The isolation and characterization of the gene encoding malic enzyme in the chicken

Fantozzi, Dominic Alexander, Ph.D.

Case Western Reserve University (Health Sciences), 1990
THE ISOLATION AND CHARACTERIZATION OF
THE GENE ENCODING MALIC ENZYME IN THE CHICKEN

by

DOMINIC A. FANTOZZI

Submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Thesis Advisor: Alan G. Goodridge, Ph.D.

Department of Biochemistry
CASE WESTERN RESERVE UNIVERSITY
January 11, 1990
CASE WESTERN RESERVE UNIVERSITY
GRADUATE STUDIES

We hereby approve the thesis of

Dominic A. Fantozzi

candidate for the Ph.D.
degree.*

Signed:  
John W. Nelson
(Chairman)

Cecil Cooper

David Sauers

William E. Menden

Date August 28, 1989

*We also certify that written approval has been obtained for any proprietary material contained therein.
I grant to Case Western Reserve University the right to use this work, irrespective of any copyright, for the University's own purposes without cost to the University or to its students, agents and employees. I further agree that the University may reproduce and provide single copies of the work, in any format other than in or from microforms, to the public for the cost of reproduction.

[Signature]
THE ISOLATION AND CHARACTERIZATION OF
THE GENE ENCODING MALIC ENZYME IN THE CHICKEN

Abstract
by
DOMINIC A. FANTOZZI

The abundance of the mRNA encoding malic enzyme (L-malate,NADP⁺-oxidoreductase (decarboxylating), EC 1.1.1.40) is regulated by the nutritional status of the whole animal and the hormonal milieu of cultured primary hepatocytes. Accumulation of the mRNA is stimulated by consumption of high carbohydrate, low fat diets. Accumulation of the mRNA is inhibited by fasting. In avian hepatocytes cultured in chemically-defined media, accumulation of the mRNA is stimulated by insulin plus thyroid hormone and inhibited by glucagon or cAMP. Abundance of malic enzyme mRNA is controlled by regulating both the rate constant for its degradation and the rate of initiation of transcription of the cognate gene. The purpose of these studies was to isolate and partially characterize genomic sequences potentially required for regulation of the accumulation of malic enzyme mRNA by both diet and hormones. Toward that end, we have initiated an analysis of the single-copy gene encoding malic enzyme in the avian liver.

A malic enzyme cDNA clone derived from the goose uro-
pygial gland was used to isolate homologous cDNA clones from duck liver and chicken liver cDNA libraries. These clones, in turn, were used to isolate and partially characterize additional cDNA clones derived from embryonic chick hepatocytes cultured in the presence of both insulin and thyroid hormone. One of these clones, pCME5, contains an insert of ~2000 bp and represents >95% of chicken malic enzyme mRNA.

The availability of malic enzyme cDNA clones facilitated the isolation and partial characterization of a series of genomic clones constituting part of the malic enzyme locus in the chicken. These clones span ~100 kb of non-overlapping genomic DNA. Gaps exist between some clones, hence the exact length of the transcription unit is unknown. The 3'-most genomic clone, λCME1, contains a consensus sequence for polyadenylation, thus likely encodes the 3' end of the mRNA. The 5'-most 254 nt of the cDNA are encoded by two exons that are separated by an intron of at least 39 kb. Hybridization and sequencing analyses of the genomic clones suggest the presence of at least 11 exons, thus at least 10 introns, within the malic enzyme gene.

The 5'-most genomic clone, λCME245, was shown by primer-extension and S1 nuclease analyses to encode the 5' end of the mRNA. Transcription initiates at multiple sites in liver and hepatocytes, but primarily at a site corre-
sponding to the 5' end of the cDNA in pCME5. Sequence analysis of the putative promoter region revealed the absence of "CCAAT box" and "TATA box" homologies, consistent with multiple sites of transcription initiation. The 5' flanking region also lacks putative binding sites for the transcription factor Sp1. The 5' flanking region contains distinct sequences that exhibit (a) similarity to T3-response elements found in other T3-responsive genes, (b) similarity to the AP-2 sites within the genes encoding rat prolactin and tyrosine aminotransferase, and (c) identity to the AP-2 site of human proenkephalin. Also identified were a putative binding site for the transcription factor AP-1 and a poly(pyr·pur) sequence which may be involved in the regulation of transcription.
Dedication

To Rosamaria, without whose encouragement, understanding and affection this endeavor would have never been completed.

Quotation

"...Don't pay the ferryman, don't even fix a price, Don't pay the ferryman, 'til he gets you to the other side..."

Chris DeBurgh in "Don't Pay the Ferryman" from The Getaway
Acknowledgements

I thank the following people who enhanced my experiences as a graduate student:

Dr. Alan Goodridge, my advisor, for placing his confidence in me, being patient, and aiding in my maturation as a scientist by serving as a model of discipline, persistence, and intellectual integrity; Drs. John Nilson, Marcelo Jacobs-Lorena, and Cecil Cooper, the other members of my pre-thesis committee, for providing constructive criticisms throughout the course of my project;

Drs. Bill Merrick and David Samols, for serving on my defense committee;

my associates in Alan's laboratory: Judy Fisch, Tim Uyeki, Dave Mitchell, Mike McDevitt, Paul Huang, Steven Klautky, Manuel Glyniass, Xiao-jun Ma, and Drs. Mitchell Goldman, Brian Wilson, Bill Carpenter, Sue Stapleton, Jim Crish, Lisa Salati, Brad Hillgartner, Ken Kameda, Nat Ross, Susumu Kawamoto, Ray Ochs, Donald Back, Larry Winberry, Bob Jenik, and Sid Morris for providing a unique combination of scientific insight, humor, culinary expertise, and unsolicited financial advice (the latter prior to the crash of the stock market); all members of John Nilson's laboratory, for sharing their facilities with me during the final months of my training;
Allan Parrish, Bill Magner, Lori Rusterholz, and Drs. Dave Reinhold, Julie C. Kennedy, Sharon Weldon, Don Anthony, Joyce Jentoft, Dave Samols, Ray Goodwin, Tim Nilsen, Ken Neet, Story Landis, Ruth Siegel, and Richard Hanson for sharing valuable insights and friendships;

the remaining members of the departments of Biochemistry and Pharmacology, too numerous to mention, for establishing and maintaining a professional environment in which to work and learn;

the National Institutes of Health, from whom funds were obtained by Alan to support my training.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION AND QUOTATION</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td><strong>CHAPTERS</strong></td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A. Perspective</td>
<td>1</td>
</tr>
<tr>
<td>B. Biosynthesis of Saturated Fatty Acids</td>
<td>2</td>
</tr>
<tr>
<td>C. Sites of Biosynthesis of Fatty Acids</td>
<td>4</td>
</tr>
<tr>
<td>D. Nutritional Regulation of the Biosynthesis of Fatty Acids</td>
<td>5</td>
</tr>
<tr>
<td>E. Hormonal Regulation of the Biosynthesis of Fatty Acids</td>
<td>7</td>
</tr>
<tr>
<td>F. The Role of Malic Enzyme in Hepatic Lipogenesis; Nutritional and Hormonal Regulation of Enzyme Activity</td>
<td>11</td>
</tr>
<tr>
<td>G. Molecular Cloning of DNA Sequences Complementary to Malic Enzyme mRNA</td>
<td>13</td>
</tr>
<tr>
<td>H. Regulation of the Abundance of Malic Enzyme mRNA by Nutritional Status in vivo and by Hormones in Cultured Primary Hepatocytes</td>
<td>14</td>
</tr>
<tr>
<td>I. Statement of the Problem</td>
<td>21</td>
</tr>
<tr>
<td>II. EXPERIMENTAL PROCEDURES</td>
<td>23</td>
</tr>
<tr>
<td>A. Materials</td>
<td>23</td>
</tr>
<tr>
<td>B. Methods</td>
<td>26</td>
</tr>
<tr>
<td>1. Identification of Malic Enzyme cDNA Clones</td>
<td>26</td>
</tr>
<tr>
<td>2. Identification of Malic Enzyme Genomic Clones</td>
<td>27</td>
</tr>
<tr>
<td>3. Characterization of Cloned DNAs Derived from the Malic Enzyme Locus</td>
<td>28</td>
</tr>
<tr>
<td>4. Isolation and Southern Analysis of DNA</td>
<td>28</td>
</tr>
<tr>
<td>5. Isolation and Northern Analysis of RNA</td>
<td>29</td>
</tr>
<tr>
<td>6. Sequence Analysis</td>
<td>30</td>
</tr>
<tr>
<td>7. Primer-extension Analysis</td>
<td>30</td>
</tr>
<tr>
<td>8. S1 Nuclease Analysis</td>
<td>31</td>
</tr>
<tr>
<td>9. Preparation of Molecular Probes</td>
<td>33</td>
</tr>
</tbody>
</table>

