Gene Expression, Vol. 4, pp. 337-344, 1995
Printed in the USA.

1052-2166/95 $5.00 + .00

Gene Expression as a Target for New Drug Discovery

ADRIANA HEGUY,[1] ALEXANDER A. STEWART, JOHN D. HALEY, DAVID E. SMITH, AND J. GORDON FOULKES

Oncogene Science, Inc., 106 Charles Lindbergh Blvd., Uniondale, NY 11553

Over the last 7 years we have carried out a major research effort focused on gene transcription as a novel approach to drug discovery. The goal is to identify small molecular weight compounds that modulate the expression of a target gene in a specific manner, thereby either increasing or decreasing the concentration of the corresponding protein product. Transcriptional modulation not only provides a potential means to replace recombinant proteins as drugs, but also provides a novel approach to manipulate key gene targets in many therapeutic areas. This article describes some of the features and advantages of transcription-based pharmaceuticals and illustrates how this approach can be applied to drug discovery with a program we are pursuing to identify new treatments for sickle cell disease and β-thalassemia.

Transcriptional modulation High throughput screening Drug discovery Sickle cell anemia

TRADITIONAL DRUG DISCOVERY APPROACHES

Until recently, a major focus of the pharmaceutical industry was centered around improving either the efficacy or safety of existing drugs, agents for which the mechanism of action was often unknown. Chemistry efforts to support this approach during this century have been prolific, and more than 3 million unique chemical structures have now been cataloged.

Within the last decade, however, molecular biologists have begun to unravel the detailed molecular aspects of normal cellular physiology and the metabolic abnormalities associated with human diseases. To date, the application of molecular biology to drug discovery has taken three basic directions: i) the discovery of protein factors that act directly as pharmaceutical agents; ii) the combination of x-ray crystallography and computer modeling to design organic compounds based upon a detailed understanding of protein structure; or iii) the use of recombinant proteins for limited in vitro screening of natural products or synthetic organic compounds. Each of these approaches, however, suffers from significant limitations.

The development of protein-based pharmaceuticals is perhaps the best known application of molecular biology to the drug industry. Major successes have included the use of growth hormone for the treatment of pituitary dwarfism, tissue plasminogen activator to treat myocardial infarction, and erythropoietin and G-CSF for the treatment of anemia and neutropenia, respectively. All of these agents, however, possess a number of major limitations characteristic of protein-based therapeutants. These include the route of administration, typically intravenous, the high cost of manufacture, the technical difficulties relating to formulation and stability, and the relative complexity of patenting recombinant biologicals. Such problems also apply to therapeutic methods based upon the use of monoclonal antibodies.

The second approach to drug discovery is contingent upon a detailed knowledge of specific protein structure. Interactive computing using modern supercomputers theoretically allows the design of small molecular weight compounds that are predicted to bind to specific sites on the target

Accepted December 6, 1994.
[1]Address correspondence to Adriana Heguy.

protein. This approach requires that the three-dimensional structure of the target protein be known at very high resolution. The generation of high-resolution protein structure is a very difficult and lengthy process, and the structure of the protein as determined by crystallography may not reflect the native conformation of the protein within the cell. For example, many proteins are now known to exist in macromolecular complexes and not as isolated entities. As more protein structures are solved, and as modeling programs become more powerful, rational drug design is likely to make increasingly important contributions to drug discovery. To date, however, there are no drugs on the market that have been developed de novo using this technology.

The third traditional method has involved the screening of several thousand fermentation broths, natural product extracts or synthetic organic compounds, for their ability to influence the function of a purified or partially purified protein (e.g., a hormone receptor or an enzyme). In many instances, however, it can prove difficult to isolate a complex protein with retention of its activity. In addition, the critical biochemical activity of many disease targets may be unknown, thereby preventing the design of an appropriate in vitro screen.

GENE EXPRESSION AS A TARGET FOR DRUG INTERVENTION

In marked contrast to traditional drug discovery, transcription-based approaches have the potential to overcome many of these limitations and lead to the development of novel pharmaceuticals targeting most of the major disease markets.

Over the last few years, it has become clear that many normal physiological pathways, as well as the body's response to a wide variety of disease states, involve specific changes in gene transcription. For example, the amount of phosphoenolpyruvate carboxykinase (PEPCK) produced by the liver is tightly controlled at the transcriptional level by physiological signals. Transcription of the PEPCK gene is activated by glucagon and glucocorticoids, and inhibited by insulin (Granner and O'Brien, 1992). The level of the hemopoietic growth factor erythropoietin is controlled by the level of oxygen in the blood: the erythropoietin gene is specifically activated in response to anoxia (Semenza, 1991). T cells are activated by ligand interactions at the cell surface with the T-cell receptor and CD28, events that result in transcriptional activation of both IL-2 and its receptor (Schwartz, 1992). Similarly, most of the antiviral effects of the interferons are mediated by changes in transcription of key genes (Sen and Lengyel, 1992). The major goal of transcription-based drug discovery is the identification of drugs that modulate such normal physiological responses to produce the desired pharmacological end point.

