Metabolic Regulation of Gene Transcription in Mammals*

Howard C. Towle

From the Department of Biochemistry and Institute of Human Genetics, University of Minnesota Medical School, Minneapolis, Minnesota 55455

All cells regulate gene expression in response to changes in the environment. For unicellular organisms, specific mechanisms have evolved to allow these cells to metabolize various fuels based on their availability in the external milieu. In part, these mechanisms involve conditional transcription of genes encoding enzymes unique to specific metabolic pathways in the presence of appropriate nutrients. The study of such control mechanisms has led to several of the classic paradigms for transcriptional regulation present in today's textbooks. Thus, the lac operon of Escherichia coli and the gal regulon of Saccharomyces cerevisiae are among the best understood regulatory pathways of gene expression.

In multicellular organisms, the needs of not only the individual cell but also the whole organism must be managed. Consequently, much of the task of interpreting environmental cues in mammals is handled by hormonal and neuronal pathways. For example, the counterbalancing hormones insulin and glucagon play a major role in maintaining blood glucose levels within fairly narrow limits by controlling glucose utilization in several different tissues. Although not as widely appreciated, nutritional and metabolic signals also play an important role in controlling gene expression in multicellular organisms. This review will summarize recent work on two metabolic signals, cholesterol and glucose metabolism, which can lead to altered gene expression in mammals.

Cholesterol Metabolism

All mammalian cells require cholesterol for biogenesis of membranes. Cholesterol can be derived externally via uptake of cholesterol-containing lipoprotein particles or from de novo biosynthesis. Mammalian cells regulate these two pathways to ensure an appropriate supply of cholesterol by feedback repression of several key genes of cholesterol metabolism. The low density lipoprotein (LDL)* receptor plays a critical role in cellular uptake of cholesterol. HMG-CoA reductase and HMG-CoA synthase provide control points for the de novo biosynthetic pathway. When cellular sterol levels are low, expression of the genes involved in cholesterol biosynthesis and uptake is activated. Conversely, when sufficient cholesterol is present, the biosynthesis of these pivotal proteins is repressed. Although both transcriptional and post-transcriptional regulation is involved, transcriptional regulation is better understood and will be the focus of this review.

Goldstein and Brown (1) and their collaborators defined the critical regulatory sequences for transcriptional regulation of several sterol-regulated genes. When chimeric constructs containing the 5'-flanking regions of the genes for LDL receptor, HMG-CoA synthase, or HMG-CoA reductase linked to a reporter gene were introduced into cultured cells, promoter activity was elevated in the medium. Through mutagenesis of specific sequences in the individual promoters, sterol regulatory elements were identified. Fusion of these sequences to a heterologous promoter conferred sterol regulation. Comparison of the regulatory regions of these three genes indicated a consensus sequence: 5'CACCC/G/CAC. Mutations within this motif blocked the ability of these promoters to respond to sterols. This motif was termed the sterol regulatory element-1 or SRE-1. These observations led to the proposal that the SRE-1 represents the binding site for a common factor whose activity is modulated by sterols.

A nuclear factor from HeLa cells that bound to the SRE-1 of the gene for the LDL receptor gene was subsequently purified (2, 3). This effort was complicated by the existence of several cellular proteins capable of recognizing the SRE-1. To identify the appropriate protein, binding of various factors to mutated SRE-1 sites was correlated with the effects of the mutations on sterol regulation of promoter activity in the transfection assay. By this means, a specific factor, designated SREBP-1, was identified that was responsible for transcriptional regulation. The gene for this factor was subsequently cloned and found to be a member of the c-Myc family of transcription factors (4). The gene for a second structurally related form of SREBP, SREBP-2, has also been cloned (5). Both proteins bind to the SRE-1 element and are capable of stimulating transcription from promoters containing this element. Members of the c-Myc family contain the basic region/helix-loop-helix/leucine zipper DNA binding motif and recognize an E-box motif related to the sequence CACGTG. Thus, the recognition sequence of the LDL receptor gene, CACCC/G/CAC, differed from that recognized by other family members. However, Kim et al. (6) have recently found that SREBP-1 (which is identical to the factor ADD-1) can also recognize the CACGTG motif. The ability to recognize two related, but distinct, DNA-binding sites may allow SREBP a greater flexibility in control of gene transcription.