viii
III. RESULTS ......................................................... 36

A. Characterization of Cloned DNA Sequences
   Complementary to Malic Enzyme mRNA ............... 36
B. Assessment of the Copy Number of the Gene
   Encoding Malic Enzyme ................................. 44
C. Isolation and Characterization of Cloned
   Genomic DNA from the Malic Enzyme Locus ....... 44
D. Structural Analysis of the Putative Promoter .. 72

IV. DISCUSSION ..................................................... 87

A. Introduction ................................................. 87
B. Transcription of Eukaryotic Genes and Its
   Regulation .................................................. 90
   1. Regulation by Thyroid Hormone .................... 92
   2. Regulation by Glucagon .............................. 94
   3. Transcription mediated by the AP-1 motif ...... 96
   4. Regulation involving conformational changes
      in DNA .................................................. 96
C. Regulation of Stability of mRNA ..................... 99
D. Future Prospects .......................................... 101

V. REFERENCES .................................................... 103
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Determination of the size of malic enzyme mRNA by Northern analysis</td>
<td>16</td>
</tr>
<tr>
<td>2. Restriction maps of cloned DNAs complementary to avian malic enzyme mRNA</td>
<td>38</td>
</tr>
<tr>
<td>3. Identities in the nucleotide sequences for malic enzyme cDNAs of rat, mouse, chicken, duck, and goose</td>
<td>43</td>
</tr>
<tr>
<td>4. Determination of the copy number of the malic enzyme gene in the haploid genome of the goose</td>
<td>46</td>
</tr>
<tr>
<td>5. Southern analyses of total genomic DNA isolated from both duck liver and chicken liver</td>
<td>49</td>
</tr>
<tr>
<td>6. Partial restriction map of λCME1, a recombinant bacteriophage that encodes the 3' end of chicken malic enzyme mRNA</td>
<td>53</td>
</tr>
<tr>
<td>7. Southern analysis of total genomic DNA from chicken liver using an homologous probe representing the 5'-most 0.3 kb of pCME5</td>
<td>57</td>
</tr>
<tr>
<td>8. Partial restriction maps of genomic clones encoding portions of chicken malic enzyme mRNA</td>
<td>61</td>
</tr>
<tr>
<td>9. Evaluation of potential restriction site polymorphisms involving the malic enzyme locus of the chicken</td>
<td>64</td>
</tr>
<tr>
<td>10. Southern analysis of total genomic DNA from chicken liver using the 5'-most EcoRI fragment of λCME20B.</td>
<td>69</td>
</tr>
<tr>
<td>11. Partial restriction map of λCME245, a recombinant bacteriophage that encodes the 5' end of chicken malic enzyme mRNA</td>
<td>74</td>
</tr>
<tr>
<td>12. Consensus nucleotide sequence of the coding strand of the promoter region of the chicken malic enzyme gene</td>
<td>77</td>
</tr>
<tr>
<td>13. Primer-extension analysis of chicken liver malic enzyme mRNA</td>
<td>80</td>
</tr>
</tbody>
</table>
14. S1 nuclease analyses of malic enzyme mRNA isolated from both chick-embryo hepatocytes and chicken liver ........................................... 84
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>cAMP</td>
<td>adenosine 3',5' monophosphate</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>CRE</td>
<td>sequence motif mediating response to cAMP</td>
</tr>
<tr>
<td>CREB</td>
<td>the trans-acting factor cognate to the CRE</td>
</tr>
<tr>
<td>dATP</td>
<td>2'-deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>2'-deoxycytidine 5'-triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>2'-deoxyguanosine 5'-triphosphate</td>
</tr>
<tr>
<td>(nx) Denhardt's</td>
<td>indicated multiple of 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% BSA</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>ds</td>
<td>double-stranded</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>hr</td>
<td>hour(s)</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>ME</td>
<td>malic enzyme</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>MOPS</td>
<td>morpholinopropane sulfonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotides</td>
</tr>
<tr>
<td>OL</td>
<td>oligodeoxynucleotide</td>
</tr>
<tr>
<td>Pipes</td>
<td>1,4-piperazinediethanesulfonic acid</td>
</tr>
<tr>
<td>poly(A)^+ RNA</td>
<td>preparation of RNA molecules containing a series of adenylate residues at their 3' ends</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>rp</td>
<td>randomly primed</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>(nx) SSC</td>
<td>indicated multiple of 0.15 M NaCl, 0.015 M sodium citrate</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>T₃</td>
<td>3,5,3'-triiodo-L-thyronine</td>
</tr>
<tr>
<td>T₃RE</td>
<td>sequence motif mediating response to T₃</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)amino methane</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TTP</td>
<td>2'-deoxythymidine 5'-triphosphate</td>
</tr>
</tbody>
</table>
Chapter I

INTRODUCTION

A. **Perspective**

The cell-specific modulation of gene expression requires coordinate regulation of the ensemble of genes defining a particular phenotype. The mechanisms by which regulation is achieved can be addressed, in part, by isolating and assessing the function of sequences that potentially mediate control. The advent of recombinant DNA techniques has permitted the relatively facile preparation and manipulation of cloned sequences, including the evaluation of function in appropriate systems of expression.

The vertebrate liver is an ideal organ from which to obtain both general and specific insights into the mechanisms of gene regulation. It exhibits developmental changes in isoenzymes, synthesis of serum proteins, and alterations in metabolism mediated by both hormones and nutrients. Accordingly, several genes have been cloned and characterized including those encoding albumin(1), alpha-fetoprotein (2), glycolytic enzymes (3,4), gluconeogenic enzymes (5,6), and the hormonally regulated α2μ globulins (7). Structural and functional analyses of the cloned sequences have advanced our understanding of the cellular mechanisms underlying responses to changes in the extracellular environment.

Changes in the nutritional status of vertebrates pro-
foundly affect the expression of various hepatic genes. Expression of the genes involved in the biosynthesis of glucose (5,6), cholesterol (8,9), and fatty acids (reviewed in 10,11) has been studied extensively, exemplifying mechanisms of regulation of gene expression mediated by both hormones and metabolites. In each case, changes in the rates of enzyme synthesis reflect changes in the abundance of the corresponding mRNA, indicating control by one or more pretranslational processes.

The mechanisms underlying the coordinate regulation of the lipogenic enzymes are the focus of our laboratory. Specifically, our long-term objectives are to (a) identify the hormones that alert hepatocytes to changes in the rate of alimentation of the intact animal and (b) deduce the molecular events coupling the binding of each hormone to its receptor with changes in the abundance of specific mRNAs. Toward that end, series of both cDNA and genomic clones encoding malic enzyme have been isolated and partially characterized. We are now poised to assess their functional significance, both in cultured hepatocytes and in transgenic animals.