In addition to the physiological evidence that demonstrates specific regulation of gene transcription, there are already a number of clinically approved drugs that have now been shown to act by modulating the expression of specific genes. Steroid and thyroid hormones, as well as vitamins A and D, bind to receptors that are transcription factors (McDonnell et al., 1993). The steroid receptor binding drugs Premarin (conjugated estrogens) and Nolvadex (tamoxifen) modulate the expression of estrogen-responsive genes to control osteoporosis and breast cancer, respectively (McDonnell, 1993; Mitlak and Nussbaum, 1993). Similarly, Eulexin (flutamide) demonstrates dramatic palliative efficacy in the treatment of prostate cancer by blocking the action of the testosterone receptor and thus modulating the expression of androgen-responsive genes (McDonnell et al., 1993). Steroid receptors are also the targets of drugs used as anti-inflammatory agents and contraceptives.

A number of drugs that act independently of steroid hormone receptors have also been demonstrated to mediate their effects by modulating gene expression. One example of note is aspirin, which has been postdated to mediate its anti-inflammatory effects by inhibiting transcriptional activation of prostaglandin synthesis by IL-1 (Wu, 1991). These effects have been demonstrated to take place by the specific action of aspirin on one of the two promoters of the prostaglandin H synthase-1 (Xu, 1993). The potent nonsteroidal, anti-inflammatory compound naproxen may also act by inhibition of prostaglandin H synthase-1 transcription (Zyglewska, 1992). The concentration of aspirin required to mediate these transcriptional responses is some three orders of magnitude below that required to cause acetylation of prostaglandin synthase, previously considered to be its classical mechanism of action (Sanduja et al., 1990). Recently, it has also been shown that aspirin directly influences the activity of the transcription factor NF-κB, by preventing the degradation of its inhibitor, IκB (Kopp and Ghosh, 1994).

The immunosuppressive drugs cyclosporin A and FK506 inhibit the transcription of several genes involved in T-cell activation, particularly IL-2 (Schreiber and Crabtree, 1992). Cyclosporin A and FK506 block the activation of the nuclear

transcription factor of activated T cells (NF-AT), which is essential for expression of the IL-2 gene upon T-cell activation. These drugs form complexes with cyclophilin and FK506 binding proteins, respectively, which in turn bind to and potently inhibit the activity of the calcium- and calmodulin-dependent phosphatase, calcineurin. This phosphatase is responsible for the dephosphorylation and activation of the DNA binding component of NF-AT (McCaffrey et al., 1993).

Other examples of drugs that act by influencing the transcriptional state or levels of genes involved in human disease are the cholesterol-lowering drug lovastatin and the triglyceride-lowering fibrates. It is now clear that lovastatin's ability to regulate cholesterol levels is primarily due to its indirect action in upregulating the hepatic LDL receptor (Drug & Market Development, 1993), and not inhibition of HMGCoA reductase per se (as evidenced by the fact that lovastatin is ineffective in regulating cholesterol levels in patients with defective LDL receptors). Fibrates may act by specific transcriptional activation of hepatic lipoprotein lipase (Staels and Auwerx, 1992), an enzyme not normally expressed in adult liver. This induction is tissue specific because fibrates have no effect on lipoprotein lipase mRNA levels in the heart or adipose tissue. The exact mechanism of action of fibrates has not been elucidated but may involve their activation of the transcription factor PPAR (peroxisomal proliferator-activated receptor), a member of the steroid receptor gene superfamily (Auwerx, 1992).

ADVANTAGES OF TARGETING GENE TRANSCRIPTION

One of the major advantages of the use of pharmaceuticals that alter the transcriptional status of a given gene will be the replacement of recombinant proteins as therapeutic agents (e.g., replacing the erythropoietin or growth hormone with small molecular weight pharmaceuticals that act to increase the expression of the endogenous gene). This approach could potentially overcome all of the major limitations associated with the use of biologicals (e.g., patents, the high costs of production, formulation, route of administration, etc). Both recombinant biologicals and monoclonal antibodies require their target to be either present in the circulation or on the surface of cells. Transcriptional modulation can act to affect either intracellular or extracellular targets. Most traditional drugs act to inhibit the activity of a given protein, either as receptor antagonists or enzyme inhibitors. In contrast, the transcriptional approach allows the absolute concentration of a target protein to be either increased or decreased. This also represents a significant advantage compared to current antisense-based approaches. Conventional antisense molecules are inactive orally, very expensive, and act only to decrease (not increase) protein levels.

Another advantage of transcription-based pharmaceuticals is when the target protein exists as a member of a multigene family. Modulating the activity of an entire family of structurally related proteins could be deleterious to the organism, but modulation of a specific member of a group of highly homologous proteins can be difficult to achieve pharmacologically. One example would be the specific transcriptional modulation of a CNS receptor where although the family of polypeptides may exhibit a high degree of homology, the regulatory regions (and the corresponding transcription factors of the gene families that bind to these regions) are typically distinct to allow cell type-specific expression. The individual transcription factors that regulate the expression of any one gene can also vary dramatically between different tissues. Thus, in addition to the ability to selectively regulate the expression of specific members of multigene families, transcription drugs have the potential to modulate the expression of target genes in a tissue-specific manner. Again, this is not usually possible with conventional drugs or antisense approaches, because the mRNA sequence for an individual gene is normally identical in every cell type. One example of a transcriptionally active drug that acts in a tissue-specific manner is provided by the anticancer drug tamoxifen, which acts as an antagonist of the estrogen receptor (a transcription factor) in breast, and acts as an agonist of the receptor in bone (Henderson et al., 1993).