How does SREBP-1 sense the concentration of sterol and control gene expression of the LDL receptor gene? SREBP-1 acts as a positive transcriptional factor that is active in conditions when sterols are low (Fig. 1). It is synthesized as a precursor protein of 125 kDa that is found in the endoplasmic reticulum (7). When sterols are depleted, the membrane-bound precursor is cleaved to generate a 68-kDa NH2-terminal fragment, which enters the nucleus and binds to the SRE-1 of the LDL receptor. Transfection of a gene encoding a truncated form of SREBP-1 that removes the membrane-spanning and COOH-terminal portion leads to nuclear localization and strong activation of constructs containing SRE-1 elements in a constitutive manner (8). Interestingly, a similar truncated form of SREBP-2 is produced in a sterol-resistant cell line as a result of a gene rearrangement (9). These cells do not repress transcription of the LDL receptor in the presence of sterols, thereby implicating the SREBPs in sterol regulation. Thus, control of this pathway involves cleavage of SREBP from an inactive membrane form to an active nuclear form. The actual mechanism by which sterols control this cleavage is not yet understood. Cholesterol, or a derivative of cholesterol, as a normal substituent of the membrane, may directly influence the specific protease responsible for cleaving SREBP. Another question of importance is whether cholesterol acts directly or is metabolized to an active form. Studies using cultured cells have shown that several oxysterols, notably 25-hydroxycholesterol, are more active than cholesterol itself in promoting repression of gene transcription. Thus, it is possible that metabolism of cholesterol is required to generate the proximal signaling molecule.

The pathways for controlling gene expression by sterols are significantly more complicated. For example, the SRE-1 element does not work independently to stimulate transcription of the gene for the LDL receptor. Instead, sterol regulation requires that the SRE-1 element be situated adjacent to a binding site for Sp1. SREBP-1 and Sp1 act synergistically to activate transcription from the promoter for the LDL receptor through cooperative interactions (10). Further complexities were revealed by analysis of the HMG-CoA reductase gene. In this case, mutation of the sterol regulatory...
sequences led to constitutively high levels of transcription, which were not repressed by sterols. This is in contrast to the constitutively low transcription from promoters of the LDL receptor containing SRE-1 mutations. Thus, a fundamentally different mechanism may function for the HMG-CoA reductase gene. Saturation mutagenesis of the HMG-CoA reductase gene led to the finding that the putative SRE-1, originally defined by sequence similarity to the LDL receptor gene SRE-1, is not critical for regulation (11). Indeed, two distinct sites, one of which overlaps the SRE-1 homology region, were identified. Similarly, the regulatory region of the gene for farnesyl diphosphate synthase, another gene repressed when cholesterol levels are elevated, has been identified (12). In this case, no sequence homology to the SRE-1 was noted, and again, a unique mechanism of control was suggested. The nuclear factors responsible for controlling the transcription of these genes have not yet been identified. Given the importance of the cholesterol biosynthetic pathway in generating multiple products essential for normal cell function, it is not surprising that complex regulatory pathways have evolved to control expression of the various genes encoding enzymes of this pathway.

**Carbohydrate Metabolism**

The liver plays a key role in the handling of ingested carbohydrate in vertebrates. Excessive intake of dietary carbohydrate leads to increased triglyceride biosynthesis in the liver. This occurs through both a rapid activation of enzymes catalyzing the key rate-limiting steps as well as a longer term induction in enzyme synthesis (13–15). These responses are presumably an adaptive response that allow the organism to better utilize limiting carbohydrate in the environment by converting it to the preferred energy storage form of triglyceride. This section will focus on the mechanisms involved in induction of these enzymes of triglyceride biosynthesis, termed lipogenic enzymes, in response to carbohydrate.