B. Biosynthesis of Saturated Fatty Acids

The de novo synthesis of saturated fatty acids is catalyzed by two multifunctional enzymes, acetyl-CoA carboxylase and fatty acid synthase. Acetyl-CoA carboxylase cata-

lyzes the biotin-dependent conversion of acetyl-CoA to malonyl-CoA, the immediate precursor of 14 of the 16 carbon atoms of palmitate:

\[
\text{acetyl-CoA} + \text{HCO}_3^- + \text{H}^+ + \text{ATP} \rightleftharpoons \text{malonyl-CoA} + \text{ADP} + \text{P}_i
\]

Fatty acid synthase, using a single molecule of acetyl-CoA as primer, catalyzes the successive condensation of seven malonyl residues with the concomitant release of seven molecules of CO₂:

\[
\text{acetyl-CoA} + 7 \text{malonyl-CoA} + 14 \text{NADPH} + 14 \text{H}^+ \rightleftharpoons
\]

\[
\text{CH}_3(\text{CH}_2)_{14}\text{COO}^- + 7 \text{CO}_2 + 8 \text{CoA} + 14 \text{NADP}^+ + 6 \text{H}_2\text{O}
\]

Palmitate is the major product of these sequential reactions. Additional fatty acids required for the synthesis of phospholipids or other complex lipids are synthesized from palmitate or stearate by elongation and desaturation.

The NADPH required for the two reductive steps of the overall reaction is obtained from two sources: the dehydrogenases of the pentose phosphate pathway and malic enzyme. The former, glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase, participate in the conversion of glucose 6-phosphate to ribulose 5-phosphate. The latter catalyzes the oxidative decarboxylation of malate to pyruvate and CO₂:

\[
\text{L-malate} + \text{NADP}^+ \rightleftharpoons \text{pyruvate} + \text{CO}_2 + \text{NADPH} + \text{H}^+
\]

In the livers of vertebrates subjected to both dietary
and hormonal manipulations, the activities of these enzymes correlate positively with the rates of de novo biosynthesis of fatty acids (12). This investigation was directed exclusively toward the regulation of the activity of avian hepatic malic enzyme. Hence, it will be the specific subject of the balance of this dissertation.

C. Sites of Biosynthesis of Fatty Acids

Excluding the lactating mammary gland, liver and adipose tissue are the principal sites of synthesis of fatty acids. The apportioning of lipogenic capacity between the two tissues differs between mammals and birds. In ruminants, both liver and adipose tissues contribute significantly to the total biosynthesis of fatty acids (13). In rodents, more than half of total synthesis of fatty acids occurs in the adipose tissue while the liver contributes no more than five percent (14,15). In contrast, hepatic tissue is the major, if not only, site of synthesis in the pigeon (16) and chicken (17,18).

A proper understanding of the regulation of fatty acid synthesis in birds clearly requires examination of hepatic metabolism. Indeed, fat deposition in mammals is primarily regulated by insulin acting upon the adipose tissue. Such control is not operative in birds because adipose tissue lacks adequate capacity for the conversion of glucose to fatty acids (16), reflecting its poor or complete lack of
sensitivity to the lipogenic action of insulin (17). This difference in capacity is reflected by the activity of malic enzyme, which correlates with the capacity of a tissue to synthesize fatty acids. Producing cytoplasmic reducing equivalents, as NADPH, for the synthesis of fatty acids is a critical function of malic enzyme in avian liver. In contrast to rats (19), the activities of the pentose phosphate dehydrogenases are low and are not induced significantly by feeding a previously fasted animal (20,21).

D. Nutritional Regulation of the Biosynthesis of Fatty Acids

The requirements of most cells for fatty acids are satisfied primarily by dietary lipids rather than by de novo biosynthesis of fatty acids. In the liver, however, the maximum rate of de novo synthesis can exceed that required for membrane biosynthesis by several orders of magnitude. The primary function of hepatic lipogenesis is converting excess dietary carbohydrate or protein to fatty acids. The fatty acids are stored as triglyceride in adipose tissue and used as a source of energy when absorption of calories from the gut is limited. In man, the quantity of energy stored in adipose tissue is equivalent to forty or more days of active metabolism (22) and essential for survival under conditions of starvation.

Regulation of fatty acid synthesis in the liver is consonant with this process. Thus, hepatic synthesis of
long-chain fatty acids is inhibited by fasting and the consumption of a high-fat, low-carbohydrate diet (23,24). Conversely, refeeding a previously fasted animal stimulates fatty acid synthesis to normal levels or supranormal levels if the diet is high in carbohydrate (23-25). Thus, nutritional regulation of flux through the lipogenic pathway confers tolerance to variations in the availability of food. This may occur seasonally, or, in animals with relatively high metabolic rates, daily during periods when food is not consumed.

These metabolic responses are exemplified by the embryonic and growing chick. The rate of fatty acid synthesis in birds at the time of hatching is very low. This is consistent with the nutritional content of the egg which is high in fat and low in carbohydrate (26,27,28). The rate of lipogenesis increases approximately 1,000-fold in chicks that are fed a diet high in carbohydrate(26,27). No changes are observed in chicks that are not fed (26,27). In older chicks, fatty acid synthesis is decreased by fasting and increases upon refeeding following a fast (20,28). The total activities of the lipogenic enzymes and the rates of biosynthesis of fatty acids have been positively correlated upon feeding of neonatal chicks and fasting, or fasting and then refeeding, of older animals (29,30,31). The physiological significance of altering the activities of the en-
zymes is exemplified by the following finding: although total activity of malic enzyme in unfed hatchlings exceeds the maximum rate of fatty acid synthesis from acetate or glucose in liver slices from the same animals, it would be incapable of supporting the rate in slices from fed animals (26,27).

Similar results have been observed in both the rat and mouse. The capacity of rat liver to synthesize fatty acids declines markedly upon fasting. Realimentation returns the capacity to normal levels, or levels far above normal if diets lacking fat are used (32). Hepatic lipogenesis in vivo is significantly reduced if a diet based on glucose or fructose includes fat (33).

This sensitivity of lipogenesis to dietary fat is consistent with its regulation in the perinatal rat. Hepatic lipogenesis declines dramatically at parturition, presumably since fat is the major source of energy consumed by a suckling rat (34,35). Spontaneous weaning to the diet of the adult involves a progressive increase in the proportion of energy available as carbohydrate at the expense of fat and correlates temporally with significant increases in both lipogenesis and the activity of fatty acid synthase (36). These significant changes in metabolic flux can be initiated by weaning suckling pups prematurely (36).

E. Hormonal Regulation of the Biosynthesis of Fatty Acids

In vivo, nutritional regulation of lipogenesis is
likely mediated by hormones. The plasma concentrations of insulin, glucagon, and thyroid hormone fluctuate in response to changes in nutritional status. The changes are the bases for their proposed roles as positive (insulin and T₃) or negative (glucagon) effectors of hepatic lipogenesis during the transitions between the fed and fasted states. The levels of insulin and glucagon are low and high, respectively, in the animal which is fasting or consuming a diet that is low in carbohydrate. Conversely, their levels are high and low, respectively, in an animal feeding ad libitum. Simply stated, rates of secretion of insulin and glucagon are directly and inversely related, respectively, to the concentration of glucose in the blood. The serum concentration of T₃ decreases with fasting and increases in parallel with that of glucose upon feeding a previously fasted animal.

Insulin communicates the nutritional status of the whole animal to both the liver and adipose tissue. In tissues isolated from diabetic animals, rates of synthesis of fatty acids are comparable to those determined for tissues from starved animals, hence very low relative to those in tissues from normal animals (37-40). Administering either insulin or a diet abundant in fructose to the diabetic animal restores rates of fatty acid synthesis to normal, as assessed in the liver slices of the animals (38-40). In vitro addition of insulin to isolated adipose tissue and
liver preparations from normal rats, including cultures or suspensions of isolated hepatocytes, stimulates synthesis of fatty acids (41-43). Curiously, insulin fails in vivo or in vitro to restore fatty acid synthesis in tissues from fasted rats (44,45). Hence, additional factors are required for regulation.

