

# Transcription Factor-Based Drug Design in Anticancer Drug Development

**Athanasios G. Papavassiliou**

**Department of Biochemistry, School of Medicine, University of Patras, Patras, Greece**

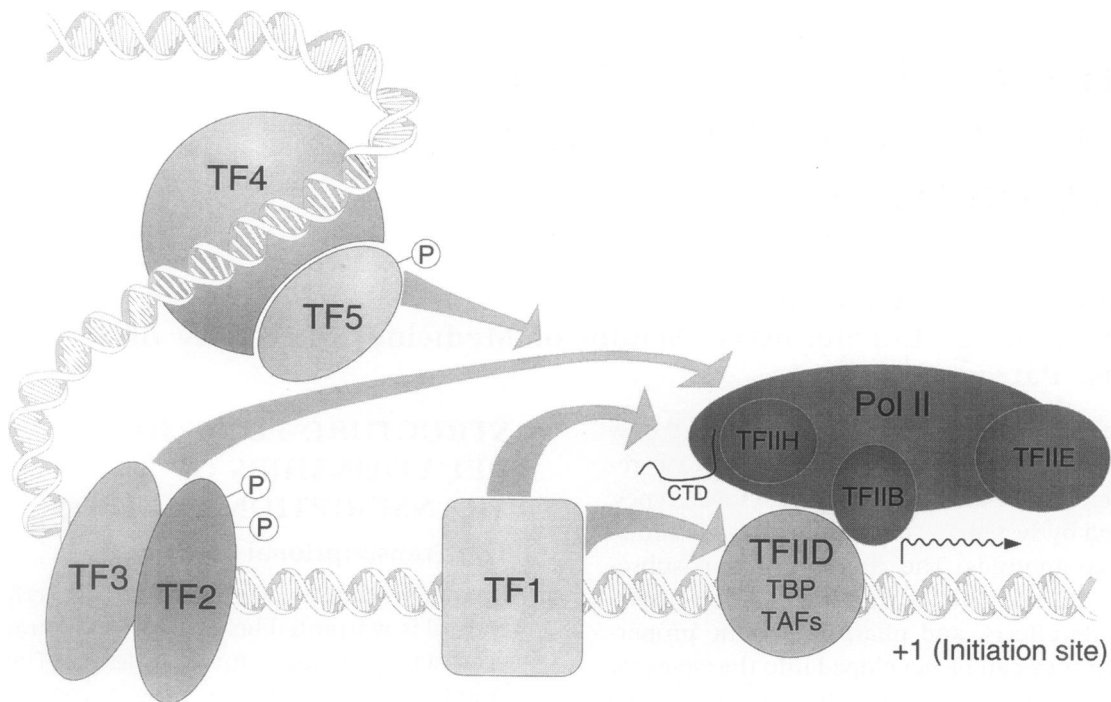
The traditional approach to the discovery of new anticancer drugs involves the use of disease-related screens for a wide gamut of chemical types, followed by systematic chemical manipulation of active compounds. The compounds can subsequently be evaluated for antineoplastic activity, toxic side effects, and pharmacokinetic properties, and they can be developed into therapies for cancer treatment. Increasingly, and in striking contrast, defined selective macromolecular targets are being used as the basis for new drug discovery. The importance of this trend will increase as the results of studies on the molecular basis of oncogenesis are applied to cancer diagnosis and cure. The enormous advances in recombinant DNA techniques over the past decade have led to the identification of key proteins that are intimately involved in the regulation of cancer growth and invasion control at the level of gene expression. In particular, the pivotal role of specific transcription factors in certain cancers, either as mutants or in overexpressed levels, highlights them as rational targets for chemotherapeutic intervention. The large volume of data available on the molecular anatomy of transcription factors and the biochemical pathways that modulate their function offer opportunities for the design of structure-based, small organic molecules targeting oncogenic transcription factors (oncogene or tumor suppressor gene products) selectively, thus creating powerful new pharmaceuticals that inhibit malignant cell growth and tumor metastasis.

Address correspondence and reprint requests to: Athanasios G. Papavassiliou, M.D., Ph.D., University of Patras, School of Medicine, Department of Biochemistry, GR-26110 Patras, Greece. Tel: +30-61-996144; Fax: +30-61-996110; E-mail: papavas@med.upatras.gr

## **STRUCTURE-FUNCTION RELATIONSHIPS OF TRANSCRIPTION FACTORS**

### **The Transcriptional Machinery**

A brief overview of eukaryotic transcriptional control is warranted before we discuss transcription factors as targets for drug design. The regulation of transcription of protein-coding genes involved in normal cellular metabolism or correlated with cancer development and progression is achieved by an ensemble of proteins whose central component is the enzyme RNA polymerase II (pol II). Although pol II has an intrinsic capacity to synthesize RNA, it is unable to efficiently recognize gene promoters or transcription initiation sites and accurately initiate transcription on its own. To accomplish this, it requires the close collaboration of a battery of accessory proteins, collectively termed transcription factors. Transcription factors are generally divided into two groups: the basal transcription factors and the gene-specific transcription factors (1). Basal transcription factors are ubiquitous, DNA- or non-DNA-binding proteins essential for the transcription of all protein-coding genes, which recruit and align pol II at the core promoter region of the gene. This region encompasses the so-called TATA box (a recognition sequence around which basal transcription factors/pol II amalgamate) and the transcription initiation site (Fig. 1) (2). Gene-specific transcription factors, which are only required for a subset of genes transcribed by pol II, recognize and bind to a second promoter region (typically composed of multiple short—6–12 bp—*cis*-acting DNA sequence elements) located within a few hundred base pairs upstream from the transcription initiation site or positioned many kilobases away (in which case it is called enhancer) (1). These control regions may bind numerous, different gene-



**FIG. 1. Multiple interactions between the transactivation domains of DNA-bound transcription factors and RNA polymerase II (Pol II)/basal transcription factors facilitate gene transcription**