Finally, it is now well documented that coordinated changes in gene expression regulate a large number of metabolic processes in the body. Thus, it may be possible to use a single transcription drug to regulate the expression of small groups of genes simultaneously, rather than using a conventional pharmaceutical to simply target one specific enzyme or receptor. Given that many diseases, including atherosclerosis, cancer, and diabetes, are now known to be multifactorial in origin, a transcription approach may ultimately prove to be essential to develop truly effective drugs in these areas. For example, the ubiquitous transcription factor NF-κB, is involved in the activation of

many different genes, most of which are involved either directly or indirectly in immune responses and inflammatory processes (e.g., cytokine and cytokine receptor genes, MHC class I and serum amyloid protein genes) (Baeuerle and Baltimore, 1991). Thus, an inhibitor of this transcription factor may be useful as an immunosuppressant, against septic shock, and for tissue damage induced by inflammation.

APPROACHES TO TRANSCRIPTION-BASED DRUG DISCOVERY: IN VITRO VERSUS CELL BASED

Over the past decade in vitro transcription systems have been developed using purified or partially purified components of the basic transcriptional machinery coupled to specific transcription factors. Numerous binding assays to detect DNA–protein interactions have also been developed. Although these systems are extremely useful for dissecting transcriptional processes and identifying specific components of a transcription complex, in most cases it has not yet proven possible to faithfully recapitulate the physiological regulation of a gene outside the context of an intact cell environment.

Our approach to transcription-based drug discovery has been to utilize genetically engineered live cells lines containing reporters for the expression of single specific genes. Thus, once a particular gene has been chosen as a target, the first step of the process is to clone the regions of DNA that regulate expression of that gene. These sequences are then fused to a highly sensitive reporter gene, the firefly luciferase gene, which generates a readily measurable light signal in response to changes in transcription of the target gene. This genetically engineered DNA construct is then introduced into a relevant cell type in which the regulation of gene expression is closest to the in vivo situation, and cell lines are isolated. Such cell lines generate a light signal that reflects a specific change in gene expression when an appropriate compound is added to the tissue culture media.

Although both in vitro and cell-based assays can be automated to allow the rapid screening of a large number of samples, it is clear that cell-based approaches provide a number of major advantages. Cell-based assays provide screens that are considerably more physiological than those using purified enzymes or receptors in vitro, yet avoid the biological complexity and ethical issues associated with animal-based screens. The cell-based approach also provides multiple target sites for pharmacological intervention, because each step of the complete signal transduction pathway from receptor to transcription factor is simultaneously assayed, thereby significantly increasing the possibility of finding lead compounds. Another major advantage of cell-based screens relates to the considerable amount of time that can be lost in attempting to develop in vitro leads that fail to exhibit whole cell activity. Purification and reconstitution of components for in vitro assays can also be time consuming, expensive, and, in many cases, not possible due to lack of detailed knowledge about the biochemical activities involved. By employing live cells in the primary screen, compounds that are either cytotoxic or fail to interact with the cell in a functional manner (due to lack of permeability or degradation) are readily excluded. In addition, pro-drugs (i.e., compounds that are metabolized into the active drug product) can also be identified using cell-based screens.

One final advantage of the cell-based assays is the possibility of screening for compounds that act in a tissue-specific manner to regulate transcription of a target gene. It is becoming increasingly apparent that the mechanisms that regulate any one gene at the transcriptional level vary between different tissues (i.e., the transcription factors that bind to the regulatory region of a gene can vary widely from one cell type to the next). Thus, by inclusion of the same tester construct in different cell types in the primary screen (or by subsequent analysis in secondary screens), this approach has the potential to identify drugs that modulate the expression of a target gene in a tissue-specific manner. The potential of this approach for reducing unwanted side effects is considerable, because many of the problems associated with otherwise efficacious drugs are due to effects on nontarget tissues (e.g., the agonistic action of tamoxifen on the endometrium has been reported to significantly increase the incidence of endometrial cancer) (van Leeuwen, 1994; Marshall, 1994).

ULTRA HIGH THROUGHPUT SCREENING AS AN APPROACH TO TRANSCRIPTION-BASED DRUG DISCOVERY PROGRAM

Traditional approaches to drug screening have relied heavily on structure–activity studies and incremental improvements of drug leads by evaluation of analogues produced by medicinal chemistry. This is a time-consuming approach that may

well overlook the potential of identifying leads with completely novel structures. Recent advances in laboratory automation at Oncogene Science now allow the ultra high throughput screening of extremely large libraries, making it possible to consider lead structures representing as much molecular diversity as possible.

