REVIEW: TRANSCRIPTION

Signal Transduction and the Control of Gene Expression

Ali H. Brivanlou[1] and James E. Darnell Jr.[2]*

More than 2000 transcription factors are encoded in the human genome. Such proteins have often been classified according to common structural elements. But because transcription factors evolved in the service of biologic function, we propose an alternative grouping of eukaryotic transcription factors on the basis of characteristics that describe their roles within cellular regulatory circuits.

All cellular life can recognize and properly respond to molecules in the extracellular environment. Indeed, an increased repertoire of recognized extracellular signaling molecules matched with increasingly sophisticated intracellular responses was the central requirement for the evolution of metazoan life. Two very broad fields of research, which are often described as "signal transduction" and "control of gene expression," have merged recently to become a pivotal arena for developmental genetics as well as cellular biochemistry. In this review we present a classification of transcription factors that is intended to organize thinking about the connection of extracellular signaling to the regulation of transcription in eukaryotic cells.

A host of proteins crucial to transcription initiation are assembled into the RNA polymerase, the general transcription factors, coactivators, corepressors, chromatin remodelers, histone acetylases, deacetylases, kinases, and methylases, to list the main participants (*1–5*). These crucial proteins are present in all eukaryotic cells and contribute to the initiation of every RNA polymerase II primary transcript that eventually becomes messenger RNA.

As important as the ~200 to 300 proteins that constitute the coactivators and the transcriptional machinery may be to the survival of cells and organisms, the regulation of the choice of specific initiation sites for transcription is not vested in these proteins. Rather, transcriptional regulation depends on members of an even larger number of proteins, in mammals perhaps 2000 to 3000 (*6, 7*), with two characteristic domains: a DNA binding domain that binds gene-specific regulatory sites directly, and a second domain that exhibits transcriptional activation potential. In some cases this dual requirement is shared between partner proteins, so that the site-specific binding domain and transcription activation domain occur on separate proteins. These site-specific transcription factors recruit coactivators and the transcription machinery to initiate gene-specific transcription (*1–5*). As development and cell specialization occurs, selection among these 2000+ transcription factors for the regulation of cell-specific gene expression involves (i) a cascade of transcriptional control of transcription factor genes, and (ii) signals from outside the cell that activate, posttranscriptionally, already formed transcription factors. In the regulatory regions in the DNA of a few well-studied vertebrate genes (*8, 9*), as many as six to eight different protein chains, acting on one enhancer (together forming an "enhanceosome"), are required for gene-specific regulation, and this is likely true for many other genes. The combinatorial use of subsets of the 2000+ proteins could easily mean that the complete set of regulators for each gene is unique, ensuring the right amount of the right protein at the right time as development proceeds.

There are also proteins (most likely a smaller total number) that interact with coactivators or positive-acting transcription factors and/or a member of the transcription machinery to interdict their activity. But in general the negative-acting proteins in animals do not bind DNA in a gene-specific manner to sterically inhibit polymerase binding. This is quite different from the classic Jacob-Monod model of bacterial repression (*10*). When not specifically activated or lacking a proper partner, some site-specific DNA binding proteins can also have an inhibitory function. Comprehensive reviews of eukaryotic proteins that negatively affect transcription are available (*11–15*).

Positive-Acting Eukaryotic Transcription Factors

We have divided the positive-acting transcription factors into groups on a functional basis. By function we do not mean a molecular definition, such as the structure of the DNA binding motif or association with particular cofactors, but rather a definition based on cell biological and regulatory function. For example, does the factor directly participate in changing the rate of transcription? Does the factor, once synthesized, automatically enter the nucleus to act in transcription? If the factor requires a signal to become active in transcriptional regulation, what is the nature of that signal? If we group factors on the basis of the answers to these questions, we have a useful framework in which to discuss the role of transcription factors in the design of regulatory circuits (Fig. 1).

I. Constitutively active nuclear factors. A sizable group of site-specific DNA binding proteins are present in the cell nucleus of all cells at all times, and they have transcriptional activating potential as assessed by in vitro transcription or in vivo synthetic reporter gene assays. However, these activators have not been, by themselves, implicated in changing the rates of individual gene transcription in a chromosomal context. They very likely play an important facilitating role in the transcription of many chromosomal genes, possibly in genes that seem to be always transcribed [e.g., structural proteins like tubulin and actin, and ubiquitous metabolic enzymes such as glyceraldehyde phosphate dehydrogenase (GAPDH)]. These constitutively active transcription factors can also participate in composing enhanceosomes together with transcription factors that are regulated. This group includes Sp1 (*3, 16*), CCAAT binding protein (*17*), NF1 (*18, 19*), and many others (Fig. 1).

II. Regulatory transcription factors. We distinguish two broad classes of regulatory transcriptional activators.

IIA. Developmental ("cell type–specific") transcription factors. Beginning in a fertilized egg (or germinated seed), the cell-specific accumulation during development of this group of factors is dependent largely on sequential waves of regulated transcription of the genes encoding these transcription factors (*20–22*). The cells in which such factors are generated may require extracellular signals to make these factors, but members of this group of proteins enter the nucleus as soon as they are made without requiring any regulated posttranslational event. Examples include

[1]Laboratory of Molecular Vertebrate Embryology, [2]Laboratory of Molecular Cell Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: darnell@rockefeller.edu

many early embryonic factors in *Drosophila* [e.g., bicoid (*22, 23*)], the Hox cluster of homeobox genes sequentially expressed in the anteroposterior axis of vertebrates (*24*), a series of helix-loop-helix factors (MyoD, Myf5, and myogenin) that appear in sequence in the control of muscle differentiation (*25*), and a series of factors required for β cell differentiation in pancreatic islets (*26*). Developmental factors are not strictly cell-, tissue-, or region-specific. Rather, combinatorial distribution of groups of these factors in different cell types helps to direct cell determination (choice of cell fate) and differentiation (synthesis of recognized cell-specific proteins).