In the model depicted here a number of potential mechanisms are illustrated whereby DNA, gene-specific transcription factors (TF1-5) and Pol II/basal transcription factors (TFIIB, TFIID-TBP-TAFs, TFIIE, and TFIIH are the only ones shown here) are brought together to form an active transcription complex. TF1 binds close to the transcription initiation site and is able to communicate directly with components of the basal transcriptional machinery. TF2 and TF3 bind DNA as a heterodimer to a sequence distant from the transcriptional start, thus making indirect (via protein-protein interactions with TF1, for example) contacts with Pol II and associated basal transcription factors. Heterodimerization and/or the transactivation potential of TF2 are subject to regulation by inducible phosphorylations. TF4 interacts with the basal transcriptional machinery via a second, phosphorylation-modulated protein, TF5, that does not itself bind to DNA. The intervening DNA loops out to accommodate the interaction. CTD, Pol II carboxy-terminal domain; TBP, TATA box-binding protein; TAFs, TBP-associated factors; P, phosphoryl group.

specific transcription factors, each of which acts, either positively or negatively, to influence transcription initiation and rate (Fig. 1). Cross-talk between DNA-bound, gene-specific transcription factors may also have synergistic or antagonistic effects on transcription kinetics. Thus it is the orchestrated action of these factors binding at a given promoter/enhancer that ultimately determines the spatial and temporal expression patterns of genes during development and differentiation, and the subsequent homeostatic regulation of cellular metabolism under the regime of a variety of environmental cues and intracellular signals (1).

### Principles of Gene-Specific Transcription Factor Structure

Gene-specific transcription factors are composed of functionally distinct, independently folding

domains (1). One domain, termed the DNA-binding domain, recognizes and binds to specific DNA sequences within promoter/enhancer elements of genes. This domain typically includes a helical unit ( $\alpha$ -helix) within or adjacent to positively charged (basic) amino acids. Four classes of structural protein motifs characterize 80% of transcription factor DNA-binding domains (3-5):

1. In the *zinc finger* a small group of conserved amino acids (including a pair of cysteines and a pair of histidines with characteristic spacing) acts together to coordinate a zinc ion. The motif takes its name from the loop of amino acids that protrudes from the zinc-binding site. Zinc fingers may form  $\alpha$ -helices that insert into the major groove of DNA, and they are usually organized as a single series of tandem repeats. A distinct form

- of the motif is found in the steroid-thyroid-retinoid nuclear hormone receptors, where its presence confers specificity to both DNA binding and dimerization.
2. In the *helix-turn-helix* (*HTH*) binding to DNA is mediated by alignment of each of the two  $\alpha$ -helices of the motif within and alongside the grooves of the DNA helix. A related form of the motif is present in the homeodomain, a sequence first characterized in several regulatory proteins concerned with the execution of particular differentiation and developmental programs in the organism.
  3. The *basic region leucine zipper* (*bZIP*) consists of a periodic array of leucine residues, at every seventh position, along an  $\alpha$ -helix. A leucine zipper in one polypeptide interacts with a zipper in another polypeptide forming an  $\alpha$ -helical "coiled-coil," which serves to align the DNA-contacting motifs (adjacently nested stretches of basic amino acids) of two interacting proteins.
  4. The *basic region helix-loop-helix* (*bHLH*) is comprised of two amphipathic  $\alpha$ -helices, each of which presents a face of hydrophobic residues on one side and charged residues on the other side. The motif enables proteins to dimerize, and a basic region near this motif contacts DNA.

As a rule, DNA-binding domains mediate unspecific or "positioning contacts" that provide a general, moderate affinity to DNA and help fix the orientation of the whole structural element, and base-specific contacts that ensure high-affinity binding to specific target sequences (6). Positioning contacts are by and large interactions with the deoxyribose-phosphate backbone of the DNA and frequently involve electrostatic attractions between positively charged amino acid side chains and negatively charged phosphate groups (hence the high frequency of lysine and arginine residues in many DNA-binding domains). Base specificity is governed by hydrogen bonds and van der Waals contacts between amino acid side chains of the binding domain and the exposed chemical groups on the edges of the base pairs in the DNA target sequence (6).

High-resolution atomic structures for several members of the above (as well as additional) classes, developed from either X-ray diffraction analysis of single crystals and/or nuclear mag-

netic resonance (NMR) spectroscopy, have been solved. A thorough description of the structure, function, and evolution of transcription factor DNA-binding domains can be found in Pabo and Sauer, and Ouzounis and Papavassiliou (7,8).

Aside from their DNA-binding domain, transcription factors may have domains for reversible, noncovalent protein-protein interactions called dimerization domains (often closely juxtaposed to the DNA-binding domain). Interactions between one specific protein via a similar dimerization domain on another protein is a prerequisite for binding of some transcription factors to DNA (see *bZIP* and *bHLH* DNA-binding motifs above). Therefore the binding of transcription factors to DNA is often cooperative, that is, the presence of one protein can change the affinity of other proteins for binding to DNA. Many transcription factor dimers contain two of the same protein species (homodimers), while others are formed between two nonidentical proteins (heterodimers) (Fig. 1). Consequently, different transcription factor homodimers and heterodimers may recognize the same or related DNA sequences and this allows the generation of combinatorial diversity in the regulation of transcription (1).

The DNA-binding and dimerization domains of transcription factors are necessary but not sufficient for transcriptional activity. Yet another domain, the so-called transactivation (or, in the case of negatively acting factors, transrepression) domain, rich in negatively charged (acidic) amino acids or glutamine/proline residues, is required for interacting (directly or indirectly) with one or more components of the transcriptional apparatus (pol II and basal transcription factors) and thus facilitating (or inhibiting) transcription initiation from a given gene promoter (Fig. 1) (1,9). Several transcription factors appear to have multiple transactivation domains that are of more than one type. Detailed information about the tertiary structure of these domains is difficult to obtain, as they are loosely ordered in solution and adopt a rigid conformation only when they contact their appropriate target within the transcription complex.