To facilitate the screening of large numbers of test samples, a proprietary robotic system has been developed, which enables complete automation of every step in the drug screening process. On-line cell incubation facilities interface with liquid handling systems for the dilution and addition of test samples, and an array of other units are also robotically manipulated in an assay loop that transfers microplates between the plate washer, the reagent addition work station, and a 96-well luminometer. Two robotic arm assemblies are employed in each system. Data are captured automatically into a processing network that performs quality control evaluations on each individual microplate assay, as well as rapid data reduction and analysis.

By modifying the luciferase reaction we have been able to stabilize the light signal and produce assays that are not only highly quantitative but of sufficient sensitivity to allow the detection of fewer than 25 molecules per cell. This provides both a number of major technical advantages, such as the inclusion of weakly active promoters, while enabling the screen to detect initial lead compounds with apparent limited activity (e.g., potent compounds present at low concentrations in fermentation broths). This technology is currently employed in transcriptional reporter assays, cell-based immunoassays, protein kinase assays, and protein–protein interaction screens. Each of these robotic systems can assay up to 10,000 compounds per week. Over the next year, the screening facility at Oncogene will screen over 3 million compounds, against a total of 30 different drug targets.

By using multiple targets in the primary screen, efficacy, cytotoxicity, and initial specificity are evaluated rapidly. For example, the presence of a chemical that specifically activates transcription of the human growth hormone gene is detected by an increase in the reporter signal only in a cell that contains a growth hormone fusion construct. Automated on-line data reduction and statistical analysis allow rapid quantitative determination to identify lead compounds. Compounds of interest identified by the primary screen are then further evaluated in secondary screens (e.g., by quantitative PCR, protein-based assays, etc.) for their ability to regulate the endogenous gene in a similar fashion. A variety of appropriate tertiary assays and animal models can then be employed for the final stages of lead development.

SICKLE CELL ANEMIA AS A TARGET FOR TRANSCRIPTION-BASED DRUG DISCOVERY

As one example of a transcription-based drug discovery program, this section reviews our program on the activation of γ-globin transcription as an approach to treat sickle cell anemia. Sickle cell disease is an often fatal anemia afflicting 1 in 400 African-Americans. This genetic disease results from a glutamic acid to valine change at codon 6 of the adult β-globin gene (Hb to HbS). The altered globin protein aggregates to produce an irreversible deformation of red blood cell morphology. The resulting "sickled" cells then become trapped in the spleen and lyse to produce symptoms of hemolytic anemia. Splenic atrophy is a long-term consequence. Blood vessel occlusion is also a major clinical problem, resulting in increased ischemic infarction. Approximately 10% of African-Americans are asymptomatic heterozygotes, whereas homozygosity is usually associated with severe morbidity. Therapeutic agents developed using the approach of γ-globin induction should also be useful for the treatment of β-thalassemias.

REGULATION OF THE HUMAN β-LIKE GLOBIN GENES

The human β-like globin genes (ϵ, γ, δ, β) exist in a discrete locus on chromosome 11 (Fig. 1). These duplicated globin genes were derived from a common ancestral globin. During embryonic and fetal development, globin genes are expressed in a temporal and tissue-specific fashion. This involves a switch in globin expression from embryonic ($\zeta_2\epsilon_2$, $\alpha_2\epsilon_2$, and $\zeta_2\gamma_2$) to fetal ($\alpha_2\gamma_2$) and then finally to the adult forms ($\alpha_2\beta_2$ and $\alpha_2\delta_2$). Embryonic hemoglobin is synthesized in the yolk sac early in development. During the fifth week of gestation a switch in β-chain subtype occurs and fetal hemoglobin (HbF; $\alpha_2\gamma_2$) synthesis is initiated in the

FIG. 1. Schematic representation of the 100-kb human β-globin locus.

well overlook the potential of identifying leads with completely novel structures. Recent advances in laboratory automation at Oncogene Science now allow the ultra high throughput screening of extremely large libraries, making it possible to consider lead structures representing as much molecular diversity as possible.

To facilitate the screening of large numbers of test samples, a proprietary robotic system has been developed, which enables complete automation of every step in the drug screening process. On-line cell incubation facilities interface with liquid handling systems for the dilution and addition of test samples, and an array of other units are also robotically manipulated in an assay loop that transfers microplates between the plate washer, the reagent addition work station, and a 96-well luminometer. Two robotic arm assemblies are employed in each system. Data are captured automatically into a processing network that performs quality control evaluations on each individual microplate assay, as well as rapid data reduction and analysis.

By modifying the luciferase reaction we have been able to stabilize the light signal and produce assays that are not only highly quantitative but of sufficient sensitivity to allow the detection of fewer than 25 molecules per cell. This provides both a number of major technical advantages, such as the inclusion of weakly active promoters, while enabling the screen to detect initial lead compounds with apparent limited activity (e.g., potent compounds present at low concentrations in fermentation broths). This technology is currently employed in transcriptional reporter assays, cell-based immunoassays, protein kinase assays, and protein–protein interaction screens. Each of these robotic systems can assay up to 10,000 compounds per week. Over the next year, the screening facility at Oncogene will screen over 3 million compounds, against a total of 30 different drug targets.