Glucagon is also critical to the regulation of synthesis of fatty acids. Glucagon inhibits synthesis in liver in vivo (46,47), slices of liver, and in isolated hepatocytes (48,49). Glucagon inhibits stimulation of hepatic fatty acid synthesis that normally results from refeeding starved rats (50), but has no such effect upon liver slices obtained from animals that were previously fasted (51). These effects may be dominant to those of insulin. The high glucagon level in the diabetic animal is lowered by the administration of insulin (52), suggesting that insulin deprivation may be secondary to hyperglucagonemia. Since glucagon levels are elevated by fasting, the predominant role might account for the inability of insulin to stimulate biosynthesis of fatty acids in tissues of fasted animals (53).

A significant interrelationship exists between the metabolism of thyroid hormone and the conversion of carbohydrate to lipid. The hyper- or hypothyroid states are associated with rates of fatty acid synthesis that are enhanced or reduced, respectively (54,55). The thyroid hor-
mones dramatically stimulate synthesis of fatty acids in cultured hepatocytes (56). In the rat, T₃ and a fat-free, carbohydrate-rich diet interact synergistically to enhance flux through the pathway (57). Lipogenesis is also enhanced by T₃ in the diabetic animal, albeit on a delayed basis relative to normal animals (55). Although the ultimate source of carbon for synthesis is unknown in this case (58), the data underscore the stimulatory effects of T₃ upon the pathway.

Despite these documented direct effects of T₃ upon flux through the pathway and evidence that serum T₃ fluctuates in parallel with serum glucose, other evidence suggests that T₃ does not play a major role in the nutritional regulation of the activities of the lipogenic enzymes (25 and ref.23,25,34, and 40 therein). Thyroid hormone in vivo increases the total activities of enzymes constituting diverse pathways. In addition to lipogenesis, hyperthyroidism accelerates glucose turnover, glycolytic flux, and gluconeogenesis from both pyruvate and glycerol (59). The role of thyroid hormone may thus be permissive, required for the synthesis of various enzymes. Such a capability would be consistent with the hypothesis (54) that T₃ effects thermogenesis by stimulating futile biochemical cycles. In any event, T₃ clearly collaborates with insulin and glucagon in regulating the activity of hepatic malic enzyme (Section
To this point, the regulation of fatty acid synthesis by both diet and hormones has been described in terms of flux through the pathway. Such regulation involves both acute and relatively slower changes in flux. The former result from alterations in the catalytic efficiency of rate-limiting enzymes via covalent modification or allosterism, exemplified by the regulation of acetyl-CoA carboxylase by glucagon via cAMP (60) or citrate (61), respectively. The latter result from coordinate changes in total activities of the lipogenic enzymes responding to prolonged stimuli, due mainly to altered rates of enzyme synthesis rather than degradation (62). These are not insignificant adjustments secondary to mechanisms of acute control. The total activities of the lipogenic enzymes before the induction prompted by feeding are insufficient for the incorporation of precursors at the rates observed after induction (53). The balance of this introduction is limited primarily to malic enzyme, the rate of synthesis of which is regulated by both diet in vivo and hormones in hepatocytes in culture.

F. The Role of Malic Enzyme in Hepatic Lipogenesis; Nutritional and Hormonal Regulation of Enzyme Activity

The relative rate of synthesis of hepatic malic enzyme correlates positively with the rate of de novo fatty acid biosynthesis in birds and mammals subjected to various nu-
tritional and hormonal conditions. Synthesis of the enzyme is stimulated by both consumption of high carbohydrate, low fat diets and the administration of thyroid hormone. Their synergistic effects do not result from either a decrease in the clearance rate of T₃ or an increase in the nuclear binding of T₃ (57). Synthesis of the enzyme is inhibited by both fasting and thyroid hormone deprivation. Since starvation inhibits the conversion of T₄ to T₃, the possibility exists that the decrease in induction may involve the reversal of the mechanisms required for induction. In avian hepatocytes cultured in chemically defined media, synthesis of the enzyme is stimulated by insulin and thyroid hormone and inhibited by glucagon or cAMP.

The relative rate of synthesis of malic enzyme reflects the abundance of the cognate mRNA in total hepatic RNA. Both Towle et al. (63) and Siddiqui et al. (64) measured the amount of malic enzyme mRNA competent for translation in vitro using a mRNA-dependent rabbit reticulocyte lysate. In both the mammalian (63) and avian (64) systems, the abundance of hepatic malic enzyme mRNA was assessed as a function of nutritional state in vivo. In addition, the abundance was assessed in primary hepatocytes isolated from 17- to 19-day old neonatal chick embryos as a function of treatment with T₃ or T₃ plus glucagon. In both species under all conditions, the activity of the malic enzyme mRNA
template changed in parallel with the relative synthesis of the enzyme. The template activity of avian albumin mRNA was either unaffected or changed in a direction opposite to that of malic enzyme mRNA. Significantly, treatment of the rat resulted in insignificant changes in the (a) mass of total, cytoplasmic poly(A)^+ RNA and (b) proportion of malic enzyme mRNA bound to polyribosomes. In sum, these results suggested that rates of synthesis of malic enzyme, under the different conditions described above, are regulated by varying the abundance of malic enzyme mRNA. Since evaluating the template activity of a mRNA in vitro is indirect, confirmation of these results awaited the application of molecular hybridization techniques to quantitate directly the levels of malic enzyme mRNA.

G. Molecular Cloning of DNA Sequences Complementary to Malic Enzyme mRNA

Winberry et al. (65) cloned a cDNA corresponding to the malic enzyme mRNA of goose uropygial gland. A cDNA library was constructed in pBR322, using a heterogeneous population of mRNA including malic enzyme mRNA. The library was screened using a "plus-minus" strategy. Duplicate filters representing the library were hybridized independently with ^32P-labeled cDNAs synthesized from hepatic poly(A)^+ RNA isolated from neonatal goslings which had been previously fasted or fed, respectively. Recombinant plasmid DNA from one putative positive, pGME1, contained an insert of 970 bp
and selected an mRNA which directed the synthesis of malic enzyme in a cell-free system. Northern analysis using a single-stranded derivative of the sequences revealed that hepatic malic enzyme mRNA in the chicken is ~2100 nt in length (Fig.1). The availability of this clone facilitated the (a) isolation and characterization of additional cDNA clones (Section III A.), hence the determination of the primary structure of malic enzyme mRNA (66), (b) initial isolation and characterization of an exon-containing clone comprising a fraction of the malic enzyme gene (Section III C.), and (c) detailed analyses of pretranslational mechanisms potentially involved in the regulation of accumulation of malic enzyme mRNA.

H. Regulation of the Abundance of Malic Enzyme mRNA by Nutritional Status in vivo and by Hormones in Cultured Primary Hepatocytes

In duck liver, the synthesis and degradation of malic enzyme mRNA are regulated by nutritional status (67). When previously fasted ducklings were fed a high-carbohydrate diet, the amount of malic enzyme mRNA increased rapidly and reached an apparent steady state (about twenty times the initial level) at 24 hours. The half-life of malic enzyme mRNA in the livers of fed ducklings is 3-5 hours. When fed ducklings were starved, the amount of malic enzyme mRNA decreased rapidly with a half-life of approximately 1 hour. Apparently, realimentation inhibits the degradation of malic enzyme mRNA. The magnitude of the apparent inhibition
Fig. 1. Determination of the size of malic enzyme mRNA by Northern analysis. Polyadenylated RNA samples were subjected to electrophoresis under denaturing conditions. The RNA was blot-transferred to nitrocellulose and hybridized with $^{32}$P-labeled, single-stranded cDNA probe representing the insert of pGME1. The size scale was determined by electrophoresis of total duck reticulocyte poly(A)$^+$ RNA through the same gel. Migrations of 27 and 18 S ribosomal RNAs and globin mRNA were linear with respect to the logarithms of their molecular weights. Lanes A and B each contained 20 μg of RNA isolated from the livers of 2-days-old chicks which were either fasted (A) or fed (B) for 24 hours after hatching. OR, origin; DF, dye front.
of degradation is not sufficient to account for the twenty-fold increase in the level of the mRNA caused by refeeding.