Although all the proteins mentioned here automatically enter the nucleus upon being made and thereby immediately contribute to transcription, a number of developmental factors do shuttle in and out of the nucleus and can be phosphorylated in the cytoplasm, leading to a block in their reentry to the nucleus. Such regulation occurs in muscle to Mef-2, a MADS box protein, and to at least several different forkhead (winged helix) proteins (Fig. 2) (*27–29*). In addition, some factors (e.g., MyoD and myogenin) can be phosphorylated on serine while in the nucleus, resulting in decreased transcriptional activation potential (*30*).

IIB. Signal-dependent transcription factors. These proteins (or their precursor polypeptides) may be either developmentally restricted in their expression or present in most or all cells, but all are inactive (or minimally active) until cells containing such proteins are exposed to the appropriate intra- or extracellular signal. Three broad classes of these signal-dependent transcription factors are recognized and outlined in Figs. 1 and 2.

IIB(1). The steroid receptor superfamily. Steroids dissolve in the lipid bilayer of the plasma membrane and enter cells, where they bind and activate one of the many [50 in humans (*31*)] specific steroid receptors that then participate in activating specific gene transcription (*32, 33*). Some members of this extensive family are restricted to specific cell types, and many are known to become phosphorylated on serine, although the importance of this phosphorylation remains unknown in most cases. All except the glucocorticoid receptor (GR) are found primarily in the nucleus before the appearance of their cognate hormone. GR is held in a cytoplasmic complex until glucocorticoids bind it and release the GR dimers, which then enter the nucleus (*34, 35*).

IIB(2). Transcription factors activated by internal (cell-autonomous) signals. Small intracellular signaling molecules that regulate preexisting transcription factors have been recently recognized. Internal sterol concentrations regulate the proteolysis of a membrane protein precursor within the membrane to release sterol response element binding protein (SREBP) (*36*) (Fig. 2). In yeast, low internal unsaturated fatty acid concentration leads to the juxtamembrane cleavage of precursors to Spt23 and MGA2, liberating for nuclear entry these two transcription factors that regulate genes of unsaturated fatty acid synthesis. Internally generated lipid compounds activate steroid-like receptors (*37, 38*), and genome damage that demands DNA repair enzyme function increases p53 concentration (*39*). Another recently recognized pathway in yeast is the unfolded protein response, one arm of which induces differential splicing to produce an active transcription factor. This pathway probably exists in mammalian cells, because a homologous DNA binding protein exists (*40*). Transcription factor regulation by internal signals is the most recently recognized regulatory group and will almost certainly grow in the future.

IIB(3). Transcription factors activated by cell surface receptor–ligand interactions. There are two major routes from extracellular signaling proteins to genes. First, there are dozens of intracellular serine phosphorylation cascades that end at hundreds of resident nuclear protein substrates (*41, 42*). Second, a more limited number of latent cytoplasmic transcription factors are activated after cell surface receptor–ligand interaction and then accumulate in the nucleus to drive transcription.

IIB(3)(a). Constitutive nuclear factors activated by serine phosphorylation. Small molecules such as epinephrine and leucotrienes or peptides such as thyroid-stimulating hormone and adrenocorticotropic hormone (*43, 44*), when bound to their cognate G-coupled cell surface receptor proteins (GCRPs), lead to increases in intracellular second messengers [adenosine 3′,5′-monophosphate (cAMP), phosphoinositides, diacylglycerol, and Ca^{2+}] that trigger serine kinase cascades and phosphorylation of resident nuclear transcription factors (Fig. 2). A second route of signaling with a similar outcome is triggered by polypeptide ligands that bind cell surface receptors possessing intrinsic tyrosine kinases (receptor tyrosine kinases or RTKs) (*45, 46*). Thus, both liganded GCRPs and RTKs regulate transcription through a multitude of serine kinase cascades that finally terminate in serine phosphorylation of the abundant resident nuclear transcription factors. The substrate proteins include the Ets family, the c-Jun–c-Fos–ATM family, the cAMP response element–binding protein (CREB)–cAMP response element modulator (CREM) families, and the MADS box family of transcription factors [including serum response factor (SRF)], to list some of the better studied groups. Many of these target proteins are not only constitutively nuclear but are bound to DNA at all times (*41, 47–51*).

Serine phosphorylation of resident nuclear factors is evolutionarily ancient and occurs in fungi, plants, and single-celled organisms such as *Saccharomyces cerevisiae* (*52–54*). Because the resident nuclear factors are so numerous and the number of serine kinases

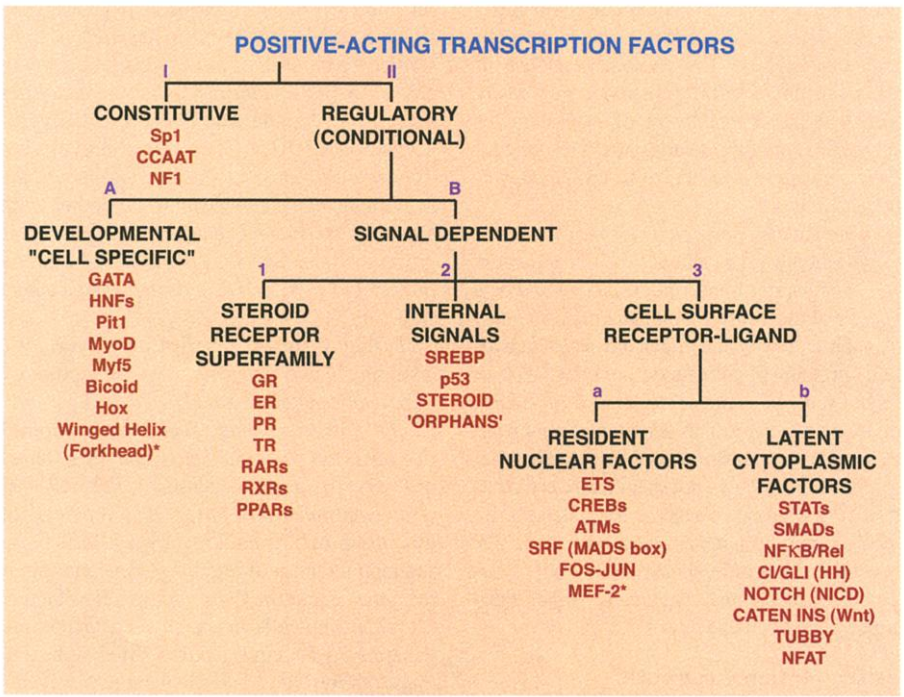

Fig. 1. Functional classification of positive-acting transcription factors. Major functional groups are shown in black; specific examples are illustrated in red. The list of examples is not exhaustive. An asterisk indicates that the indicated factors can be trapped in cytoplasm by serine phosphorylation.