By using multiple targets in the primary screen, efficacy, cytotoxicity, and initial specificity are evaluated rapidly. For example, the presence of a chemical that specifically activates transcription of the human growth hormone gene is detected by an increase in the reporter signal only in a cell that contains a growth hormone fusion construct. Automated on-line data reduction and statistical analysis allow rapid quantitative determination to identify lead compounds. Compounds of interest identified by the primary screen are then further evaluated in secondary screens (e.g., by quantitative PCR, protein-based assays, etc.) for their ability to regulate the endogenous gene in a similar fashion. A variety of appropriate tertiary assays and animal models can then be employed for the final stages of lead development.

SICKLE CELL ANEMIA AS A TARGET FOR TRANSCRIPTION-BASED DRUG DISCOVERY

As one example of a transcription-based drug discovery program, this section reviews our program on the activation of γ-globin transcription as an approach to treat sickle cell anemia. Sickle cell disease is an often fatal anemia afflicting 1 in 400 African-Americans. This genetic disease results from a glutamic acid to valine change at codon 6 of the adult β-globin gene (Hb to HbS). The altered globin protein aggregates to produce an irreversible deformation of red blood cell morphology. The resulting "sickled" cells then become trapped in the spleen and lyse to produce symptoms of hemolytic anemia. Splenic atrophy is a long-term consequence. Blood vessel occlusion is also a major clinical problem, resulting in increased ischemic infarction. Approximately 10% of African-Americans are asymptomatic heterozygotes, whereas homozygosity is usually associated with severe morbidity. Therapeutic agents developed using the approach of γ-globin induction should also be useful for the treatment of β-thalassemias.

REGULATION OF THE HUMAN β-LIKE GLOBIN GENES

The human β-like globin genes (ϵ, γ, δ, β) exist in a discrete locus on chromosome 11 (Fig. 1). These duplicated globin genes were derived from a common ancestral globin. During embryonic and fetal development, globin genes are expressed in a temporal and tissue-specific fashion. This involves a switch in globin expression from embryonic ($\zeta_2\epsilon_2$, $\alpha_2\epsilon_2$, and $\zeta_2\gamma_2$) to fetal ($\alpha_2\gamma_2$) and then finally to the adult forms ($\alpha_2\beta_2$ and $\alpha_2\delta_2$). Embryonic hemoglobin is synthesized in the yolk sac early in development. During the fifth week of gestation a switch in β-chain subtype occurs and fetal hemoglobin (HbF; $\alpha_2\gamma_2$) synthesis is initiated in the

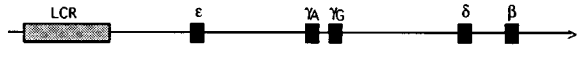

FIG. 1. Schematic representation of the 100-kb human β-globin locus.

liver. HbF remains the predominant form of hemoglobin until birth, when adult hemoglobin (HbA) concentration sharply increases (Stamatoyannopoulos and Niehuis, 1992).

The developmental regulation of the human β-like globin genes occurs at the transcriptional level. Sequences within or immediately flanking the genes themselves confer both tissue and temporal specificity (Chada et al., 1985, 1986; Kollias et al., 1986; Rutherford and Niehuis, 1987). In addition, the region that flanks the β-globin gene cluster, the locus control region or LCR, has been shown to be critical for high level, tissue-specific globin gene expression, and appears to play a role in hemoglobin switching. The LCR has been defined by four major DNase I hypersensitive sites (HS I–IV) distributed over 20 kb 5′ of the globin locus (Tuan et al., 1985; Grosveld et al., 1987; Talbot et al., 1989). Individual hypersensitive sites and combinations of sites have been found to retain much of the function of the full LCR (Talbot et al., 1989). Only the HS II is active as a single site (Sorrentino et al., 1990). The erythroid-specific activity of the HS2 enhancer is dependent on the erythroid-specific protein NF-E2 (Ney et al., 1990). The γ-genes (γ^A and γ^B) appear to be silenced independently of other globin genes, presumably by stage-specific factors binding to sequences flanking the genes (Dillon and Grosveld, 1991). Competition of linked γ and β genes may be mediated in part by a sequence element termed the stage selector element (SSE) identified in the human γ-globin promoter located at position −35 to −53 (Jane et al., 1992). The SSE fosters preferential interaction with the 5′ HS2 enhancer and promotes effective competition with the β promoter. Work by Hanscombe et al. (1991) has shown that the relative position of the β gene with respect to LCR plays an important role in the competition between γ and β genes.

Experimental evidence strongly suggest that the fetal γ gene is actively suppressed in the majority of adult erythroid cells. Within the γ-globin gene promoter there are two CAAT boxes, located upstream of the TATA box. Mutation of the distal CAAT box results in γ-globin gene expression in adult erythroid cells at approximately 60% of the fetal level. Promoter mutations have been defined that: i) cause binding of new transacting factors to the promoter DNA; ii) heighten affinity for existing factors; or iii) mutate the distal CAAT box. The distal CAAT box may be, indirectly, a binding site for the transcription factor GATA-1. It has been postulated that protein–protein contact of GATA-1 and a CAAT box binding protein is responsible for suppression of γ-globin gene expression in adult erythroid cells.