### Regulation of Transcription Factor Activity

The function of any of the aforementioned transcription factor domains may be subject to regulation by reversible, covalent modifications induced by a broad spectrum of physical and

chemical stimuli (e.g., mechanical forces, osmotic stress, ultraviolet light, growth factors/mitogens, cytokines, hormones, etc.). Among them, protein (de)phosphorylation at specific sites is the post-translational modification of choice when rapid modulation of transcription factor activity in response to changes in environmental conditions, metabolic activity, and growth signals is required (10). Alterations in the phosphorylation state of transcription factors may affect their function either positively or negatively by eliciting conformational changes that expose, mask, or remodel a particular domain/region of the protein (11). Since phosphorylation of an acceptor amino acid (serine, threonine, or tyrosine) changes its charge to negative, a decrease in the phosphorylation of the DNA-binding domain would increase its net positive charge and thus enhance the interaction of the DNA-binding domain with the negatively charged phosphodiester backbone of the DNA duplex. On the other hand, increased phosphorylation of amino acids within or in the vicinity of the dimerization or the transactivation (transrepression) domains could augment the ability of the transcription factor to homo- or heterodimerize or stimulate (inhibit) transcription (Fig. 1) (11,12). Phosphorylation can also affect the function of transcription factors in other fashions. It is now well established that the subcellular localization and stability of several transcription factors (thus their steady-state level) are subject to regulatory influences by phosphorylation events occurring at specific parts of the molecule; these places usually overlap with other defined domains of the transcription factor (12,13).

Ligand binding is another mode of transcription factor activation that is typical for the large superfamily of steroid-thyroid-retinoid nuclear hormone receptors. Upon binding its ligand (hormone), a typical steroid receptor activates expression of particular target genes by binding to its specific response element in a promoter or enhancer. Finally, a variety of different (cytoplasmic and nuclear non-DNA-binding) proteins interacting specifically with transcription factors may control their activity in indirect ways, adding another dimension to the regulatory repertoire and signal integration. Tethering of such proteins to transcription factors may exert a diverse range of functions, e.g., serving as a bridge between the transcription factor and the basal transcriptional machinery or unrelated transcription factors; stabilizing the DNA-bound form;

changing the specificity of the target recognition sequence; sequestering the factor in an inactive complex (dissociation of the complex by signal-dependent (de)phosphorylation of the anchor protein allows translocation of the factor to the nucleus); enhancing its degradation, etc. (13).

A single transcription factor may be regulated by one or more systems at multiple steps along the way to transcription activation. There is probably a hierarchy in the importance of the control steps for a particular transcription factor, some of which may mainly serve for "fine tuning" by coupling to other regulatory pathways.

## ONCOGENIC TRANSCRIPTION FACTORS

Cancer often results from the aberrant activation of specific genes known as oncogenes, which encode components of the cellular machinery that regulates normal growth processes (14,15). Either overexpression of these genes or mutations that lead to the formation of a more active product can result in deregulated control of cellular proliferation and conversion to the malignant phenotype (16). Considering their importance in the control of cell behavior, it is not unexpected that a large group of oncogenes (currently one-third of the known oncogenes) have been found to encode transcription factors (or nuclear proteins that regulate transcription factors) engaged in the expression of genes whose products are required to initiate the cascade of events that lead to passage through the cell cycle (17). Moreover, a number of developmental decisions as well as the execution of coordinated programs of differentiation of different specialized cell types are under the control of oncogenic transcription factors, which are often the final targets (integration centers) of signal transduction pathways (18). The concentration and activity of oncogenic transcription factors are normally tightly controlled at several points (and by one or more systems) down the pathway to their ultimate action in gene transcription, and they can be greatly affected by a wide spectrum of extracellular signals. Several of these transcription factors can themselves function as oncoproteins (5,13,16,19).

### Mechanisms of Transcription Factor Oncogenicity

The role of oncogenic transcription factors in contributing to malignancy by altering programs

of cell growth, differentiation, and development is most clearly demonstrated in the case of human leukemias. Abnormal expression or inappropriate activation of specific oncogenes encoding transcription factors cause a variety of leukemias (but also other tumor types), which are characterized by genetic rearrangements induced by particular chromosomal translocations and gene amplifications (14,20,21). In the simplest case, the translocation results in the relocation of the gene encoding the transcription factor to a position where it is adjacent to a highly expressed gene (i.e., an immunoglobulin gene in B-cell leukemias or a T-cell receptor gene in T-cell leukemias). Under the influence of the transcriptional activity of such loci, the transcription factor becomes constitutively expressed and does not respond to the signals that normally regulate its expression. A second form of damage inflicted by chromosomal translocations on oncogenes results when the breakpoints of the translocations fuse portions of two genes together in a manner that creates a hybrid protein comprised of domains derived from both genes. Such rearrangements presumably result in leukemia because the hybrid protein exhibits properties distinct from those of either protein alone that render it capable of transforming the cell into a malignant phenotype. Translocations of this type in human leukemias frequently involve fusion of a transcription factor–coding gene with either a gene that does not encode a transcription factor or with a different transcription factor–coding gene, generating in both cases a novel, oncogenic fusion protein with abnormal biochemical properties (22–24). These molecular “cut-and-paste” maneuvers give rise to hybrid transcription factors that possess the DNA-binding specificity of one parental protein and the activation characteristics of another. In some instances it may be necessary for only one of the genes to contribute a biochemical activity (DNA binding, transactivation) to the hybrid protein. The contribution from the other gene may be passive, such as disruption or replacement of a regulatory domain/region.

The *myc* oncogene (which encodes a bHLH/ZIP transcription factor, which, together with its partner Max, is required for cell proliferation, prevention of differentiation in response to mitogenic stimuli, and induction of apoptosis) was identified as a site of chromosomal translocation in Burkitt’s lymphomas in humans (where its expression is dramatically increased) and as an amplified locus in some human tumors (17,25).