Enzymes that are induced by feeding of a high carbohydrate, low fat diet are listed in Table I. In every case shown, the induction is due to an increase in mRNA levels. Transcription has been shown to be responsible at least in part for the regulation of many of these genes. However, post-transcriptional control is also likely to be important (16–18). The production of enzymes indicated in Table I is also decreased by fasting or glucagon administration or during diabetes. Polyunsaturated fatty acids repress the expression of many genes in this group (19), whereas insulin or thyroid hormone increases their synthesis. Thus, the expression of the lipogenic enzymes appears to be coordinately controlled.

Feeding of a high carbohydrate diet increases levels of plasma insulin and decreases plasma glucagon. For many years, it was thought that these hormones were directly responsible for regulating gene expression of the lipogenic enzymes. This indeed appears to be the case for glucokinase expressed from its liver-specific promoter (20). However, for several other enzymes listed in Table I, carbohydrate metabolism has been implicated in the generation of the transcriptional response. This hypothesis was most strongly supported by work in primary cultured hepatocytes. Many of the lipogenic enzymes are induced by increasing the glucose concentration (and hence glucose metabolism) in cultured hepatocytes in the face of a fixed concentration of insulin (21–23). For example, several different carbohydrate substrates capable of being metabolized in the glycolytic pathway are capable of supporting the increased biosynthesis of malic enzyme, whereas non-metabolizable analogs of glucose are not (24). Thus, the primary signal for enzyme induction was generated in response to increased carbohydrate metabolism, and insulin acted indirectly to support this process.

Glucokinase may represent the key insulin-dependent step in supporting the metabolic response of the hepatocyte for L-type pyruvate kinase (L-PK) gene expression (15). If hepatocytes are treated with low concentrations of fructose to stimulate glucokinase activity independently of insulin, the induction of transcription from the L-PK promoter does not require insulin (25). Similarly, introduction into hepatocytes of a vector that constitutively expresses glucokinase results in an insulin-independent glucokinase activity. Thus, the role of insulin in the transcriptional induction of the L-PK gene is to promote glucose metabolism via the stimulation of glucokinase.

**The Carbohydrate Response Element**—To explore the pathway by which increased carbohydrate metabolism leads to altered gene transcription of lipogenic enzymes in the liver, we set out to define the critical regulatory sequences for this response. We chose two genes for these efforts: L-PK and S14. Both of these genes are induced at the level of transcription (27, 28). S14 is a gene originally detected based on its rapid response to thyroid hormone. S14 mRNA is also rapidly induced by feeding rats carbohydrate, with changes in mRNA levels detectable within 30 min of treatment (22). The physiological role of the S14 gene product is unknown. Based on its tissue distribution and responses to various effectors, S14 has been suggested to play a role in some aspect of lipid metabolism (29).

Chimeric constructs containing the 5'-flanking region of the L-PK or S14 genes linked to a reporter gene were introduced into primary hepatocytes by lipofection. Cells grown in the presence of elevated glucose had an increased expression of reporter gene transcription compared with cells grown in low glucose (30, 31). This increase was specific, as the promoters of several other genes expressed in hepatocytes did not show this response when similarly tested. Using this assay, the control elements of both of these genes were mapped; for L-PK, the critical region included sequences from 172 to 124 (32, 33), while for S14, the essential sequences were from 1457 to 124 (34). Comparison of the regulatory regions of the L-PK and S14 genes revealed a sequence with 34 out of 36 bp
identity, suggesting that a common regulatory factor is involved in controlling expression of these two genes. This region of similarity is centered by a CACGTG motif, the core binding site for the c-Myc family of transcription factors (34).