DISCOVERY OF INDUCERS OF γ-GLOBIN GENE EXPRESSION BY ULTRA HIGH THROUGHPUT SCREENING

Current "activators" of fetal globin gene expression such as hydroxyurea, aza-cytidine, and butyrate are relatively nonspecific and toxic, whereas erythropoietin elicits only a transient rise in fetal globin expression. Although the mechanism of action by which these agents stimulate globin expression is not fully understood, and none of these agents are entirely suitable for chronic treatment of sickle cell disease or β-thalassemia, these agents provide proof of principle that transcriptional modulation of γ-globin by pharmacological intervention is feasible. Gene therapy approaches to sickle cell disease treatment also have been pursued, but await the development of efficient transfer vector systems.

We have undertaken high throughput drug screening using a genetically engineered erythroid cell line, K562/HS2γ, containing an integrated reporter plasmid in which the luciferase gene is under the control of the regulatory regions of the γ-globin gene, including HS II, to identify compounds that selectively activate fetal globin gene expression. The DNA constructs used in this screen are depicted in Fig. 2.

Promoter reporter plasmids were constructed as follows: the 1.5-kb KpnI-BglII HS2 element of the LCR was positioned 5′ of a Hinf1-AluI γ^A promoter fragment (− 260 to + 36; Ney et al., 1990) and cloned upstream of the luciferase reporter gene from *P. pyralis*, generating the plasmid HS2γ-luci. The selectable marker TK-neo,

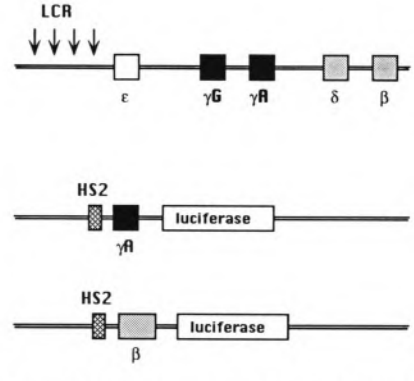

FIG. 2. Schematic representation of the δ- and β-globin promoter/luciferase fusion plasmids.

liver. HbF remains the predominant form of hemoglobin until birth, when adult hemoglobin (HbA) concentration sharply increases (Stamatoyannopoulos and Niehuis, 1992).

The developmental regulation of the human β-like globin genes occurs at the transcriptional level. Sequences within or immediately flanking the genes themselves confer both tissue and temporal specificity (Chada et al., 1985, 1986; Kollias et al., 1986; Rutherford and Niehuis, 1987). In addition, the region that flanks the β-globin gene cluster, the locus control region or LCR, has been shown to be critical for high level, tissue-specific globin gene expression, and appears to play a role in hemoglobin switching. The LCR has been defined by four major DNase I hypersensitive sites (HS I–IV) distributed over 20 kb 5′ of the globin locus (Tuan et al., 1985; Grosveld et al., 1987; Talbot et al., 1989). Individual hypersensitive sites and combinations of sites have been found to retain much of the function of the full LCR (Talbot et al., 1989). Only the HS II is active as a single site (Sorrentino et al., 1990). The erythroid-specific activity of the HS2 enhancer is dependent on the erythroid-specific protein NF-E2 (Ney et al., 1990). The γ-genes (γ^A and γ^B) appear to be silenced independently of other globin genes, presumably by stage-specific factors binding to sequences flanking the genes (Dillon and Grosveld, 1991). Competition of linked γ and β genes may be mediated in part by a sequence element termed the stage selector element (SSE) identified in the human γ-globin promoter located at position −35 to −53 (Jane et al., 1992). The SSE fosters preferential interaction with the 5′ HS2 enhancer and promotes effective competition with the β promoter. Work by Hanscombe et al. (1991) has shown that the relative position of the β gene with respect to LCR plays an important role in the competition between γ and β genes.

Experimental evidence strongly suggest that the fetal γ gene is actively suppressed in the majority of adult erythroid cells. Within the γ-globin gene promoter there are two CAAT boxes, located upstream of the TATA box. Mutation of the distal CAAT box results in γ-globin gene expression in adult erythroid cells at approximately 60% of the fetal level. Promoter mutations have been defined that: i) cause binding of new transacting factors to the promoter DNA; ii) heighten affinity for existing factors; or iii) mutate the distal CAAT box. The distal CAAT box may be, indirectly, a binding site for the transcription factor GATA-1. It has been postulated that protein–protein contact of GATA-1 and a CAAT box binding protein is responsible for suppression of γ-globin gene expression in adult erythroid cells.