The Myc protein family (including c-Myc, N-Myc, and L-Myc) serves as an archetype for the activation of an oncogenic transcription factor by chromosomal rearrangement. This model has since been successfully applied to the study of genes that are associated with chromosomal translocations in human leukemias, lymphomas, and several solid tumors. The characterization of transcription factor–coding oncogenes that are rearranged in chromosomal translocations includes the following (17,18,26, and references therein): *hox-11* (which encodes a homeodomain-bearing transcription factor whose expression is activated by translocation to the T-cell receptor locus in cases of acute childhood T-cell leukemia); *tal-1* (which encodes a bHLH transcription factor whose expression is stimulated in acute lymphoblastic leukemia); *bcl-3* (a member of the I $\kappa$ B family, which interacts with the NF $\kappa$ B transcription factors [see below], whose aberrant expression is involved in some B-cell chronic leukemias); *ets* (encoding the Ets transcription factor, which is fused to the platelet-derived growth factor receptor in patients with chronic myelomonocytic leukemia); *erg* (an *ets*-related gene, which is activated by translocation in human myeloid leukemias); *fli-1* (another *ets*-related gene, which is fused to the transcription factor–coding gene *ews* in Ewing’s sarcomas); the retinoic acid receptor  $\alpha$  (*RAR $\alpha$* ) gene (which encodes a transcription factor of the steroid-thyroid-retinoid nuclear hormone receptor gene superfamily, which is fused to the zinc finger-containing PML transcription factor in all cases of promyelocytic leukemia); *E2A* (which encodes a bHLH transcription factor, which is fused to the homeodomain-containing Pbx-1 transcription factor in pre-B-cell leukemias); *aml-1* (involved in a chromosomal translocation, which fuses it to the transcription factor–coding gene *mtg-8*, present in a large fraction of cases of acute myeloid leukemia); *pax-3* (which participates in dermomyotome formation during development and is fused to the gene encoding the FKHR transcription factor in alveolar rhabdomyosarcomas); *hrx/enl* and *hrx/af-4* (fused transcription factor–coding genes generated by translocations in some acute leukemias); and many others. These genes are of special interest because of their direct association with specific human tumors.

Other oncogenic transcription factors whose excessive or inappropriate function is likely to underlie many forms of cancer by disrupting control mechanisms of intracellular signaling pathways, of the cell cycle, and of cell differen-

tiation and developmental programs include the following (17,18, and references therein): Fos and Jun family members (they form the homo- and heterodimeric bZIP transcription factor AP-1 that regulates a variety of genes associated with proliferation and differentiation); members of the Rel/NF $\kappa$ B family of transcription factors (implicated in lymphoid development); ErbA (which encodes the thyroid hormone receptor that normally acts to induce transcription of genes that function in erythroid differentiation, thereby serving as a negative regulator of cell proliferation); Myb (an HTH-containing transcriptional activator involved in the control of hematopoiesis—erythroid lineage); PU.1 (a member of the Ets family of transcription factors involved in the control of hematopoiesis—myeloid lineage); Rbtn2 (which encodes a cysteine-rich LIM domain-containing protein that interacts physically with Tal-1 [see above], generating a protein complex that functions in transcriptional activation during erythroid development); Ski; Qin; Gli; and others.

A distinct class of genes, known as anti-oncogenes or tumor suppressor genes, encode proteins that normally function in a manner opposite that of oncogenes and act to restrain cellular growth. The mutational inactivation or deletion of these genes can therefore result in cancer (15,27). A number of anti-oncogenes of this type have been defined, and four of them (this number continues to grow) encode transcription factors (17,28). Three of these, p53 (which participates in the cellular response to DNA damage and acts as a tetramer), its newly discovered structural relative p73 (acting also as an oligomer), and the Wilms' tumor gene product (WT1, a zinc finger-containing transcription factor), function by binding directly to their target DNA and either up-regulating the expression of growth-inhibitory/apoptosis-promoting genes (p53), or down-regulating the expression of growth-inducing genes (WT1) (17,29,30). More research is definitely needed to determine the role(s) of p73 in cell growth control (31). In contrast, the product of the retinoblastoma susceptibility gene (pRb) exerts its growth-inhibitory effects primarily via protein-protein interactions with other DNA-binding transcription factors (e.g., E2F), preventing them from stimulating the transcription of genes that encode growth-promoting products (e.g., Myc, Myb, thymidine kinase) (17,32,33). Interestingly (for the purpose of this review), association of the human papillomavirus—types 16 and 18—E6

oncoproteins with p53 inactivates the latter (by targeting it for degradation), thus interfering with its function as a tumor suppressor (34). Germ-line mutations of the anti-oncogenes encoding the above transcription factors are responsible for one or more types of inherited cancers (e.g., Li-Fraumeni cancer family syndrome, Wilms' tumor, retinoblastoma), whereas somatic mutations of the same genes appear to play prominent roles in the development of a wide variety of more common sporadic human cancers (various carcinomas, brain tumors, sarcomas, lymphomas, and leukemias) (34).

### ONCOGENIC TRANSCRIPTION FACTORS AS TARGETS FOR DRUG DESIGN

Traditionally, drug discovery programs have long relied upon systematic screening of libraries of naturally occurring products (including plant, fungal, and bacterial extracts) and synthetic chemicals in biological and pharmacological assays that utilize whole animals or isolated tissues. Progress in anticancer drug development has been largely confined to the more classic targets of hormone, DNA and nucleotide metabolism (e.g., dihydrofolate reductase, thymidylate synthase), as well as DNA itself, with almost complete ignorance regarding the precise mechanisms of the underlying molecular machinery involved in their potency. Medicinal chemists and pharmacologists had not ventured into the field of transcription control because they feared that drugs that interfere with the transcriptional apparatus may not be selective or efficacious. The rapid pace at which discoveries have been made over the past decade in the areas of signal transduction and transcriptional regulation has opened the possibility of selectively switching a gene off or on directly by rational targeting of specific transcription factors involved in human cancers. In marked distinction to basal transcription factors, gene-specific transcription factors and, therefore, oncogenic transcription factors have the advantage of being promoter-selective and of modulating the expression of only a limited number of genes. Furthermore, the remarkable diversity in their structure (no two oncogenic transcription factors are exactly alike) offers unique biological surfaces to target. A plethora of studies have underlined the paramount importance of three-dimensionality in

molecular recognition and discrimination of these regulatory proteins (35). As a result of technological advances in X-ray crystallography/NMR spectroscopy and computing, highly refined information for the three-dimensional (3-D) atomic conformation of oncogenic transcription factors can be now utilized to create molecules that are "custom built" to complex with the critical surfaces of these factors and render them inert (or reactive) in a selective (tumor-specific) manner. Integrated approaches that combine modeling with experimental methodologies may be especially powerful in forecasting the affinity of novel molecules designed this way (36).