The rate of transcription of the cognate gene was measured using nuclei isolated from duck livers and cloned malic enzyme cDNAs as probes. When fasted ducklings were fed, the rate of transcription increased 3- to 5-fold. When fed ducklings were fasted, transcription was inhibited 55-65%. The synthesis of albumin mRNA was affected little under these conditions, hence the effects upon the synthesis of malic enzyme mRNA were selective. The 3- to 5-fold increase in rate of transcription plus a similar increase in stability of the mRNA could account for the 20-fold increase in the level of malic enzyme mRNA caused by feeding the fasted animal. A major role for transcription is supported by analyses of hepatic nuclear RNA. The levels of putative nuclear precursors of malic enzyme mRNA correlate positively with changes in the rate of transcription of the gene. This suggests that the processing of nuclear precursors is not regulated by nutritional status. The possibility remains, however, that decreases in the degradation of nuclear malic enzyme RNA species contribute to increases in production of cytoplasmic malic enzyme mRNA.

These transcription experiments have been repeated using restriction fragments of malic enzyme genomic clones, the isolation of which is described in Chapter III. In both ducklings and chicks, refeeding a starved animal caused a
≥ 60-fold increase in transcription (68). Thus, nutritional status regulates the abundance of malic enzyme mRNA primarily by changing the rate of initiation of transcription. These data contrast quantitatively with those obtained with cDNA probes. The latter contain sequences that are not specific to malic enzyme mRNA (e.g., homopolymeric tails) that might contribute to artifactual results.

Mechanisms involved in the multihormonal regulation of synthesis of hepatic malic enzyme have been investigated (65,69). The stimulatory effects of insulin and triiodothyronine and the inhibitory effect of glucagon have been evaluated by assessing both the abundance and rate of synthesis of malic enzyme mRNA in primary cultures of hepatocytes isolated from embryonic chicks. Insulin and T₃ in combination increase the abundance of malic enzyme mRNA as much as 100-fold (65). Glucagon causes a 93% decrease in the accumulation resulting from treatment with insulin plus T₃. These effects are similar to those exerted by the hormones upon the synthesis of malic enzyme (69), suggesting that regulation involves pretranslational process(es).

Regulation of accumulation of malic enzyme mRNA by T₃ involves changes in transcription. The rate of initiation of transcription, assessed using genomic fragments as probes, increases ≥ 30-fold following treatment with T₃ (70). This fully accounts for the accumulation of malic enzyme mRNA assayed in parallel. Treatment with T₃ results
in relatively slow accumulation of the mRNA which appears to require continuous synthesis of protein (69,71). Both puromycin and pactamycin inhibit the response to T₃. The lack of effect of puromycin upon the half-life of β-tubulin mRNA provides evidence for its specificity in inhibiting the accumulation of malic enzyme mRNA (69). These data support a model in which T₃ enhances the synthesis of a polypeptide intermediate which, in turn, stimulates transcription of the malic enzyme gene. Alternatively, a protein with a short half-life may be required for the transcriptional response to T₃, although the concentration of the protein may not be regulated by T₃.

Glucagon may function at two levels in cultured avian hepatocytes to regulate the abundance of malic enzyme mRNA. Glucagon decreases the T₃-induced rate of transcription by ≥ 90% (70). Glucagon also appears to alter the stability of the mRNA (69). The half-life for malic enzyme mRNA in cells treated with T₃ and glucagon was no greater than 1.5 ± 0.3 hours. The half-life of malic enzyme mRNA in cells treated with T₃ alone, estimated by adding either α-amanitin or actinomycin D to cells previously incubated with insulin and T₃, was 8 or 11 hours, respectively. Thus, glucagon apparently stimulates the degradation of malic enzyme mRNA by at least five-fold.

The foregoing data are inconsistent with the indepen-
dent findings of Nikodem et al. (72, 73). In euthyroid rats fed a fat-free, high carbohydrate diet for ten days, the abundance of hepatic malic enzyme mRNA increased 7- to 8-fold compared to that in control animals consuming standard chow. The increase did not result from increases in either the rate of transcription of the malic enzyme gene or the abundance of polyadenylated nuclear sequences encoding malic enzyme, suggesting that consumption of the high carbohydrate diet decreases the rate of degradation of cytoplasmic malic enzyme mRNA. The amplitude of the response is a function of the thyroid status of the animal, which alters the abundance of malic enzyme mRNA by nuclear processes.

In rat liver, T₃ acts at multiple levels, via mechanisms distinct from those mediating the effects of diet. Following treatment of euthyroid rats with T₃ for ten days, the rate of transcription, assessed using homologous cDNA probes, increased 3- to 4-fold (74). Since the abundance of the cytoplasmic mRNA increased 10- to 15-fold under identical conditions (75), the operation of one or more additional pretranslational mechanisms was invoked to account for the discrepancy. To address the role of nuclear processes, both cDNA and intron probes specific for the malic enzyme gene were employed to quantitate the time courses of the effects of T₃ on both the rate of transcription of the gene and the accumulation of malic enzyme RNA in the nucle-
us. The abundance of both nuclear and cytoplasmic malic enzyme RNA sequences increased approximately 10- to 12-fold, exhibiting a half-time of accumulation of ~60 hours. Since the increase in transcription rate occurred with a half-time of ~18 hours and results were identical regardless of the type of probe used, it appears that T₃ both activates the initiation of transcription of the malic enzyme gene and decreases the rates of degradation of nuclear precursors of malic enzyme mRNA. These data contrast with those obtained from avian systems in which diet and hormones regulate accumulation of malic enzyme mRNA primarily, if not exclusively, by changing the rate of initiation of transcription.

I. Statement of the Problem

Our objective is to define the molecular bases for the regulation of accumulation of avian hepatic malic enzyme mRNA by diet, triiodothyronine, and glucagon. The effects of both perturbation of diet in vivo and hormone concentrations in vitro are pretranslational, exerted primarily at the level of initiation of transcription. Additional processes which might be regulated include elongation or termination of transcription, covalent modification or splicing of nuclear precursors, vectorial transport of mRNA from nucleus to cytoplasm, and turnover of functional mRNA. Consequently, a major prerequisite for identifying the sequences
critical for regulation is to isolate and analyze genomic sequences that constitute the malic enzyme locus. Accordingly, a family of cDNA clones and a series of genomic clones have been isolated and partially characterized.
EXPERIMENTAL PROCEDURES

A. **Materials**

Chemicals were of reagent grade or the highest quality available commercially. Agar and agarose (Seakem LE, Seaplaque, and GTG) were purchased from Difco and FMC Corporation, respectively. Nucleotides were purchased from Sigma, Pharmacia, and BRL. Radiolabelled nucleotides were obtained from ICN and NEN. Phenol (Fisher Scientific) was redistilled, buffered with Tris·HCl (pH 8.0), and stored under 100 mM Tris·HCl (pH 8.0) -20°C. Formamide (MCB) was deionized by stirring in the dark with AG501-X8 resin (BioRad) for several hours. The mixture was then filtered by suction, and the filtrate was stored at -20°C. Oligo(dT)-cellulose was obtained from Collaborative Research. Nitrocellulose membranes were purchased from Schleicher and Schuell.