DISCOVERY OF INDUCERS OF γ-GLOBIN GENE EXPRESSION BY ULTRA HIGH THROUGHPUT SCREENING

Current "activators" of fetal globin gene expression such as hydroxyurea, aza-cytidine, and butyrate are relatively nonspecific and toxic, whereas erythropoietin elicits only a transient rise in fetal globin expression. Although the mechanism of action by which these agents stimulate globin expression is not fully understood, and none of these agents are entirely suitable for chronic treatment of sickle cell disease or β-thalassemia, these agents provide proof of principle that transcriptional modulation of γ-globin by pharmacological intervention is feasible. Gene therapy approaches to sickle cell disease treatment also have been pursued, but await the development of efficient transfer vector systems.

We have undertaken high throughput drug screening using a genetically engineered erythroid cell line, K562/HS2γ, containing an integrated reporter plasmid in which the luciferase gene is under the control of the regulatory regions of the γ-globin gene, including HS II, to identify compounds that selectively activate fetal globin gene expression. The DNA constructs used in this screen are depicted in Fig. 2.

Promoter reporter plasmids were constructed as follows: the 1.5-kb KpnI-BglII HS2 element of the LCR was positioned 5′ of a HinfI-AluI γ^A promoter fragment (− 260 to + 36; Ney et al., 1990) and cloned upstream of the luciferase reporter gene from *P. pyralis*, generating the plasmid HS2γ-luci. The selectable marker TK-neo,

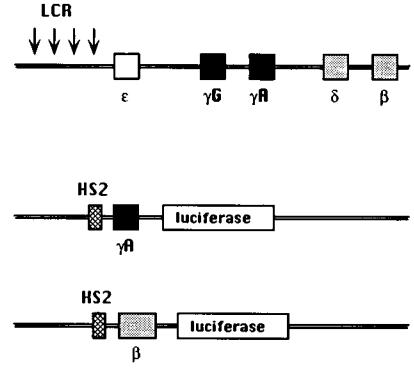

FIG. 2. Schematic representation of the δ- and β-globin promoter/luciferase fusion plasmids.

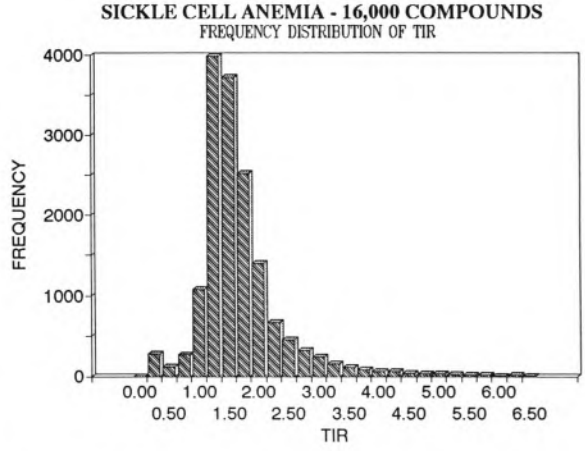

FIG. 3. Frequency distribution of 16,000 fungal extracts screened. As expected, the majority of the extracts do not have any effect on δ-globin luciferase transcription. TIR = transcription induction ratio, or the signal of the sample signal of DMSO control.

conferring resistance of transfected cells to the antibiotic G418, was inserted in *cis* at the SFiI site of HS2γ-luci. In addition, the HS2 enhancer/β-globin promoter reporter construct (HS2β-luci) was prepared to serve as a specificity control in lead follow-up assays. The TK-neo cassette was similarly inserted into the SfiI site of this plasmid. Stably transfected cell lines were generated and high throughput screening initiated. Compounds that induce γ-globin gene expression but have minimal effect on the β-globin gene will be selected for further evaluation. Extracts showing a transcription induction ratio (TIR) of > 4.0 are rescreened. The TIR is generated from the signal, in light units for luciferase, obtained from the test sample versus a solvent control. Thus, a TIR of 4.0 equates to fourfold induction.

To date, approximately 25,000 fungal extracts have been screened against this target. As predicted, the majority of extracts have little or no effect on HS2γ-luci and show a TIR at or around 1. A histogram depicting the frequency distribution of the first 16,000 compounds is shown in Fig. 3. As observed in Fig. 3, 330 compounds induced HS2γ-luci > fourfold, representing a 1.4% hit rate. These leads can be stratified on the basis of potency. For example, of the first 10,406 compounds screened, 11 compounds induced γ-globin gene expression > 10-fold. Leads can also be ranked on the basis of potency and on selectivity. Data from the sickle cell screen can be directly compared with data from other related and unrelated targets that are currently being screened using the same extracts. Thirty-six leads exhibited greater than a fourfold induction of K562/γ-globin reporter gene expression, with minimal effect on the expression of four other promoter/reporter-containing cell lines, including an additional promoter in the same K562 cell background. The further in vitro and in vivo evaluation of these active extracts is currently in progress.

Currently there are no therapeutic agents useful in treatment of sickle cell disease and β-thalassemia. Despite initial patient responses to hydroxyurea or butyrate, they are unlikely to be suitable for chronic treatment of sickle cell disease or β-thalassemia. Current sickle cell patient costs in the U.S. are estimated to be $61,000 per year. There is clear need for agents or therapeutic approaches that reduce these costs and provide effective outpatient support. The approach of transcriptional induction of γ-globin expression in human erythroid cells has the potential to overcome many of the limitations of conventional drug discovery and could lead to the development of novel pharmaceuticals targeting sickle cell anemia.