that affect them so large, it has been difficult to trace precise protein pathways to the activation of individual genes in a natural context.

For example, mammalian CREB protein must be phosphorylated on Ser^{133} for maximal activity in transfection assays (*46, 55*). Phosphorylation allows binding to CREB of a coactivator CBP (CREB binding protein, or a relative p300). At least five different kinases [cAMP-dependent protein kinase (PKA), ribosome S6 kinase, and multiple mitogen-activated protein (MAP) kinases and calcium/calmodulin–dependent (CAM) kinases] can catalyze this phosphorylation. Other examples of proteins that can be substrates of multiple kinases are the Ets and MADS box families (*51, 52, 55*).

A further word about the extremely important MAP kinase group of serine kinases is in order (*46, 53, 54*). There are in yeast at least five separate serine kinase cascade pathways that end at five different MAP kinases, which can enter the nucleus and phosphorylate many different resident nuclear proteins. In mammals, more than 10 MAP kinases are already known, and more are suggested from genomic sequence data (*42, 56*). Prominent among the many possible gene targets are genes containing in their DNA so-called AP-1 sites that bind dimers of c-Jun, c-Fos, and ATMs among other proteins.

IIB(3)(b). Latent cytoplasmic factors. The hallmark of this group of transcription factors is residence in an inactive form in the cytoplasm until they are activated by proteins that bind to cell surface receptors. Because the intracellular biochemistry that activates proteins in this group varies widely, these factors are not often thought of together. However, they all share the property that a highly specific polypeptide-receptor interaction at the cell surface is succeeded by delivery of a now-activated transcription factor to the nucleus. The numbers of proteins discovered to act in each of the listed pathways (Figs. 1 and 2) have increased in animal evolution, but most of the pathways are present in invertebrates, which suggests that the majority of such pathways may now be known. However, these proteins and the receptors that activate them are not present in fungi or plants, hinting at an important evolutionary divergence leading to animals.

There are eight currently understood pathways for activation of latent cytoplasmic transcription factors (Fig. 2). Proteins in two of these pathways are directly activated after receptor-ligand interaction by association of the transcription factor with the activated receptor where phosphorylation of the factor occurs. In four of the pathways, the factors undergo a more tortuous route of activation involving cytoplasmic serine phosphorylations and either promotion or inhibition of proteosomal proteolysis. Finally, there are two pathways where fluctuations in either Ca^{2+} ions or phosphoinositides cause transcription factors to be released and move to the nucleus. We discuss each of these latter pathways briefly to illustrate that a wide variety of cell surface signals from liganded receptors result in nuclear delivery of one or more particular transcription factors.

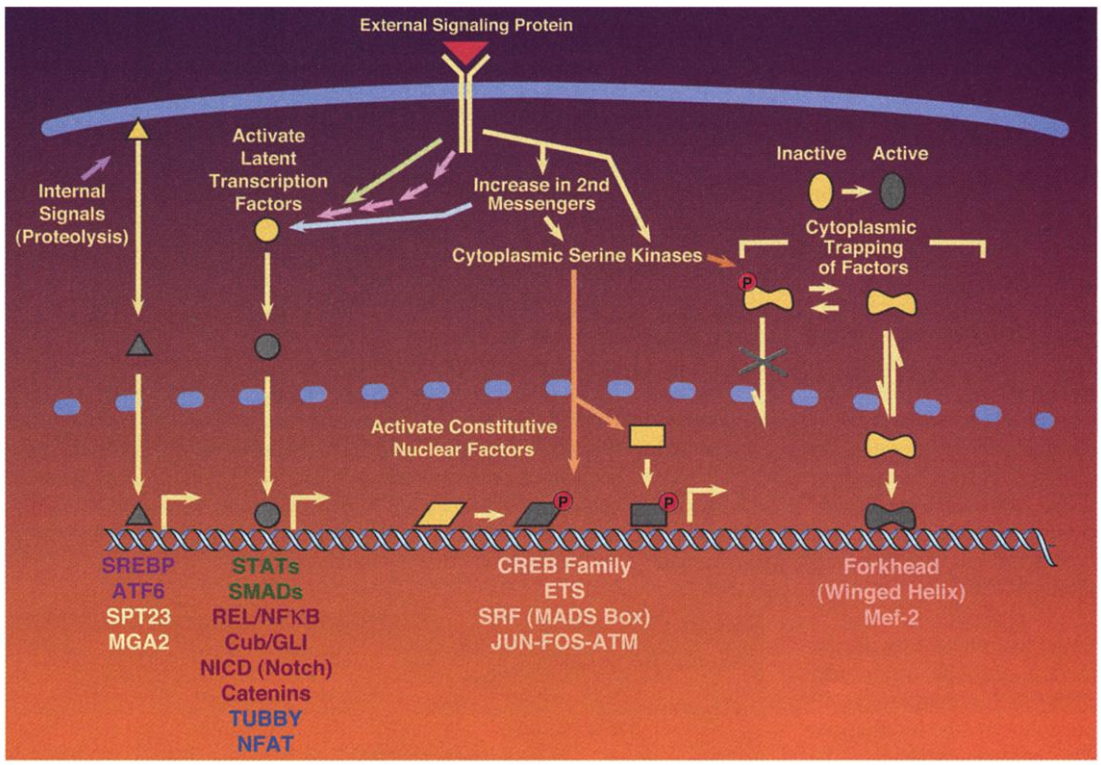

Fig. 2. General scheme of action of signals from extracellular signaling proteins through cell surface receptors to the nucleus. Latent cytoplasmic factors can be activated in a variety of ways [directly by phosphorylation (green), through regulated phosphorylation or proteolysis (light purple), and through second messenger fluctuations (light blue)]. Internal signals regulate proteolysis at the plasma membrane to liberate some transcription factors (dark purple). Resident nuclear proteins (▱, □, yellow) can be phosphorylated on serine and thus activated either when bound to DNA or free in the nucleus.