### Computational Aspects of Small-Molecule Drug Design

Computational drug design allows the simulation of the physicochemical properties of drug molecules and their targets. The drug industry's need to develop products systematically, together with the inefficiency of classical drug discovery, has led to the progressive development of the technique of "rational" drug design, i.e., the use of computers to literally design drugs atom by atom. Since the first developments over 20 years ago, computational drug design has expanded to include new compound discovery (by computer searching of chemical databases), compound optimization (the systematic modification of functional groups to maximize potency and minimize or eliminate side effects such as toxicity), and even *de novo* drug design (the ability to generate entirely new molecules that might fit a target site and act as specific antagonists or inhibitors; see below) (37,38).

Formally, computational chemistry is the quantitative modeling of chemical behavior on a computer by the formalisms of numerical methods. This includes database systems that can search for structures, activities, or properties; tools for analyzing experimental data such as those from diffraction or spectroscopic studies; modeling and visualization systems for exploring and predicting chemical properties and structures; and computer-assisted synthesis and planning systems that propose synthesis or reaction pathway schemes for small-molecule drug development. Molecular modeling itself encompasses the generation and representation of the 3-D structure of molecules and their physicochemical properties. This incorporates structure building and conversion from one or two dimensions into

three dimensions, simulation of chemical properties and behavior, analysis of structures and their associated stereoelectronic properties, and quantitative methods to compare chemical structures for similarities or differences that may be related to their physical properties (37–39). The role of computational drug design is to aid in the discovery and optimization of new candidate drug molecules.

### Structure-Based Drug Design Targeting Transcription Factors

The most significant problem arising in this approach is the ligand (drug)-protein (transcription factor domain/region) "docking" problem. The docking problem is really a search problem involving two steps: the first is generating all potential solutions, and the second is eliminating the improbable or incorrect ones. Small-molecule drug-transcription factor recognition is achieved through the molecular surface structures and the implicit interaction energies with the most commonly associated minor conformational changes. The factors contributing to complementarity include the size and shape of interacting surfaces (geometric properties) as well as special features of a surface structure, such as charge distribution and hydrophobicity (gross chemical properties). A common molecular docking procedure can be divided into two stages. The first is a selection of a population of complexes by geometric docking in which surface structures of two interacting molecules are matched with each other, allowing minor conformational changes implicitly on the basis of complementarity in size and close packing in shape. Searching for the optimal match in a chosen solution space (which has to be delicately balanced to maintain both computational efficiency and completeness in searching) is accomplished by the use of elaborated combinatorial algorithms (40).

After the potential solutions are generated in a chosen solution space in the first stage of the docking procedure, the precision of matching two surface structures (i.e., the "fitness" of candidate ligands [drugs] into the targeted transcription factor site) is evaluated in the second stage from the energetic point of view. This stage utilizes a detailed knowledge of the atomic interactions involved in molecular association, much of which has been obtained by high-resolution X-ray crystallographic and/or NMR analyses of macromolecular complexes (transcription factor

alone and, ideally, of the DNA–transcription factor or the transcription factor–other protein complexes). Since the putative drug is not present in the solved structure, one must be “docked” into the target site by considering the energy of drug binding (41–44). Any calculation of the binding energy of the simulated docking of a drug to a transcription factor target site must take into account the “cavitation” effect (the loss of bound solvent from the drug and targeted transcription factor sites), van der Waals/hydrophobic effects, electrostatic (hydrogen bonding, charge–charge/dipole–dipole interactions) and induced electrostatic effects, conformational effects (induced fit) in the target, and the entropic effect resulting from the restriction of several degrees of freedom (e.g., vibrational, rotational, translational) in the drug molecule. Since there is no method to accurately calculate each of these effects, approximations must be made taking into consideration that multiple binding modes of the drug to its target transcription factor site are also possible. Even in the absence of the 3-D structure of the target transcription factor, drug design that takes into account the 3-D flexibility of candidate ligands can help revolutionize the discovery of new compounds (38).

The great advantage of structure-based design is that rather than trying to find molecules that adopt suitable geometries for a single binding mode, the entire transcription factor site can be explored, allowing a diversity of transcription factor domain/region–drug interactions to be considered. A number of different methodologies and algorithms have been adopted for docking organic molecule databases into known target structures (44–46).

### **De novo Drug Design Based on Target Transcription Factor 3-D Structures**

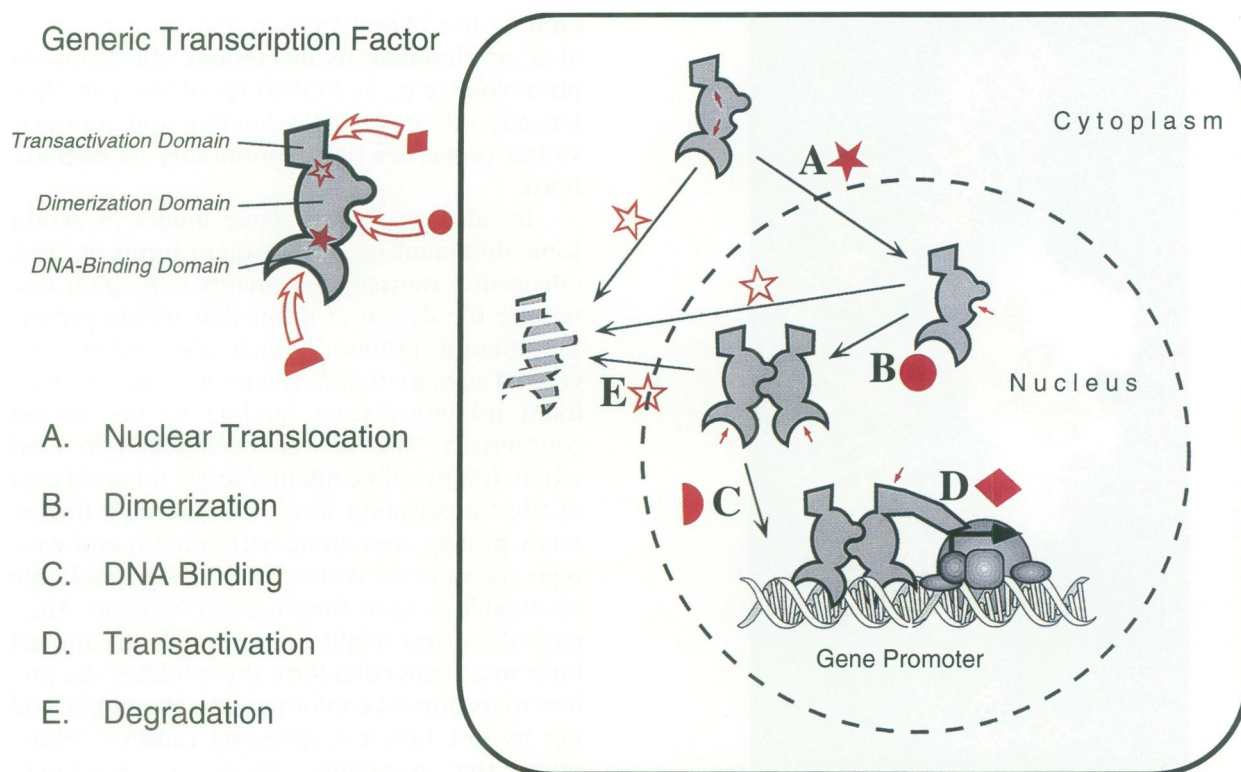
The most ambitious route to designing appropriate transcription factor inhibitors is to create completely new compounds as drugs. The new molecules may be based on existing inhibitors or antagonists, or they may be created from scratch, atom by atom. At least three distinct approaches to de novo design (the design of novel compounds against a target based on structural information about that target) have emerged: directed design, random design, and grid-based design. For each of these, the de novo paradigm can be split into two phases: structure generation (either atom by atom or by linking together existing fragments and templates), and structure