FINAL COMMENTS

To meet future clinical demand for new and cost-effective therapies will require the continuing development of novel approaches to drug discovery. We are optimistic that the transcription-based approach that we have discussed will be a valuable contribution in this area, and will lead to the development of a new generation of specific and effective pharmaceutical agents that can modulate human metabolism with a precision previously only dreamed of. Oncogene Science and its collaborators are pursuing this goal in areas as diverse as cancer, atherosclerosis, diabetes, muscular dystrophy, asthma, Alzheimer's disease, inflammation, AIDS, anemia, osteoporosis, and sickle cell disease.

REFERENCES

J. Auwerx (1992), Horm Res 38, 269–277.

P. A. Baeuerle and D. Baltimore (1991), in The Hormonal Control Regulation of Gene Transcription (P. Cohen and J. G. Foulkes, eds). Elsevier, Amsterdam.

K. Chada, J. Magnam, and F. Constantini (1986), Nature 319, 685–689.

K. Chada, J. Magnam, K. Raphael, G. Radice, E. Lacy, and F. Constantini (1985), Nature 314, 377–380.

N. Dillon and F. Grosveld (1991), Nature 350, 252–254.
Drug & Market Development (1993), 4(1/2), 8–14.
D. K. Granner and R. M. O'Brien (1992), Diabetes Care 15(3), 369–395.
F. Grosveld, G. B. van Assendelft, D. R. Greaves, and G. Kollias (1987), Cell 51, 975–985.
O. Hanscombe, D. Whyatt, P. Fraser, N. Yannoutsos, D. Greaves, N. Dillon, and F. Grosveld (1991), Genes Dev 5, 1387–1394.
B. E. Henderson, R. K. Ross, and M. C. Pike (1993), Science 259, 633–638.
S. Jane, P. Ney, E. Vanin, D. Gumucio, and A. W. Nienhuis (1992), EMBO J 11, 2961–2969.
G. Kollias, N. Wrighton, J. Hurst, and F. Grosveld (1986), Cell 46, 89–94.
E. Kopp and S. Ghosh (1994), Science 265, 956–959.
E. Marshall (1994), Science 264, 1524–1527.
P. G. McMaffrey, B. A. Perrino, T. R. Soderling, and A. Rao (1993), J Biol Chem 268(5), 3747–3752.
D. P. McDonnell, E. Vegeto, and M. A. G. Gleason (1993), Biotechnology 11, 1256–1261.
B. H. Mitlak and S. R. Nussbaum (1993), Annu Rev Med 44, 265–277.
P. Ney, B. Sorrentino, C. Lowrey, and A. W. Nienhuis (1990), Nucleic Acids Res 18, 6011–6017.
T. Rutherford and A. W. Nienhuis (1987), Mol Cell Biol 7, 398–402.
R. Sanduja, D. Loose-Mitchell, and K. K. Wu (1990), Adv Prostaglandin Thromboxane Leukot Res 21, 149–152.
S. L. Schreiber and G. R. Crabtree (1992), Immunol Today 13(4), 136–142.
R. H. Schwartz (1992), Cell 71, 1065–1068.
G. L. Semenza, S. T. Koury, M. K. Nejfelt, J. D. Gearhart, and S. E. Antonarakis (1991), Proc Natl Acad Sci USA 88, 8725–8729.
G. C. Sen and P. Lengyel (1992), J Biol Chem 267(8), 5017–5020.
B. Sorrentino, P. Ney, D. Bodine, and A. W. Niehuis (1990), Nucleic Acids Res 18, 2721–2731.
B. Staels and J. Auwerx (1992), Development 115, 1035–1043.
G. Stamatoyannopoulos and A. W. Niehuis (1992), Annu Rev Med 93, 497–521.
D. Talbot, P. Collis, M. Antoniou, M. Vidal, F. Grosveld, and D. R. Greaves (1989), Nature 338, 352–355.
D. Tuan, W. Solomon, Q. Li, and I. M. London (1985), Proc Natl Acad Sci USA 82, 6384–6388.
F. E. van Leeuwen, J. Benraadt, J. W. W. Coebergh, L. A. L. M. Kiemeney, C. H. F. Gimbrere, R. Otter, L. J. Schouten, R. A. M. Damhuis, M. Bontenbal, F. W. Diepenhorst, A. W. van den Belt-Dusebout, and H. van Tinteren (1994), Lancet 343, 448–452.
K. K. Wu, R. Sanduja, A-L. Tsai, B. Ferhanoglu, and D. S. Loose-Mitchell (1991), Proc Natl Acad Sci USA 88, 2384–2387.
X-M. Xu, A. Hajibeigi, A. Loose-Mitchell, D. Wang, and K. K. Wu (1993), Blood 82(10 Suppl. 1), 342A.
T. Zyglewska, R. Sanduja, K. Ohashi, D. S. Loose-Mitchell, S. K. Sanduja, and K. K. Wu (1992), Biochim Biophys Acta 1131, 78–82.