Direct Activation of Latent Transcription Factors at the Cell Surface Receptor

Serine phosphorylation in the SMAD pathway. The SMAD proteins are the only known latent cytoplasmic transcription factors that become directly activated by serine phosphorylation at their cognate receptors. This family of transcription factors transduce signals on behalf of the transforming growth factor–β (TGF-β) superfamily of ligands (*57*) (Fig. 3).

There are three types of SMADs, most extensively studied in mammalian cells. The effector SMADs (also called the R-SMADs) become serine-phosphorylated in the COOH-terminal domain by the activated receptor. Smad1, Smad5, Smad8, and Smad9 become phosphorylated in response to bone morphogenetic protein (BMP) and growth and differentiation factor (GDF), and Smad2 and Smad3 become phosphorylated in response to the activin/nodal branch of the TGF-β pathway. The second group in this family are called regulatory or co-SMADs. There are two regulatory SMADs: Smad4 and Smad4β (also called Smad10) (*58, 59*). Smad4 binds to, and is essential for, the function of Smad1 and Smad2 (*60*). The regulatory Smad4 binds to all effector SMADs in the formation of transcriptional complexes (*59, 60*), but it does not appear to be required for nuclear translocation of the effector molecules (*61*). Finally, two inhibitory SMADs, Smad6 and Smad7,

provide negative regulation of the pathway (*62*) by blocking Smad4 binding.

Once an activated, serine-phosphorylated effector SMAD binds Smad4 and escapes the negative influences of Smad6 and Smad7, nuclear accumulation and regulation of specific target genes can occur. In most cases, SMADs require partner transcription factors with strong DNA binding capacity that determine the gene to be activated. The DNA binding is then strengthened by association with SMADs that on their own bind weakly to adjacent DNA sites. The SMADs furnish transcriptional activation capacity. The specificity of response among different ligands can be partially explained by the choice of DNA binding partner proteins. For example, activin activation of SMADs results in combinations with FAST1 and a particular set of genes is activated. Signaling by BMP ligands results in association of activated SMADs with a DNA binding protein called OAZ (*62*).

Tyrosine phosphorylation of the STATs. The STATs (signal transducers and activators of transcription) are the only known transcription factors that become activated from a latent state by phosphorylation on tyrosine (*63, 64*) (Fig. 3). They are activated by more than 20 different cytokines, the receptors for which are associated with Janus kinases (JAKs) that tyrosine-phosphorylate the liganded receptor and then the associated STAT. The STATs can also be activated by RTKs such as epidermal growth factor (EGF), platelet-derived growth factor, and Eyk; by non-RTKs such as Src and Abl; and through GCRPs (*65, 66*). Tyrosine phosphorylation of a STAT is followed by dimerization through reciprocal SH2 (SRC homology 2) phosphotyrosine interaction. The precise mechanism of STAT entry into the nucleus remains unknown. [One and perhaps two export signals have been described (*67, 68*), but a nuclear localization signal has not been discovered.] Nevertheless, accumulation of tyrosine-phosphorylated dimers in the nucleus—most probably through binding to an importin (*69*)—allows site-specific DNA binding, and participation in gene control follows. The transcriptional impact of activated STATs is regulated by constitutive and induced proteins (*63, 64*). These include tyrosine phosphatases and the SOCS and PIAS proteins (*70–72*). The natural activation-deactivation cycle of STAT molecules is quite short, about 15 min for an individual molecule (*73*). There are seven known STATs in mammals, each of which has separate in vivo functions, as revealed by knockout experiments (*63, 64*).

Activation of Latent Transcription Factors That Require Cytoplasmic Serine Phosphorylation and/or Proteolysis

A number of important transcriptional regulatory pathways governed by extracellular protein receptor interaction require proteolysis before the delivery to the nucleus of a transcription factor. In turn, many of the proteolytic steps appear to be governed by cytoplasmic serine phosphorylation.

The Rel/NFκB family. The nuclear factor κB (NFκB)/Rel family of proteins has been studied extensively in *Drosophila* (*22*) and is conserved in vertebrates (*74, 75*). There are five Rel/NFκB transcription factors in mammals: NFκB1, NFκB2, c-Rel, Rel-a, and Rel-b. This family of factors can be activated by a large array of extracellular products, including tumor necrosis factor–α, interleukin-1, growth factors, bacterial and viral infections, oxidative stress, and a variety of pharmaceutical compounds (*74*) (Fig. 3). *NFκB1* cDNA encodes a larger protein (p105) that is cleaved to generate a DNA binding molecule of 50 kD. A previously sequenced retroviral oncogene, *vRel*, and its cellular counterpart, *cRel*, had a similar sequence to the DNA binding region of NFκB, and soon four other mammalian Rel/NFκB family members were recognized (*74, 75*).