Analysis of the regulatory sequences of the L-PK gene indicated two factor-binding sites (33, 35). An oligonucleotide corresponding to one of these sites bound to the factor USF, a ubiquitously expressed member of the c-myc family. An oligonucleotide comprising the other site was recognized by the hepatic enriched factor, HNF-4, an orphan receptor of the steroid/thyroid receptor family. Chimeric gene constructs in which an oligonucleotide corresponding to either the USF- or HNF-4-binding sites was linked to the basal L-PK promoter (gene sequences from -96 to +12) did not respond transcriptionally to glucose when introduced into hepatocytes. Similar constructs containing both sites did respond. Interestingly, constructs that contained multiple copies of the USF-binding site, but not multiple copies of the HNF-4 site, showed a robust transcriptional response to glucose (32, 33). Furthermore, the USF-binding site of the L-PK gene contained the region with sequence similarity to the regulatory region of the S14 gene. These observations suggested that the USF-binding site interacts with the factor that receives the signal generated by increased glucose metabolism. This site was termed a ChoRE for carbohydrate response element. It has also been called a GIRE or glucose/insulin response element (32). HNF-4 serves as an accessory factor that functions together with the carbohydrate-responsive factor to activate gene transcription. In this way, with a 5-bp spacing between the two USF-binding sites, transcription is induced by sequences that are functionally analogous to the Sp1-binding site of the LDL receptor gene in promoting a sterol response.

The S14 regulatory region also contains two sites involved in supporting the response to carbohydrate: a USF-binding site and an accessory site (36). Again chimeric gene constructs containing multiple copies of the USF-binding site were capable of giving a transcriptional response to glucose in hepatocytes, whereas multiple copies of the accessory site did not. Interestingly, the accessory factor belonging to the Sp1 family did not appear to be a functional factor that has not yet been defined. Synergistic interactions between multiple transcription factors are commonly involved in gene regulation. Such systems provide multiple sites of regulation that allow integration of metabolic and hormonal signals. For instance, Lii and coworkers (37) recently mapped a site in the L-PK promoter that is responsible for transcriptional repression by polyunsaturated fatty acids. This site corresponded to the HNF-4 site of the L-PK gene. Thus, two factors binding to adjacent regulatory sites receive distinct metabolic inputs and act in a coordinated manner to regulate gene transcription.

The nature of the carbohydrate responsive factor remains unknown. Substitution of an authentic USF-binding site from the adenovirus major late promoter in place of the ChoREs of either L-PK or S14 genes did not reconstitute a response to glucose (33, 38). Thus, binding of USF alone is not sufficient to render a response. This result is not unexpected, as many genes expressed in the liver contain USF-binding sites but do not respond to increased glucose metabolism. What then is the basis for specificity with respect to glucose responsiveness? An examination of the ChoREs of the L-PK and S14 genes provides some intriguing clues. In the L-PK USF-binding site, two imperfect CACGTG motifs separated by 5 bp are found (Fig. 2). Each motif contains a 5 out of 6-base bp match to the c-Myc family consensus-binding site. Mutations in either of the two motifs result in a loss of the glucose response (32). Base substitution mutations in the 5-bp spacer separating the two CACGTG motifs did not disrupt the response to glucose (36). On the other hand, mutations that altered the distance between the two motifs dramatically affected the ability of this element to respond. Thus, a single bp deletion or a single bp insertion essentially eliminated the ability to respond to glucose. The S14 ChoRE has a similar arrangement; a single perfect CACGTG motif is separated by 5 bp from the sequence CCTGTTG with a 4 out of 6 bp match to the consensus. Again, mutation of either motif disrupts the response to glucose. Furthermore, converting the CCTGTTG motif to the CACGTG motif in this region of the S14 gene leads to a dramatic increase in the ability of the site to respond to glucagon. The latter (two perfect CACGTG motifs separated by 5 bp) no longer requires an accessory factor in order to respond to glucose. However, the