evaluation (whereby the structures are assessed and prioritized using a scoring scheme). Each of these methodologies has its own strengths and weaknesses, depending on whether one wishes to build molecules from linked fragments that have been matched to transcription factor site points, from linked fragments grown from a seed point using a potential energy function, or from linked fragments built using irregular lattices of previously docked molecules. Today, one of the most exciting developments in de novo design is the use of the known 3-D structure of a transcription factor as a “virtual” screen against a combinatorial library, allowing the de novo design methodology to generate diversity focused toward a specific transcription factor site (DNA-binding/oligomerization/transactivation domain or other functionally defined regions). Indeed, there are a number of molecules in clinical trials that have been assisted by de novo design philosophies (38,47).

### **Mode of Action of Putative Transcription Factor Inhibitors**

The modular architecture of gene-specific transcription factors predisposes them to the effects of small-molecule drugs. An oncogenic transcription factor inhibitor may act by binding to the DNA-binding domain, the dimerization domain(s), the transactivation (transrepression) domain(s), or other defined protein/ligand (e.g., hormone)-binding regions mediating a specific biochemical function (see Figs. 2 and 3). The mode of action of the drug in all these cases could be simple steric hindrance of either the requisite DNA–protein or protein–protein/ligand interactions, resulting in a loss of function along with the associative loss of transcriptional regulation (i.e., reduced gene expression when inactivating an activator of transcription, enhanced gene expression when inactivating a transcriptional repressor). In the clinically important situation of the p53-E6 oncoprotein association (see above), steric blockage of the region mediating this interaction would abolish the ability of the papilloma virus product to degrade *in trans* the tumor suppressor. On the other hand, if both the DNA-binding and transactivation (transrepression) domains undergo conformational changes to bind tightly with their cognate DNA or protein partners, the drug might act allosterically, thereby preventing a requisite conformational alteration in a critical domain of the transcription factor (48). Finally, the drug may interfere with the function of





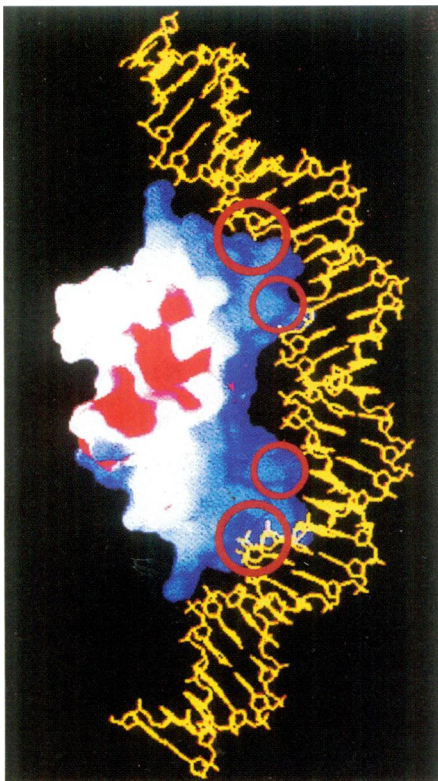
**FIG. 2. Anatomy of gene-specific transcription factors and potential ways in which small-molecule drugs could inhibit their function**

Minimally, an idealized gene-specific transcription factor contains two distinct domains, a DNA-binding domain and a transactivation (transrepression) domain. Transcription factors that have to dimerize in order to bind to their target DNA sequence additionally contain a dimerization domain. Some transcription factors bear additional domains or regions employed for specific protein/ligand (e.g., hormone) binding. After their biosynthesis in the cytoplasm, such transcription factors have to migrate to the nucleus, dimerize, bind to their target gene promoter, and interact with the basal transcriptional apparatus, consisting of basal transcription factors and Pol II (see also Fig. 1). This sequence of events ultimately causes the enhanced (or suppressed) transcription of the target gene. The process can in principle be regulated at any level, e.g., nuclear transport (A), dimerization (B), DNA binding (C), and transactivation (D), but transcription factor degradation (E) can also be subject to control mechanisms that affect the expression level of the target gene (see text for details). A small-molecule drug (indicated by the red-filled rhomboid, red-filled sphere, red-filled hemisphere, red-filled asterisk, and red asterisk) could interfere with any of the above steps by interacting with the appropriate domain or region, thus inhibiting transcription factor function. Interaction with either the DNA-binding or dimerization domain would inhibit DNA binding, whereas interaction with the transcriptional activation domain could inhibit transactivation, leaving the DNA-binding function unperturbed. Red-filled hemisphere, drug targeted to the DNA-binding domain; red-filled sphere, drug targeted to the dimerization domain; red-filled rhomboid, drug targeted to the transactivation (transrepression) domain; red-filled asterisk, drug designed to bind to the region responsible for nuclear localization; red asterisk, drug designed to bind to the region regulating transcription factor degradation (the latter two sites have been arbitrarily chosen on the surface of the transcription factor molecule). Small red arrows point to the transcription factor domain or region interacting with the corresponding drug (i.e., designed on the basis of its 3-D structure).

oncogenic transcription factors by mimicking the positive or negative effect of a regulatory (de)phosphorylation on the interaction potential of any of the aforementioned domains/regions.