A cytoplasmic inhibitor of active NFκB called IκB binds to a subunit of NFκB (*74–76*). The IκB protein has ankyrin repeats that bind to the actin cytoskeleton, both tethering the bound NFκB in the cytoplasm and blocking the nuclear localization signal of NFκB. Two serine residues in IκB when phosphorylated lead to destruction of IκB by proteosomes (*75, 76*), with NFκB then moving to the nucleus. Thus, two proteolytic events—the cleavage of p105 to p50 and the destruction of IκB governed by serine phosphorylation—are required to produce the proteins and then activate them. At present it is not clear how long an NFκB molecule continues to function once it is in the nucleus. However, IκB is a target gene of NFκB. Thus, an increase in IκB may help to again sequester NFκB in the cytoplasm (*77*). Finally, it seems likely that different serine kinases are responsible in different cells in the physiologic activation of NFκB (*74–76*).

The CI (GLI) pathway. Signaling by Hedgehog (Hh), a lipid-anchored cell surface ligand (*78*), in *Drosophila* ultimately inhibits the proteolytic cleavage of the transcription factor Ci, cubitus interruptus (*79*). In the absence of Hedgehog signaling, the DNA binding CI protein is cleaved to a smaller molecule that retains nuclear localization and DNA binding properties but has lost its transactivation domain, and hence it acts as a repressor of Hh target genes (Fig. 3). Cleavage of CI is dependent on a PKA phospho-

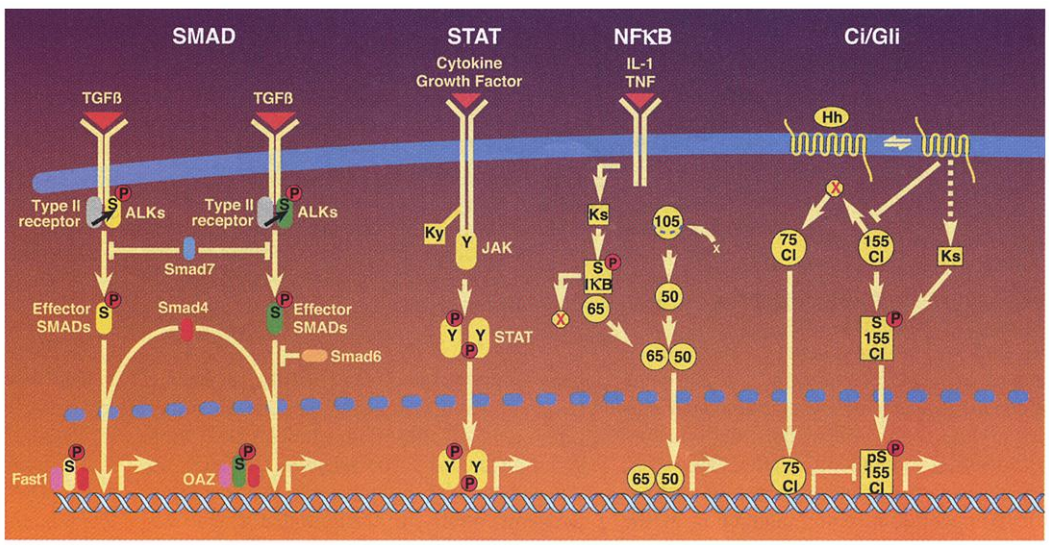

Fig. 3. Eight major signaling pathways that deliver an active transcription factor to the nucleus in response to cell surface receptor binding by an extracellular protein ligand. These factors are activated by polypeptide ligand-receptor interaction at the cell surface and then accumulate in the nucleus to regulate the transcription of target genes. From left to right are the SMAD pathway activated through serine kinases subsequent to binding ligands of the TGF-β superfamily with two branches representing the pathway used by activin/nodal group (left) and BMP/GDF subgroups (right); the STAT pathway transducing signals in response to a large number of ligands including cytokines (illustrated here), growth factors, and charged small molecules that trigger tyrosine (Y) kinases; Rel/NFκB activated by relatively few extracellular proteins but likely involving a number of different intracellular kinases in different cell

rylation (*80*). Inhibition of this proteolytic cleavage preserves the full-length (155-kD) activator and allows Hh-responsive genes to be expressed. There are at least three different homologs of Ci in humans, called Gli because of their presence in gliomas. Mutations in chickens and humans that upset this balance of the long and short forms of GLI proteins affect various developmental events (*80*).

The Wnt pathway. The Wnt proteins (more than 30 in mammals) are extracellular signaling molecules that bind receptors of the Frizzled family (*81*). The first known intracellular step in Wnt signaling operates through a protein called Disheveled (Dsh) (Fig. 2). Although the exact role of Dsh is unclear (*82*), cytoplasmic events controlling the final delivery into the nucleus of a transcriptionally active protein involve both proteolysis and phosphorylation.

There are at least three Wnt-activated pathways downstream of Dsh, only one of which is currently known to result in accumulation of a transcription factor in the nucleus (*83, 84*). In this "canonical" pathway, Dsh is thought to become phosphorylated through the GCRP transmembrane activity generated by the Wnt-Frizzled interaction; this inhibits the kinase, glycogen synthesis kinase–3 (GSK-3). GSK-3 activity maintains a proteolytic cascade that prevents any nuclear accumulation of β-catenin and γ-catenin (*85*). β-catenin binds cytoplasmic structural protein, and excess binds to a protein encoded by the *APC* (adenomatous polyposis coli) oncogene. In the absence of Wnt signaling, GSK-3 phosphorylates both APC and β-catenin, targeting the catenin for destruction by the proteosome. When Wnt binds Frizzled, activated Dsh blocks the GSK-3 phosphorylation and subsequent proteolysis of β-catenin. Cytoplasmic β-catenin levels rise, and because the protein has a nuclear localization signal, it enters the nucleus, where it participates in gene activation.

These events are now firmly established for both β-catenin and γ-catenin (earlier known as plakoglobin) (*85*). The catenins do not bind DNA, although they have NH_2-terminal transactivation domains that bind the coactivating p300/CBP histone transacetylases. However, the catenins do bind to DNA binding proteins called TCF/LEF that themselves lack transactivation capacity.