**Fig. 2. Comparison of the regulatory sequences of the L-type pyruvate kinase and S14 genes that are required for the transcriptional response to carbohydrates.** In both genes, ChoREs with similar arrangements are found that contain two CACGTG motifs (underlined) separated by 5 bp. Correct spacing of these elements is critical for carbohydrate control of gene transcription. Multiple copies of either the L-PK or S14 ChoREs can support a glucose response when linked to a basal promoter. However, in their natural context each ChoRE requires an adjacent accessory factor-binding site (AF) to support the response. For the L-PK gene, HNF-4 serves as an accessory factor, whereas the accessory factor for the S14 gene is an unidentified nuclear protein distinct from HNF-4. 5-bp spacing of these two motifs remains critical for maintaining glucose control. The strict spacing requirement of the two CACGTG motifs suggests that two identical or closely related factors may bind to provide the carbohydrate response noted for transcription of the L-PK or S14 genes. These two factors either directly contact each other or form a precise surface for interaction with a third factor. These factors have not been detected using in vitro DNA binding experiments.

- Further Directions—What is the nature of the signaling pathway that is responsible for the glucose response in hepatocytes? Two different possibilities have been suggested. Based on work in adipocytes, Girard and co-workers (39) suggested that glucose 6-phosphate may be the critical regulatory component. In this study, addition of 2-deoxyglucose, but not 3-O-methylglucose, stimulated expression of the genes for fatty acid synthase and acetyl-CoA carboxylase in cultured adipocytes. The former is capable of being metabolized to 2-deoxyglucose 6-phosphate but not further, whereas the latter is derived from the glycolytic pathway at or above pyruvate could stimulate a response. Interpretation of this result is complicated by the gluconeogenic potential of the hepatocyte. However, a mitochondrial origin for the mediator is supported by the observation that dichloroacetate, a stimulator of pyruvate dehydrogenase activity, also led to increased synthesis of malic enzyme (24). This latter response did not require insulin and was not inhibited by glucagon. Thus, the signal responsible for increased gene expression was proposed to have arisen at or downstream from pyruvate oxidation. Clearly, more work is necessary to elucidate this critical step in the process.

Equally obscure is the nature of the carbohydrate responsive factor. Since the adenovirus USF-binding site could not substitute for the ChoRE in either the L-PK or S14 genes, USF binding is not sufficient for carbohydrate signaling. However, USF could be part of a complex that receives the carbohydrate signal. Recent work from the Kahn laboratory (41) using dominant negative forms of USF suggests that this possibility may be correct. On the other hand, the c-Myc family of transcription factors is a large and growing family, and many other candidates are present in the hepatocyte. The recent observation that SREBP-1 has the ability to bind to either the SRE-1 element or to the CACGTG motif suggests this factor and its closely related forms could be candidates (6).

Furthermore, members of the c-Myc family are capable of heterodimerizing with each other in a specific combinatorial fashion, thus increasing the potential diversity of the regulatory process.