Enzymes were obtained from the sources indicated: DNAse I (Cooper-Worthington), RNAse A (Sigma), Moloney murine leukemia reverse transcriptase (BRL), T4 DNA ligase, calf intestinal phosphatase, T4 polynucleotide kinase, Klenow fragment of *E. coli* DNA polymerase I, and S1 nuclease (Boehringer Mannheim), T7 polymerase (Sequenase™, United States Biochemical Corp.), DNA polymerase from *Thermus aquaticus* (Promega Biochemicals). All restriction enzymes (IBI, BRL, New England Biolabs, and Boehringer Mannheim) were used according to the manufacturers' specifications.
Vectors and oligodeoxynucleotides were obtained from commercial and private sources. Plasmids were obtained from BRL (pBR322 and pUC19), P-L Biochemicals (replicative forms of M13mp18 and 19), Promega Biochemicals (pGEM3), and IBI (pIBI31). Primers complementary to the 3' side of the multiple cloning site of the M13 vectors were purchased from P-L Biochemicals and US Biochemical Corporation. Oligodeoxynucleotides complementary to malic enzyme sequences were synthesized in the Core Facility of Case Western Reserve University or The University of Iowa using Applied Biosystems 380A DNA Synthesizers. Two oligodeoxynucleotides were used as internal primers in the sequence analysis of the exon-containing, 0.6 kb PstI fragment of λCME245: MEOL 245gm (5'-CCTCCCCACTCTCTC-3') is complementary to the 'c' strand; MEOL 245gc (5'-GAGAGAGAGTGGGGAGG-3') is complementary to the 'm' strand. Two oligodeoxynucleotides were used in the primer-extension analysis of poly(A)⁺ RNA: MEOL 2.5 (5'-AGGTGCGGTGGACGACTCGTAGGCCCTCTTCAT-3') was used to prime RNA isolated from hepatocytes. MEOL 100 (5'-CCTGCTTG-CATGGCGACGACT-3') was used to prime RNA isolated from chicken liver.

Three cDNA libraries were employed in these studies. Two were prepared by Dr. S.M. Morris of our laboratory using poly(A)⁺ RNA isolated from the livers of either white Pekin ducklings or growing White leghorn chicks. One was
prepared by Dr. S. Kawamoto of our laboratory using poly(A)^+ isolated from embryonic chick hepatocytes that had been incubated with insulin and triiodothyronine to induce maximally the accumulation of malic enzyme mRNA. Priming of total poly(A)^+ RNA, synthesis of cDNAs, their insertion into the Pst I site of pBR322, and transformation of E. coli HB101 were completed as described by Morris et al. (76) with the following exception. Prior to insertion, the 5' ends of the hepatocyte cDNAs were modified twice in succession by the addition of homopolymeric tails (77).

Two genomic libraries were employed in these studies. One was kindly provided by Dr. J. Dodgson, then of the California Institute of Technology (78). It was constructed by partially digesting total genomic DNA from chicken liver with the restriction enzymes HaeIII and AluI. The reaction products were ligated into the λ derivative Charon 4A using EcoRI linkers and the recombinants packaged in vitro (79). The second library was generously provided by Dr. H.-J. Kung. It was constructed by partially digesting total genomic DNA from chicken erythrocytes with the restriction enzyme EcoRI. Reaction products were ligated to the arms of the λ derivative EMBL4 and the recombinants packaged in vitro (79).

Preparations of nucleic acids were generously donated by two individuals. Dr. S. Stapleton (Department of Bio-
chemistry, University of Iowa) provided total RNA isolated from chicken liver. Dr. H.-J. Kung (Department of Molecular Biology and Microbiology, Case Western Reserve University) provided genomic DNA isolated from erythrocytes of three chickens obtained from the inbred lines "72", "15", and "P", respectively.

A positive control for primer extension and nuclease protection assays was generated in the following manner. The insert of pCME5 was subcloned into the PstI site of pGEM3 by Dr. L. Salati of our laboratory. The resulting construct was then linearized with BamHI and non-radioactive "sense" strand malic enzyme RNA was synthesized using T7 polymerase.

B. Methods

1. Identification of Malic Enzyme cDNA Clones. Isolation and partial characterization of pGME1, a 970 bp cDNA clone derived from the 3' end of malic enzyme mRNA from the goose uropygial gland, has been described (65). Hybridization probes representing the insert of pGME1 were synthesized (Section B.9.) from the corresponding subclone in M13mp8 and used to screen the duck liver and chicken liver cDNA libraries described above. Colony hybridizations were performed by a modification (80) of the procedure of Grunstein and Hogness (81). This resulted in the identification of pDME1, pCME1, and pCME2. Subsequently, this paradigm was repeated using single-stranded probe representing the
insert of pDME1 to screen the cDNA library prepared from RNA isolated from chick-embryo hepatocytes. Thus, pCME3, pCME4, and pCME5 were identified.

2. Identification of Malic Enzyme Genomic Clones. The two genomic libraries were screened with the probes described in the text by the plaque hybridization procedure of Benton and Davis (82). The number of plaques required for screening was calculated as described by Clarke and Carbon (83) using $3 \times 10^6$ kb as the size of the chicken genome:

$$N = \frac{\ln (1-P)}{\ln (1-f)}$$

In this relationship, $P$ is the desired probability of detecting a unique sequence, $f$ is the fractional length of the genome contained within the average recombinant, and $N$ is the number of recombinants to be screened. 'P' was typically assigned a minimal value of 0.95. The average length of an insert was assumed to be 15 kb. Libraries were plated at a density of 30,000 plaques per 150 mm dish. Plaques corresponding to the most intense hybridization signals were selected for purification and characterization. Hybridization analysis of each putative positive was repeated until no less than 99% of the plaques yielded positive signals. Details of the identifications of the genomic clones are presented in Chapter III.
3. Characterization of Cloned DNAs Derived from the Malic Enzyme Locus. Structural features of the various clones were determined by a combination of restriction mapping (84), Southern analyses (Section B.4.) and sequence analyses (Section B.6.). To facilitate manipulating large quantities of specific fragments, the desired DNA fragment(s) were subcloned into the appropriately digested vector using standard conditions (84). The ligated recombinant molecules were used to transform competent E. coli derivatives HB101, JM101, or JM103 (85). Transformants were identified by their (a) resistance to tetracycline and sensitivity to ampicillin (pBR322) or (b) colorless phenotype when plated in the presence of isopropyl-β-D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (M13 derivatives). Complementarity of the inserts in single-stranded recombinants was determined by hybridization followed by electrophoresis through neutral agarose gels (86).

4. Isolation and Southern Analysis of DNA. Plasmid DNA was isolated from positive colonies by a modification (87) of the method of Birnboim and Doly (88). Single-stranded DNA was extracted from polyethylene glycol precipitates of M13 bacteriophage (87). Double-stranded replicative form of bacteriophage DNA was isolated from cells (87,88). Bacteriophage DNA was isolated from plate lysates (84). Total genomic DNA was extracted from the erythrocytes and livers of chickens, ducks, and geese (89). For the copy number
experiment, nuclei were isolated from erythrocytes and liver prior to extraction of DNA (90). Plasmid or genomic DNAs were separated by size by electrophoresis through neutral agarose buffered with 50 mM Tris, 10 mM sodium acetate, 5 mM EDTA (pH 7.0) using the concentration of agarose appropriate for the desired resolution. Following staining with ethidium bromide and photography, DNA was transferred to nitrocellulose (91). The membrane was then baked at 80°C for 2 hr in vacuo and hybridized with selected radio-labelled probes (Section B.9.). Typically, membranes were pre-hybridized for at least 2 hr at 43°C in a solution consisting of 50% formamide, 5x SSC, 5x Denhardt's solution, 0.1% SDS, and 200 μg/ml sonicated salmon sperm DNA. Subsequently, the solution was replaced with hybridization solution of the same composition, except that the 32P-labelled probe was at a concentration 0.5-1.0 x 10^6 cpm/ml. Following hybridization, some filters were washed initially at 43°C in 50% formamide, 5X SSC, and 0.1% SDS. This initial wash was omitted for other filters. All filters were eventually washed at 252°C in 0.1X SSC and 0.1% SDS.

5. Isolation and Northern Analysis of RNA. Hepatocytes or livers from decapitated chickens were homogenized in guanidine thiocyanate (92). Total RNA was isolated from the extracts by ethanol precipitation from guanidine hydrochloride according to the method of Chirgwin et al.(92).
Polyadenylated RNA was prepared by a modification of the method of Aviv and Leder (93). To test the specificity of a given probe for malic enzyme mRNA or the integrity of a preparation, Northern blots were prepared. RNA samples were denatured with 2.2 M formaldehyde at 65°C for 5 min and subjected to electrophoresis in 1.0 % denaturing agarose gels buffered with 20 mM boric acid and 0.2 mM EDTA, pH 8.3 (94) or 20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0 (95). Following staining with ethidium bromide and photography, the RNAs were transferred to nitrocellulose (96). The membrane was then hybridized with selected probes (Section B.9.) and processed as described for Southern blots (Section B.4.).