Masking of the nuclear localization signal(s) by direct targeting of the relevant region, or by mimicking the action of a sequestering protein or that of a regulatory (de)phosphorylation may hinder the transcription factor from reaching the

nuclear compartment and exercise its effect on gene expression. Similarly, exposing degradation signals by direct targeting of the responsible region or neighboring sites acting as molecular switches, or by designing compounds that neutralize or substitute for the effect of regulatory (de)phosphorylations may substantially decrease the half-life of the transcription factor (see Fig. 2).



**FIG. 3. Targeting transcription factor–DNA binding by “rational” design of small-molecule drugs**

A stereo-diagram depicting a representative—dimeric—transcription factor (in red, white, and blue) interacting symmetrically with a DNA double helix (in yellow). Protein contacts to DNA are mediated through a combination of coulombic, polar, and nonpolar interactions with the deoxyribose-phosphate backbone and with the pendant heterocyclic bases (see text for details). Red circles in the interface of the two macromolecules mark potential target sites within the DNA-binding domain of the transcription factor moiety for structure-based or de novo design of small-molecule drugs interfering with its DNA-binding activity.

The existence of specific gene-fusion products in particular types of human leukemia and some solid tumors (e.g., Ewing’s sarcoma, clear-cell sarcoma, alveolar rhabdomyosarcoma) points the way towards selective targeting of chimeric transcription factors by structure-based or de novo drug design. The fused proteins are characterized by novel amino acid sequences (and thus unique local configuration) at the junctions between the fusion partners, that are not found in normal cells. The tertiary structure of these regions could provide tumor-specific targets for the design of small-molecule drugs that

bind to the hybrid transcription factor and reduce or eliminate its deleterious effects on cell physiology (e.g., by inhibiting—directly or allosterically—its novel DNA-binding and transactivation properties or by promoting its degradation).

In addition to the above modes of action, some dominant negative mutant forms of “anti-oncogenic” transcription factors (e.g., p53) may require the design of drugs that would prevent the mutant protein (which also exhibits increased stability that facilitates its action as dominant inhibitor) from binding to the normal counterpart. This is well exemplified in cases where tumor cells contain a single mutated copy of the transcription factor that forms a heteromeric protein containing both mutant and wild-type subunits, in which the wild-type subunits are unable to exert their normal function. Alternatively, drugs might act to switch the mutant form to an activating form (by refolding the protein to its normal conformation), thereby restoring its lost tumor-suppressing capacity. Moreover, the possibility exists of developing transcription factor–modulating pharmacological agents that would inhibit progression through the cell cycle and/or induce apoptosis, potentially compensating for the loss of tumor suppressor gene products involved in the operations of the cell cycle clock apparatus (p53, pRb) and/or function in the regulation of programmed cell death (p53).

Inasmuch as the precise rate of cellular growth is likely to be controlled by the balance between interacting oncogene and tumor suppressor gene products, with cancer resulting from a change in this balance because of aberrant activation or increased expression of oncogenes or inactivation of anti-oncogenes, the possibility of directing structure-based or de novo drug design towards changing this balance in favor of the arrest of growth as well as manipulating individual mutant factors holds the promise of significant therapeutic advances in the future.

## CONCLUSIONS

Today, computer-aided molecular modeling and structure-based or de novo drug design are rapidly evolving methodologies that have found a niche in every large pharmaceutical company as well as in many smaller biotechnology companies. In fact, several companies have been formed with the express intention of using solely

these strategies for defining and exploiting "transcription" targets at the molecular level. There is no doubt that the use of these approaches in oncogenic transcription factor-targeted drug discovery, design, and optimization is an extraordinarily powerful tool that will surely expand the horizons of anticancer drug development. The ability to modulate the activity of oncogenic transcription factors directly and in a selective manner (and with acceptable toxicity profiles of the putative drugs) will allow the monitoring of tumor cell growth and progression with previously unattainable precision. This transcription factor-based therapeutic approach may enrich the anticancer drug quiver with a totally new spectrum of drugs; this will challenge cancer treatment in a fundamental way and will add significantly to the current clinical armamentarium.