Finally, the question of how the carbohydrate transcription factor is activated needs to be considered. Although highly speculative, two general models might be envisioned. In the first, an
intermediate or by-product of metabolism (e.g. glucose 6-phosphate) could serve as a direct activator by binding to the carbohydrate responsive factor. An example of this type of regulation is provided by the peroxisome proliferator-activated receptor (PPAR), a member of the thyroid/retinoic acid nuclear receptor family. This receptor activates several genes involved in fatty acid oxidation by binding to DNA response elements of these genes (42). In a fashion analogous to other nuclear receptors, direct binding of a ligand to the PPAR promotes a conformational change in the receptor to create the transactivation surface. Although the exact nature of the ligand binding to the PPAR is presently unknown, metabolism of various fatty acids can lead to receptor activation (42, 43). Similarly, a second member of the nuclear receptor family has recently been shown to be directly activated by farnesol and its metabolites (44). Thus, a growing class of metabolite-regulated signaling is emerging in vertebrates. In the second model, covalent modification of the carbohydrate responsive factor could be responsible for activation. It is well recognized that the activity of several of the lipogenic enzymes is regulated by phosphorylation. For example, the activity of the key rate-limiting enzyme for fatty acid biosynthesis (acetyl-CoA carboxylase) is inhibited by phosphorylation via the AMP-activated protein kinase, an enzyme activated in conditions of glucose deprivation (45). Similarly, a kinase or phosphatase regulated in response to carbohydrate metabolism might modify the carbohydrate responsive factor. Alternatively, the carbohydrate responsive factor could be activated by the redox potential of the hepatocyte, as NADPH is utilized in the reductive synthesis of fatty acids. Interestingly, the DNA binding potential of USF is strongly affected by changes in the redox state via modification of two cysteine sulfhydryl groups (46). Elucidation of the mechanism of transcriptional activation will likely await identification of the carbohydrate responsive factor and examination of its control.

Acknowledgments—I thank the many excellent students and collaborators who have worked with me over the years on this project. Special thanks to Hsiu-Ming Shih, Elizabeth Newburg, Karl Olson, and Ravindra Uppaluri for reading the manuscript.

REFERENCES
1. Goldstein, J. L., and Brown, M. S. (1990) Nature 343, 425–430.
2. Briggs, M. R., Yokoyama, C., Wang, X., Brown, M. S., and Goldstein, J. L. (1993) J. Biol. Chem. 268, 14490–14496.
3. Wang, X., Briggs, M. R., Hua, X., Yokoyama, C., Goldstein, J. L., and Brown, M. S. (1993) J. Biol. Chem. 268, 14497–14504.
4. Yokoyama, C., Wang, X., Briggs, M. R., Admon, A., Wu, J., Hua, X., Goldstein, J. L., and Brown, M. S. (1993) Cell 75, 187–197.
5. Hua, X., Yokoyama, C., Wu, J., Briggs, M. R., Brown, M. S., Goldstein, J. L., and Wang, X. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 11603–11607.
6. Kim, J. B., Spotts, G. D., Halvorsen, Y.-D., Shih, H.-M., Ellenberger, T., Towle, H. C., and Spiegelman, B. M. (1995) Mol. Cell. Biol. 15, 2582–2588.
7. Wang, X., Sato, R., Brown, M. S., Hua, X., and Goldstein, J. L. (1994) J. Biol. Chem. 269, 53–62.
8. Sato, R., Yang, J., Wang, X., Evans, M. J., Ha, Y. K., Goldstein, J. L., and Brown, M. S. (1994) J. Biol. Chem. 269, 17267–17273.
9. Yang, J., Sato, R., Goldstein, J. L., and Brown, M. S. (1994) Genes & Dev. 8, 1910–1919.
10. Sanchez, H. B., Yieh, L., and Osborne, T. F. (1995) J. Biol. Chem. 270, 1161–1168.
11. Osborne, T. F. (1991) J. Biol. Chem. 266, 13947–13951.
12. Spear, D. H., Ericsson, J., Jackson, S. M., and Edwards, P. A. (1994) J. Biol. Chem. 269, 25212–25218.
13. Goodridge, A. (1987) Annu. Rev. Nutr. 7, 157–185.
14. Lemaigre, F. P., and Rousseau, G. G. (1994) Biochem. J. 303, 1–14.
15. Vaulont, S., and Kahn, A. (1994) FASEB J. 8, 233–35.
16. Dozin, B., Rall, J. E., and Nikodem, V. M. (1996) Proc. Natl. Acad. Sci. U. S. A. 83, 4705–4709.
17. Baum, C. L., Teng, B.-B., and Davidson, N. G. (1990) J. Biol. Chem. 265, 19263–19270.