6. **Sequence analysis.** The 0.6 kb, exon-containing PstI fragment of λCME245 was subcloned into M13mp18. Its primary structure was determined by the dideoxy chain-termination method (97) using both Sequenase and Taq polymerase. Unambiguous data could not be obtained with Sequenase. Employing Taq polymerase at 70°C resolved all discrepancies. The sequences of other fragments described in the text were obtained in a similar fashion.

7. **Primer-extension analysis.** This protocol was used for extension of (a) MEOL 2.5 following hybridization to poly(A)⁺ RNA isolated from hepatocytes treated with insulin and T₃ and (b) MEOL 100 following hybridization to poly(A)⁺
RNA isolated from the liver of a chicken that had been
fasted for 48 hr, then fed for 24 hr. The indicated oligo-
deoxyribonucleotide was labelled at its 5' end (Section 9.b.)
and purified by electrophoresis through a 12% native polya-
crylamide gel (Section 9.a.). The primer (5 x 10^4 cpm) was
mixed with 20 µg of (a) poly(A)^+ RNA, (b) yeast tRNA, or
(c) RNA consisting of yeast tRNA and synthetic malic enzyme
RNA at a relative concentration of 1 part per hundred. Each
mixture was precipitated with ethanol, and the precipitate
was suspended in 20 µl of 2.5X reverse transcriptase buffer
(5X buffer = 250 mM Tris, 375 mM KCl, 50 mM DTT, 15 mM
MgCl_2, pH 8.3). Samples were heated at 65°C for 5 min and
allowed to cool to room temperature over a period of 4 hr.
Subsequently, dATP, dCTP, dGTP, and TTP were added to final
concentrations of 500 µM each along with 200 units of re-
verse transcriptase. Incubation proceeded at 37°C for 1 hr,
after which the nucleic acids were precipitated with
ethanol and analyzed as described at the end of Section 8.

8. S1 Nuclease Analysis (98). (a) HaeIII-HaeIII probe.
The exon-containing, PstI fragment of λCME245 was subcloned
into pIB131 by S. Klautky. The insert was isolated by prepar-
ative gel electrophoresis through a 6% native acrylamide
gel (Section 9.a.) and digested with HaeIII. Reaction prod-
ucts were treated with calf intestinal phosphatase, labelled
at their 5' ends with T₄ polynucleotide kinase (Section
9.b.), and separated by electrophoresis through 6% native acrylamide. The 264 bp HaeIII fragment was isolated and subjected to a second round of preparative electrophoresis. The probe (5x 10⁴ cpm) was combined with 20 µg of (a) poly(A)+ RNA isolated from chick hepatocytes, (b) yeast tRNA, or (c) RNA consisting of yeast tRNA and synthetic malic enzyme RNA at a relative concentration of 1 part per hundred. Following ethanol precipitation, the mixtures were resuspended in 20 µl of hybridization buffer (50% formamide, 0.4 M NaCl, 0.01 M Pipes, pH 6.5, 0.1% SDS, 1 mM EDTA), heated at 85°C for 3 min, and incubated at 42°C for 15 hr. Samples were cooled to room temperature and incubated with 200 µl of S1 nuclease buffer (0.05 M potassium acetate, 0.3 M NaCl, 1 mM ZnSO₄, 25 µg/ml denatured salmon sperm DNA, 1000 units of S1 nuclease (pH 4.5)) at 37°C for 1 hr.

(b) BstNI·HaeIII probe. The end-labelled HaeIII fragment described in Section 8.a. was isolated and then digested with BstNI. Radioactive species were separated by electrophoresis through a 6% native acrylamide gel from which the 215 bp product was isolated (Section 9.a.). The probe (5x 10⁴ cpm) was precipitated independently with each of the samples described in Section 8.a., as well as 20 µg of poly(A)+ RNA isolated from the liver of a chicken that had been fasted for 48 hr and then fed for 24 hr. Samples were processed as described in Section 8.a. with the following
two exceptions: the hybridization buffer contained 80% formamide and the hybridization was at 55°C. For procedures B.6., B.7., and B.8., radioactive species were resolved by size by electrophoresis at constant power (60W) through 7 M urea gels containing acrylamide at 6% (B.6., B.8.) or 8% (B.6, B.7.). The gels were then dried at 80°C in vacuo for at least 2 hr and subjected to autoradiography to permit visualization of bands. For primer-extension and S1 analyses, 32P end-labelled HaeIII fragments of φX174 and a sequencing ladder generated from M13mp18 were used as standards for length.


a. Preparative electrophoresis of DNA. For the preparation of specific labelled DNA molecules, two methods of isolation were employed. (1) Non-radioactive fragments of cDNA or genomic clones were separated by size by electrophoresis through neutral agarose. Following staining with ethidium bromide and visualization with long-wave UV light, the fragment was excised within a small agarose plug. The plug was dissolved in NaI and the fragment recovered using beads of glass (99). The entire procedure was repeated prior to radiolabelling by nick-translation or random priming (Section 9.b.). Such probes were used in Southern analyses of both cloned and total genomic DNA. (2) Restriction fragments and oligodeoxynucleotides were
separated by size by electrophoresis through native acryl-
amide gels of the appropriate percentage. Following local-
ization of the desired fragment by staining with ethidium 
bromide or autoradiography, the corresponding region of the 
gel was excised with a razor blade and crushed thoroughly 
in ~400 ul of 500 mM ammonium acetate, 1 mM EDTA (pH 5.0). 
The suspension was then incubated at 37° C with constant 
shaking (250 rpm) for ≥ 12 hr. The liquid was recovered by 
centrifugation and the nucleic acid recovered by precipita-
tion with ethanol in a bath composed of dry ice and ethan-
ol. The latter method was used to purify fragments that 
were end-labelled (Section 9.B.) for primer-extension and 
S1 nuclease analyses.

b. Radiolabelling of DNA. Unless indicated otherwise, 
probes were prepared with [α-32P]dCTP (800 Ci/mmol). Single-
stranded, uniformly-labelled primer extension products were 
synthesized according to Hu and Messing (100). Penta- or 
heptadecameric sequencing primers were annealed to the 
selected M13 recombinant. Products were resolved from the 
template by electrophoresis through 1% Seaplaque agarose 
under alkaline conditions (101). The probe was localized in 
situ by autoradiography, excised, and solubilized by heat-
ing in a water bath. Double-stranded substrates were either 
nick-translated (84) or primed randomly (102), separated 
from unincorporated nucleotides by chromatography through
Sephadex G-50, and used directly. Double-stranded and single-stranded molecules were 5'-end-labelled with [γ-32P]ATP (3000 Ci/mmol) and T4 polynucleotide kinase (84). Unincorporated nucleotides were removed by three successive precipitations with sodium acetate and ethanol at -20°C. The integrity of the product was evaluated by autoradiography following electrophoresis through native polyacrylamide. If necessary, a radioactive specie was recovered from the acrylamide as described in Section 9.a. Radioactivity was quantitated by Cerenkov radiation or liquid scintillation spectroscopy.
CHAPTER III

