller, 1987).

*fos*B

The *fos*B gene is similar to c-*fos* in both its genomic structure and with respect to the coding part (Zerial et al., 1989; Lazo et al., 1992; Mumberg and Schuermann, unpublished results). The *fos*B gene covers a region of approximately 8 kbp on mouse chromosome 7 (A1–B1 region) (Lazo et al., 1992). As can be deduced from Figure 2, the mouse *fos*B gene contains three introns and has identical exon–intron boundaries to c-*fos*. In addition, a serum response element (SRE) and an adjacent AP-1-like sequence (FAP) found in the c-*fos* promoter are also present in the *fos*B promoter. The promoter also contains several overlapping "TATA"-like elements. The start site of transcription has been located 40 nucleotides downstream (Lazo et al., 1992). A polyadenylation signal has not been mapped so far. The high degree of structural conservation between c-*fos* and *fos*B suggests that these genes have originated from a common ancestor form. Like c-*fos*, expression of *fos*B is also transiently inducible in most adult tissue by a number of growth factors and shows a comparable pattern with respect to its subsequent down regulation.

Contrary to c-*fos*, however, the primary transcript derived from the *fos*B gene is subjected to alternative splicing. The alternative splice eliminates 140 nucleotides within exon IV due to the presence of appropriate splice signals therein. Thus, two mature mRNA forms of approximately 3.9 and 4 kbp in length occur in the cytoplasm (Nakabeppu and Nathans, 1991; Mumberg et al., 1991; Yen et al., 1991; Dobrzanski et al., 1991). The mRNAs code for two versions of the same protein, both of which differ in their molecular properties, particularly in their transforming and transrepressing abilities (see below and Chapter 14, this volume). Furthermore, the two forms of *fos*B mRNA are also expressed at different times following serum stimulation, with the long form preceding the short form by approximately 30 min, thus pointing to a novel and potentially interesting way to regulate *fos*-dependent gene expression (Mumberg et al., 1991).

fra-1 AND *fra*-2

While the genomic structure of *fra*-1 has not been reported, two cDNA sequences derived from either rat embryo fibroblasts or human U937 monocytic cells have been isolated (Cohen and Curran, 1988; Matsui et al., 1990). Like c-*fos*, both *fra*-1 genes are highly conserved and cross-hybridize with avian genomic DNA. The rat *fra*-1 clone is a nearly full-length cDNA version and corresponds in size with the major 1.6-kbp transcript observed in most cells on serum or TPA stimulation. In addition, a minor 3.3-kbp transcript also has been observed in TPA-stimulated U937 cells (Matsui et al., 1990).

With respect to genomic organization, more is known about the *fra*-2 gene. As deduced from a partial sequence of chicken *fra*-2, this gene contains four exons with exon–intron boundaries identical to those of c-*fos* and *fos*B genes (Nishina et al., 1990). A *fra*-2 cDNA clone also has been obtained from human U937 cells. Both genes show a considerable degree of conservation (Matsui et al., 1990). Three mRNA forms can be distinguished following TPA treatment of U937 cells: two minor forms of 1.7 and 6 kbp and a major transcript of 2.3 kbp. In chicken embryo fibroblasts stimulated by serum, the 6 kbp form seems to be the most prominent. Whether these different mRNA transcripts result from the use of alternative start, splice, or polyadenylation signals remains to be shown. Unlike the rapid induction of c-*fos* and *fos*B mRNA, accumulation of *fra*-1 and *fra*-2 transcripts is not seen before 30 to 60 min following serum stimulation and remains elevated for approximately 1 to 3 hours thereafter. This initial lag phase in the kinetics of induction may reflect the additional requirement of factors that are newly synthesized and/or activated in response to mitogenic signals (e.g., c-Fos-regulated gene products; see Chapter 10, this volume), which may help to augment and maintain the level of transcription.

FOS PROTEIN STRUCTURE: A LESSON IN MODULAR ARCHITECTURE

In support of the concept that all *fos* genes may have been derived from a common ancestor, the coding sequences of all *fos* genes show remarkable areas of almost complete conservation on the nucleic acid and protein levels. If one omits the two viral Fos proteins, FBJ-Fos and FBR-Fos, which are c-Fos homologues that have undergone different point mutations and structural alterations, the products of the *fos*-related genes show five major stretches of significant homology: two short areas comprising about 10 and 20 amino acids in the *N*-terminal part, a central domain of about 85 amino acids, and two stretches of approximately 25 amino acids each at the *C*-terminal end (see Figure 2-3). These areas of relatively conserved sequences are interrupted by stretches of amino acids showing little or no homology at all. The nonconserved segments are of variable length and are responsible for the different protein sizes, ranging from 380 amino acids for c-Fos