## REFERENCES

- Latchman DS. (1995) *Eukaryotic Transcription Factors*, 2nd ed. Academic Press, London.
- Tjian R. (1996) The biochemistry of transcription in eukaryotes: a paradigm for multisubunit regulatory complexes. *Phil. Trans. R. Soc. Lond. B* **351**: 491–499.
- Harrison SC. (1991) A structural taxonomy of DNA-binding domains. *Nature* **353**: 715–719.
- Papavassiliou AG. (1995) Transcription factors. *N. Engl. J. Med.* **332**: 45–47.
- Papavassiliou AG. (1995) Transcription factors: structure, function, and implication in malignant growth. *Anticancer Res.* **15**: 891–894.
- Travers A. (1993) DNA-protein interactions: sequence specific recognition. In: *DNA-Protein Interactions*. Chapman & Hall, London, pp. 52–86.
- Pabo CO, Sauer RT. (1992) Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**: 1053–1095.
- Ouzounis CA, Papavassiliou AG. (1997) DNA-binding motifs of eukaryotic transcription factors. In: Papavassiliou AG (ed). *Transcription Factors in Eukaryotes. Molecular Biology Intelligence Unit*. Landes Bioscience, Austin & Springer-Verlag, Heidelberg, pp. 1–21.
- Tjian R, Maniatis T. (1994) Transcriptional activation: A complex puzzle with few easy pieces. *Cell* **77**: 5–8.
- Edwards DR. (1994) Cell signalling and the control of gene transcription. *Trends Pharmacol. Sci.* **15**: 239–244.
- Hunter T, Karin M. (1992) The regulation of transcription by phosphorylation. *Cell* **70**: 375–387.
- Karin M. (1994) Signal transduction from the cell surface to the nucleus through the phosphorylation of transcription factors. *Curr. Opin. Cell Biol.* **6**: 415–424.
- Calkhoven CF, AB G. (1996) Multiple steps in the regulation of transcription-factor level and activity. *Biochem. J.* **317**: 329–342.
- Bishop JM, Hanafusa H. (1996) Proto-oncogenes in normal and neoplastic cells. In: Bishop JM, Weinberg RA (eds). *Molecular Oncology*. Scientific American, New York, pp. 61–83.
- Macdonald F, Ford CHJ. (1997) *Molecular Biology of Cancer*. BIOS Scientific, Oxford, pp. 13–72.
- Forrest D, Curran T. (1992) Crossed signals: oncogenic transcription factors. *Curr. Opin. Genet. Dev.* **2**: 19–27.
- Cooper GM. (1995) Transcription factors. In: *Oncogenes*, 2nd ed. Jones & Bartlett, Boston, pp. 255–278.
- Wisdom RM. (1997) Oncogenic transcription factors. In: Papavassiliou AG (ed). *Transcription Factors in Eukaryotes. Molecular Biology Intelligence Unit*. Landes Bioscience, Austin & Springer-Verlag, Heidelberg, pp. 219–234.
- Pawson T. (1996) The biochemical mechanisms of oncogene action. In: Bishop JM, Weinberg RA (eds). *Molecular Oncology*. Scientific American, New York, pp. 85–109.
- Rabbitts TH. (1994) Chromosomal translocations in human cancer. *Nature* **372**: 143–149.
- Latchman DS. (1996) Transcription-factor mutations and disease. *N. Engl. J. Med.* **334**: 28–33.
- Sawyers CL, Denny CT. (1994) Chronic myelomonocytic leukemia: Tel-a-kinase what Ets all about. *Cell* **77**: 171–173.
- Dyck JA, Maul GG, Miller WH Jr, Chen JD, Kakizuka A, Evans RM. (1994) A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. *Cell* **76**: 333–343.
- Fredericks WJ, Galili N, Mukhopadhyay S, et al. (1995) The PAX3-FKHR fusion protein created by the t(2;13) translocation in alve-

- olar rhabdomyosarcomas is a more potent transcriptional activator than PAX3. *Mol. Cell. Biol.* **15**: 1522–1535.
25. Marcu KB, Bossone SA, Petel AJ. (1992) *Myc* function and regulation. *Annu. Rev. Biochem.* **61**: 809–860.
  26. Cooper GM. (1995) Oncogenes and chromosome translocation. In: *Oncogenes*, 2nd ed. Jones & Bartlett, Boston, pp. 99–112.
  27. Knudson AG. (1993) Antioncogenes and human cancer. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 10914–10921.
  28. Davies RC, Hastie ND. (1997) Tumor suppression by 'transcription' factors. In: Papavassiliou AG (ed). *Transcription Factors in Eukaryotes. Molecular Biology Intelligence Unit*. Landes Bioscience, Austin & Springer-Verlag, Heidelberg, pp. 297–316.
  29. Marx J. (1993) How p53 suppresses cell growth. *Science* **262**: 1644–1645.
  30. Hastie ND. (1993) Wilm's tumour gene and function. *Curr. Opin. Genet. Dev.* **3**: 408–413.
  31. Sarma MH. (1997) p73 gene: A new relative of p53. *Cancer Watch* **6**: 139–140.
  32. La Thangue NB. (1994) DP and E2F proteins: components of a heterodimeric transcription factor implicated in cell cycle control. *Curr. Opin. Cell Biol.* **6**: 443–450.
  33. Weinberg RA. (1995) The retinoblastoma protein and cell cycle control. *Cell* **81**: 323–330.
  34. Cooper GM. (1995) Tumor suppressor genes in human neoplasms. In: *Oncogenes*, 2nd ed. Jones & Bartlett, Boston, pp. 145–161.
  35. Marshall GR, Cramer RD III. (1988) Three-dimensional structure–activity relationships. *Trends Pharmacol. Sci.* **9**: 285–289.
  36. Peisach E, Casebier D, Gallion SL, et al. (1995) Interaction of a peptidomimetic aminimide inhibitor with elastase. *Science* **269**: 66–69.
  37. Neidle S. (1994) Discovery of new anticancer drugs by computer-aided drug design. *Ann. Oncol.* **5**(Suppl. 4): S51–S54.
  38. Maulik S, Patel SD. (1997) Protein engineering and computer-assisted drug design. In: *Molecular Biotechnology—Therapeutic Applications and Strategies*. Wiley-Liss, New York, pp. 109–153.
  39. Brint AT, Willett PJ. (1987) Upperbound procedures for the identification of similar three-dimensional chemical structures. *J. Comput. Aided Mol. Des.* **2**: 311–320.
  40. Wang H. (1991) Grid-search molecular accessible surface algorithm for solving the protein-docking problem. *J. Comput. Chem.* **12**: 746–750.
  41. Goodford PJ. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**: 849–857.
  42. Smellie AS, Crippen GM, Richards WG. (1991) Fast drug-receptor mapping by site-directed distances: A novel method of predicting new pharmacological leads. *J. Chem. Inf. Comput. Sci.* **31**: 386–392.
  43. Kuntz ID. (1992) Structure-based strategies for drug design and discovery. *Science* **257**: 1078–1082.
  44. Sudarsanam S, Virca GD, March CJ, Srinivasan S. (1992) An approach to computer-aided inhibitor design: Application to cathepsin L. *J. Comput. Aided Mol. Des.* **6**: 223–233.
  45. Desjarlais RL, Sheridan RP, Seibel GL, Dixon JS, Kuntz ID, Venkataraghavan R. (1988) Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **31**: 722–729.
  46. Lawrence MC, Davis PC. (1992) CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins* **12**: 31–41.
  47. Rotstein SH, Murcko MA. (1993) GenStar: A method for de novo drug design. *J. Comput. Aided Mol. Des.* **7**: 23–43.
  48. Peterson MG, Baichwal VR. (1993) Transcription factor based therapeutics: drugs of the future? *Trends Biotechnol.* **11**: 11–18.