The Notch pathway. The *Notch* gene encodes a single-pass membrane protein that upon appropriate cleavage ultimately delivers a transcription factor to the nucleus (*86*). Shortly after synthesis while in the secretory pathway, NOTCH is cleaved in the extracellular domain, with the two products remaining associated. An extracellular product (Notch EC), presented outside the cell, remains associated with a transmembrane product (Notch-TM) that is mainly intracellular. Delta and Serrate are both *Drosophila* ligands for the Notch EC. It is thought that the extracellular domains of Delta/Serrate, which are membrane proteins in adjacent cells, contact the Notch receptor in their EGF-like extracellular segments (Fig. 2). The liganded Notch EC then induces a proteolytic cascade, resulting in two additional closely spaced cleavages in the Notch-TM portion of the Notch receptor. The last (third) cleavage releases the Notch intracellular domain (NICD), which is then free to enter the nucleus. NICD itself does not bind DNA but acts as a partner to the DNA binding helix-loop-helix proteins in both *Drosophila* and mammals (*87*). There are a variety of transcriptionally active Notch-like proteins in mammalian cells, where gene activation by a Notch-like fragment (a non–DNA binding protein) cooperates with helix-loop-helix proteins bound to DNA.

Activation of Latent Cytoplasmic Transcription Factors by Fluctuations in Second Messengers

NFAT activation and Ca^{2+} increase. An important subdivision of the NFκB/Rel transcription factor family based on sequence similarity is the NFAT (nuclear factors in activated T cells) proteins (*88, 89*), but their regulation is by an entirely different mechanism (Fig. 3). The cytoplasmic NFAT molecules are heavily phosphorylated in resting cells, but binding of cell surface receptors (the T cell receptor by a cognate surface-expressed immunoglobulin on a B cell) causes a cyclical fluctuation in internal Ca^{2+} concentration. The increase in Ca^{2+} concentration activates the phosphatase calcineurin, which dephosphorylates NFAT, leading to an accumulation of NFAT in the nucleus (*89, 90*). More than 10 different NFAT proteins are expressed in a variety of different tissues (*89*). The NFATs represent the only known case of a cell surface protein interaction resulting in an internal Ca^{2+} ion increase that triggers activation of specific latent cytoplasmic transcription factors. The NFAT proteins have a weak affinity for DNA and usually associate with other factors such as AP-1 (Fig. 3).

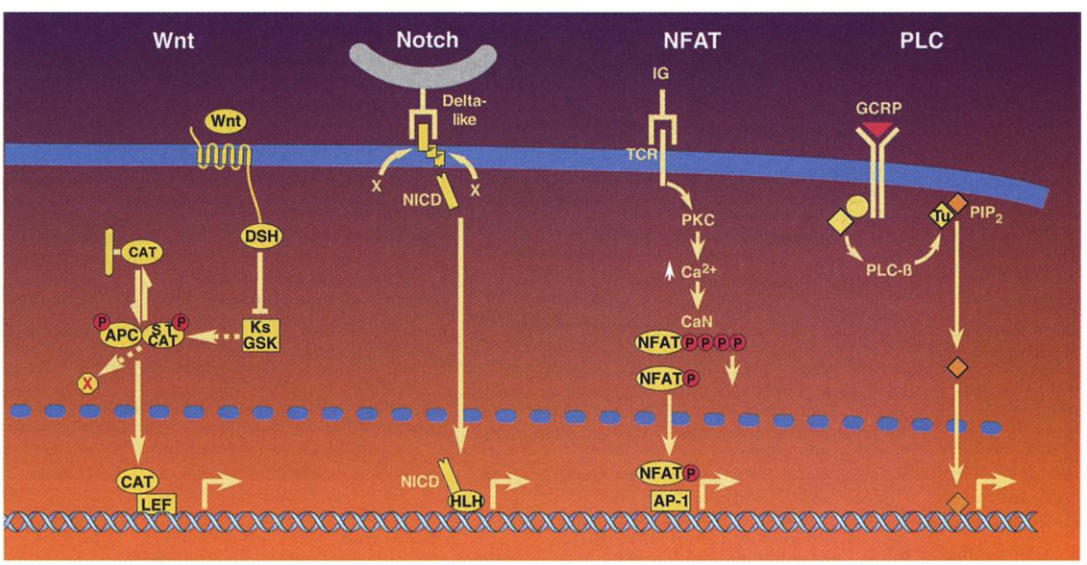

Fig. 3. (continued)—contexts; the Ci/Gli pathway activated by the Hh family of ligands; the canonical pathway activated by the Wnt/Wg ligands, resulting in catenin accumulation in the nucleus; and the Notch pathway activated by membrane-associated ligands (either Delta or Delta-like proteins), resulting in proteolysis releasing the intracellular domain of Notch (NICD) that then accumulates in the nucleus. The final two examples require second messenger triggering. NFATs are imported into the nucleus after dephosphorylation dependent on Ca^{2+} increase that activates the phosphatase calcineurin. Finally, in the phospholipase C (PLC) pathway, Tubby is released from PIP_2 binding by activation of phospholipase that cleaves PIP_2, leaving Tubby protein free to enter the nucleus. Abbreviations: SP, serine phosphate; Ky, tyrosine kinase; CAT, catenin; CaN, calcineurin; Tu, Tubby; Ks, serine kinase; X, proteosome; HLH, helix-loop-helix; TCR, T cell receptor; PKC, protein kinase C; ALKs, activin receptor-like kinases; pS, phosphoserine; IG, immunoglobulin.

Phosphoinositide-dependent release of Tubby. A large number of cell surface ligands affect intracellular signaling through changes in phosphoinositide concentration. Recently the first such change linked directly to the activation of a latent cytoplasmic transcription factor was described (*91*) (Fig. 3). The

Tubby protein, the product of a gene that when mutant causes obesity, is anchored to the plasma membrane by affinity to phosphatidylinositol 4,5-bisphosphate (PIP_2). A ligand binding to a GCRP at the cell surface causes an activation of phospholipase that cleaves PIP_2, releasing Tubby to enter the nucleus. Proof that Tubby is a transcription factor is incomplete because no target genes are yet known, but the protein definitely binds DNA and has transcription activation potential (*91*).

Conclusions

In this summary of the salient features of transcription factors, we have concentrated on those that receive signals in the wake of polypeptide receptor interaction at the cell surface because they are crucial in metazoan development and in adult function. We suggest that the evolutionarily earliest major regulatory circuits involved serine phosphorylation of resident nuclear factors, and that the duplication of serine kinases provided increasingly complex effects on nuclear function. The proper regulation of genes in fungi and plants is dependent on these pathways. Animals have not lost this dependence, but indeed have amplified the number of serine kinases crucial for proper balanced transcription (*41, 42*). Dozens (if not hundreds) of resident nuclear proteins depend on serine phosphorylation for maximal activity and are often bound to DNA without phosphorylation. During regulated transcription, these proteins often cooperate with other regulatory proteins. It is tempting to speculate that these factors modulate the rate of transcription but may not act as on-off switches. The second large class of polypeptide responsive factors are the latent cytoplasmic factors.

It seems possible that the initial trigger for cell-specific transcription in increasingly complex animals often comes from the two groups of regulatory transcription factors. These are (i) the factors synthesized in sequence as development proceeds (termed "developmental factors" in Fig. 1), and (ii) the factors activated by extracellular signals, chiefly the steroid receptor superfamily and the latent cytoplasmic transcription factors activated by extracellular polypeptides. According to the dictates of what best contributed to organism survival and selection, individual genes evolving in this mix of possible regulatory proteins have each individually acquired binding sites to accommodate an assortment of all of the different types of transcription factors—thus, the appearance of multiple proteins in an enhanceosome. So it should come as no surprise when complicated developmental events, such as eye development or wing development in flies (*92*) or cell choice in segmenting embryos or pluripotent bone marrow cells, are affected by mutations in multiple pathways. Individual genes require multiple pathways; why shouldn't developmental decisions?

References and Notes

1. S. Malik, R. G. Roeder, *Trends Biochem. Sci.* **25**, 277 (2000).
2. A. M. Naar, B. D. Lemon, R. Tjian, *Annu. Rev. Biochem.* **70**, 475 (2001).
3. A. M. Naar, S. Ryu, R. Tjian, *Cold Spring Harbor Symp. Quant. Biol.* **63**, 189 (1998).
4. K. A. Jones, J. T. Kadonaga, *Genes Dev.* **14**, 1992 (2000).
5. T. Jenuwein, C. D. Allis, *Science* **293**, 1074 (2001).
6. Special issue on The Human Genome, *Science* **291** (16 February 2001).
7. Special issue on The Human Genome, *Nature* **409** (15 February 2001).
8. R. Grosschedl, *Curr. Opin. Cell Biol.* **7**, 362 (1995).
9. D. Thanos, T. Maniatis, *Cell* **83**, 1091 (1995).
10. F. Jacob, J. Monod, *J. Mol. Biol.* **3**, 318 (1961).
11. E. Maldonado, M. Hampsey, D. Reinberg, *Cell* **99**, 455 (1999).
12. R. L. Smith, A. D. Johnson, *Trends Biochem. Sci.* **25**, 325 (2000).
13. V. Pirotta, *Cell* **93**, 333 (1998).
14. J. Torchia, C. Glass, M. G. Rosenfeld, *Curr. Opin. Cell Biol.* **10**, 373 (1998).
15. A. J. Courey, S. Jia, *Genes Dev.* **15**, 2786 (2001).
16. M. R. Briggs, J. T. Kadonaga, S. P. Bell, R. Tjian, *Science* **234**, 47 (1986).
17. L. A. Chodosh *et al.*, *Cell* **53**, 25 (1988).
18. P. F. Johnson, S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1989).
19. P. J. Rosenfeld, T. J. Kelly, *J. Biol. Chem.* **261**, 1398 (1986).
20. W. Driever, C. Nusslein-Volhard, *Nature* **337**, 138 (1989).
21. K. G. Xanthopoulos, J. Mirkovitch, T. Decker, C. F. Kuo, J. E. Darnell Jr., *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4117 (1989).
22. D. S. Johnston, C. Nusslein-Volhard, *Cell* **68**, 201 (1992).
23. R. Rivera-Pomar, H. Jäckle, *Trends Genet.* **12**, 478 (1996).
24. T. Lufkin, *Curr. Opin. Genet. Dev.* **6**, 575 (1996).
25. K. Yun, B. Wold, *Curr. Opin. Cell Biol.* **8**, 877 (1996).
26. S. A. Duncan, M. A. Navas, D. Dufort, J. Rossant, M. Stoffel, *Science* **281**, 692 (1998).
27. F. S. Naya, E. Olson, *Curr. Opin. Cell Biol.* **11**, 683 (1999).
28. A. Brunet *et al.*, *Cell* **96**, 857 (1999).
29. G. J. P. L. Kops *et al.*, *Nature* **398**, 630 (1999).
30. L. Li *et al.*, *Cell* **71**, 1181 (1992).
31. R. M. Evans, personal communication.
32. G. M. Ringold, K. R. Yamamoto, J. M. Bishop, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2879 (1977).
33. D. J. Mangelsdorf *et al.*, *Cell* **83**, 935 (1995).
34. H. Htun, J. Barsony, I. Renyi, D. L. Gould, G. L. Hager, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4845 (1996).
35. S. Mackem, C. T. Baumann, G. L. Hager, *J. Biol. Chem.* **49**, 45501 (2001).
36. M. S. Brown, J. L. Goldstein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11041 (1999).
37. B. M. Forman *et al.*, *Cell* **83**, 803 (1995).
38. S. A. Kliewer *et al.*, *Cell* **83**, 813 (1995).
39. A. J. Levine, *Cell* **88**, 323 (1997).
40. C. Patil, P. Walter, *Curr. Opin. Cell Biol.* **13**, 349 (2001).
41. T. Hunter, in *The Harvey Lectures Series 94, 1998–99*, M. M. Davis, Ed. (Jossey-Bass, San Francisco, 2000), pp. 81–120.
42. ———, *Cell* **100**, 113 (2000).
43. R. J. Lefkowitz, J. M. Stadel, M. G. Caron, *Annu. Rev. Biochem.* **52**, 159 (1983).
44. H. R. Bourne, *Curr. Opin. Cell Biol.* **9**, 134 (1997).
45. W. J. Fantl, D. E. Johnson, L. T. Williams, *Annu. Rev. Biochem.* **62**, 453 (1993).
46. T. Pawson, J. D. Scott, *Science* **278**, 2075 (1997).
47. H. Gille *et al.*, *EMBO J.* **14**, 951 (1995).
48. M. Cavigelli, F. Dolfi, F.-X. Claret, M. Karin, *EMBO J.* **14**, 5957 (1995).
49. H. Gille, T. Strahl, P. E. Shaw, *Curr. Biol.* **5**, 1191 (1995).
50. R. Janknecht, A. Nordheim, *Biochim. Biophys. Acta* **1155**, 346 (1993).
51. R. Janknecht, *Immunobiology* **193**, 137 (1995).
52. B. Wasylyk, J. Hagman, A. Gutierrez-Hartmann, *Trends Biochem. Sci.* **23**, 213 (1998).
53. R. J. Davis, in *Signaling Networks and Cell Cycle Control*, J. S. Gutkind, Ed. (Humana, Totowa, NJ, 2000), pp. 153–164.
54. A. J. Whitmarsh, R. J. Davis, *Trends Biochem. Sci.* **23**, 481 (1998).
55. A. J. Shaywitz, M. E. Greenberg, *Annu. Rev. Biochem.* **68**, 821 (1999).
56. H. J. Schaeffer, M. J. Weber, *Mol. Cell. Biol.* **19**, 2435 (1999).
57. J. Massague, *Nature Rev. Mol. Cell Biol.* **1**, 169 (2000).
58. M. Howell *et al.*, *Dev. Biol.* **214**, 354 (1999).
59. N. Masuyama, H. Hanafusa, M. Kusakabe, H. Shibuya, E. Nishida, *J. Biol. Chem.* **274**, 12163 (1999).
60. G. Lagna, A. Hata, A. Hemmati-Brivanlou, J. Massague, *Nature* **383**, 832 (1996).
61. X. Liu *et al.*, *Genes Dev.* **11**, 179 (1997).
62. A. Hata, G. Lagna, J. Massague, A. Hemmati-Brivanlou, *Genes Dev.* **12**, 186 (1998).
63. J. E. Darnell Jr., *Science* **277**, 1630 (1997).
64. G. R. Stark, I. M. Kerr, B. R. Williams, R. H. Silverman, R. D. Schreiber, *Annu. Rev. Biochem.* **67**, 227 (1998).
65. S. G. Rane, E. P. Reddy, *Oncogene* **19**, 5662 (2000).
66. J. Bromberg, J. E. Darnell Jr., *Oncogene* **19**, 2468 (2000).
67. K. M. McBride, C. McDonald, N. C. Reich, *EMBO J.* **19**, 6196 (2000).
68. A. Begitt, T. Meyer, M. van Rossum, U. Vinkemeier, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10418 (2000).
69. F. Sekimoto, N. Iwamoto, K. Nakajima, T. Hirano, Y. Yoneda, *EMBO J.* **16**, 7067 (1998).
70. R. Starr, D. J. Hilton, *Bioessays* **21**, 47 (1999).
71. C. D. Chung *et al.*, *Science* **278**, 1803 (1997).
72. B. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10626 (1998).
73. R. L. Haspel, J. E. Darnell Jr., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10188 (1999).
74. P. A. Baeuerle, D. Baltimore, *Cell* **87**, 13 (1996).
75. N. D. Perkins, *Trends Biochem. Sci.* **25**, 434 (2000).
76. M. Karin, *Oncogene* **18**, 6867 (1999).
77. T. T. Huang, S. Miyamoto, *Mol. Cell. Biol.* **21**, 4737 (2001).
78. P. W. Ingham, *EMBO J.* **17**, 3505 (1998).
79. N. Methot, K. Basler, *Cell* **96**, 819 (1999).
80. B. Wang, J. F. Fallon, P. A. Beachy, *Cell* **100**, 423 (2000).
81. G. D. Barish, B. O. Williams, in *Signaling Networks and Cell Cycle Control*, J. S. Gutkind, Ed. (Humana, Totowa, NJ, 2000), pp. 53–82.
82. K. Willert, M. Brink, A. Wodarz, H. Varmus, R. Nusse, *EMBO J.* **16**, 3089 (1997).
83. P. Polakis, *Genes Dev.* **14**, 1837 (2000).
84. M. Kuhl, L. C. Sheldahl, C. C. Malbon, R. T. Moon, *J. Biol. Chem.* **275**, 12701 (2000).
85. F. T. Kolligs *et al.*, *Genes Dev.* **14**, 1319 (2000).
86. G. Weinmaster, *Curr. Opin. Genet. Dev.* **10**, 363 (2000).
87. A. Israel, *Nature* **388**, 519 (1997).
88. G. R. Crabtree, *Cell* **96**, 611 (1999).
89. A. Rao, C. Luo, P. G. Hogan, *Annu. Rev. Immunol.* **15**, 707 (1997).
90. H. Okamura *et al.*, *Mol. Cell* **6**, 539 (2000).
91. S. Santagata *et al.*, *Science* **292**, 2041 (2001).
92. M. A. Simon, *Cell* **103**, 13 (2000).

93. We thank D. Benyaklef for assistance with the graphic figures. Supported by NIH grants AI34420 and AI32489 (J.E.D.) and NIH grant 2RO1HD/GM32105-06A1 1RO1EY12370-03 (A.H.B.).