RESULTS

A. Characterization of Cloned DNA Sequences Complementary to Malic Enzyme mRNA

Cloned DNAs complementary to malic enzyme mRNA from both avian liver and uropygial gland were isolated and partially characterized (Fig. 2). Following the identification of pGME1 by hybridization-selected translation (section I.G. and (65)), a single-stranded probe representing its insert (ssGME1) was used by S. Morris of our laboratory to screen each of two cDNA libraries he constructed from polyadenylated RNA isolated from the liver of either a single duck or single chicken. One plasmid, pDME1, was isolated from the former and two plasmids, pCME1 and pCME2, were isolated from the latter. The parts of the malic enzyme cDNA not shared by pDME1 and pCME1 were initially assumed to be 5' and 3' extensions, respectively, of the other three clones because (a) the libraries were prepared by priming RNA with oligo(dT), thus biasing their contents toward 3' transcripts, (b) sites for PvuII, XbaI, and RsaI are at similar locations in pDME1 and pGME1 DNAs, and (c) sites for RsaI are at similar locations in pGME1, pCME1, and pCME2 whereas that for SstI is absent from both pGME1 and pCME2. Confirmation of these suppositions required further analyses.
Fig. 2. Restriction maps of cloned DNAs complementary to avian malic enzyme mRNA from duck liver (D), chicken liver (C), and goose uropygial gland (G). The cross-hatched regions represent flanking regions of pBR322. A, Avai; E, EcoRI; H, HindIII; Ps, PstI; Pvu, PvuII; R, RsaI; Sp, Sphi; Ss, SstI; X, XbaI. Note: in pCME1, the PstI site nearer the Avai site is defective.
The orientation of pCME1 was determined by hybridization analysis. In pCME1, the PstI site nearer the AvaI site is defective. Thus, its insert was subcloned into M13mp8 as part of a chimaeric PstI·PvuI fragment which included flanking sequence from pBR322. A recombinant of the 'm' strand was identified by its ability to hybridize to a 'c' strand recombinant of the insert of pGME1. Truncated, single-stranded probe was synthesized using the 'm' strand template. Its hybridization to pCME1 was limited to the region flanked by the defective PstI site and the sites for AvaI and RsaI. Thus, this region is the 3' end of the segment of malic enzyme mRNA represented by pCME1. The truncated probe, ssCME1_tr, was used in a subsequent experiment described below.

The orientation of pDME1 was determined by restriction digestion. The PstI fragment of pDME1 was cloned into the PstI site of M13mp8. Thus, in single-stranded DNAs extruded from recombinants, malic enzyme sequences reside 3' of the EcoRI site in the vector. Depending on the identity of the strand borne by the recombinant, the EcoRI site derived from pDME1 would map either ~700 or ~1300 nt from the EcoRI site in the vector. A recombinant bearing the 'c' strand was identified by virtue of hybridizing to an 'm' strand M13mp8 subclone of the insert of pGME1. Replicative form was prepared from the 'c' strand and was digested with EcoRI. Release of a ~1300 bp fragment unequivocally demon-
strated that the 3' end of malic enzyme cDNA, hence the 5' end of malic enzyme mRNA, is represented in the ~700 bp PstI·EcoRI fragment of pDME1. The latter fragment, named dsDME1_PG5', was also used in subsequent experiments described below.

Screening of another cDNA library resulted in isolation of a full length cloned DNA complementary to chicken hepatic malic enzyme mRNA. A cDNA library was prepared by Dr. S. Kawamoto of our laboratory using polyadenylated RNA isolated from chick-embryo hepatocytes treated with insulin and T3. Screening and preliminary Southern analysis with single-stranded probe representing the insert of pDME1 (ssDME1) revealed three clones: pCME3, pCME4, and pCME5. ssGME1 hybridized with the inserts of pCME4 and pCME5, but not with pCME3, suggesting that pCME3 represented sequences near the 5' end of malic enzyme mRNA and that pCME5 represented nearly the entire length of malic enzyme mRNA. These relationships were subsequently verified by applying a combination of restriction mapping and Southern analyses (data not shown). Malic enzyme mRNA is ~2100 nt in length; the insert of pCME5 is about 1950 bp long and does not hybridize to end-labeled poly(U). Thus, pCME5 DNA is an optimal substrate and reagent for analyzing the structures of chicken malic enzyme mRNA and its cognate gene.

The extent of the identities between the various cloned cDNAs was established by determining their primary
structures. That analysis is the subject of another thesis (66), hence, with one exception, it will not be considered here. A comparison of the nucleotide sequences of malic enzyme cDNAs from goose, duck, chicken, rat (103), and mouse (104) revealed a region of identity in the coding region mapping near the 3' ends of the mRNAs (Fig.3). The region consists of 9 consecutive nucleotides which are identical in all five animals, followed by 9 bases which are identical among the birds and almost identical among the mammals but substantially different between the birds and mammals, followed by 29 consecutive nucleotides which are identical in all five animals with the exception of a T to C transition in chickens.

The significance of this identity was assessed by M. Glynias of our laboratory. The encoded amino acids are not unusual with respect to either degeneracy at the third base position or with respect to codon preference in either hepatic proteins or in the balance of malic enzyme. Lack of drift in the third nucleotide, despite more than 200 million years of evolution since the divergence of mammals and birds, suggests that this sequence is important to the structure, function, or regulation of malic enzyme mRNA or DNA. For example, it may contribute to the regulation of stability of malic enzyme mRNA by both diet and hormones.
Fig. 3. Identities in the nucleotide sequences for malic enzyme cDNAs of rat, mouse, chicken, duck, and goose. The sequences for rat (103), mouse (104), chicken, duck, and goose (66) were obtained as indicated. The sequences are aligned to nucleotides 1140 to 1199 of the chicken malic enzyme cDNA.
RAT
TCAGAAGATAAAACCAACCGCTCTCTAGGAGTTGCTGCAATTGGTGTGCTTTTACAGA;
MOUSE
TCAAAAAGATAAAAAACCAACTGCCCTCTAGGAGTTGCTGCAATGGTGTGCTTTTACAGA;
DUCK
TAAAGATATAAAAAACCATCTGTTATTAGGAGTTGCTGCAATTGGTGTGCTTTTACTAA;
CHICKEN
TAAAGATATAAAAAACCATCTGTTATTAGGAGTTGCTGCAATTGGTGTGCTTTTACTAA;
GOOSE
TAAAGATATAAAAAACCATCTGTTATTAGGAGTTGCTGCAATTGGTGTGCTTTTACTAA;
B. Assessment of the Copy Number of the Gene Encoding Malic Enzyme

The number of copies of the malic enzyme gene per haploid genome was determined with the assistance of J. Fisch of our laboratory. The hybridization signal for malic enzyme sequences in genomic DNA digested with EcoRI was compared with that from increasing amounts of pGME1 DNA, also digested with EcoRI (Fig.4). Fragments of about the same size were detected in digestions of both cloned and genomic DNA. Thus, the size-dependent efficiency of transfer during blotting was assumed not to be a factor in the analysis. Quantitative comparison of the hybridization signals from the genomic fragments with that from the increasing quantities of plasmid DNA revealed a copy number of ~0.25, and was considered to be 1 for practical purposes. (This discrepancy was assumed due to the instability of hybrids formed between the probe and chicken genomic DNA relative to those formed between the probe and homologous goose cDNA.) If the copy number does exceed one, the additional copies exhibit little or no sequence similarity within the exons encoding the sequence detected in the plasmid digests.

C. Isolation and Characterization of Cloned Genomic DNA from the Malic Enzyme Locus

The availability of malic enzyme cDNA clones facilitated the isolation and characterization of cloned genomic
Fig. 4. **Determination of the copy number of the malic enzyme gene in the haploid genome of the goose.** Plasmid or hepatic, genomic DNA samples were digested with EcoRI, subjected to electrophoresis through neutral agarose, transferred to nitrocellulose, and hybridized with a single-stranded, $^{32}$P-labeled probe derived from pGME1. *Lanes A-F* contain the amounts of pGME1 representing the number of genomic equivalents indicated in the figure. *Lanes G and H* each contain 5 µg of genomic DNA. *Lanes I and J* each contain 10 µg of genomic DNA. In addition, sufficient sonicated salmon sperm DNA was added to each sample to adjust the amount of DNA to 10 µg.
sequences derived from the malic enzyme locus. Two genomic libraries were screened with various probes. I isolated ten unique clones encoding portions of malic enzyme mRNA which, collectively, span more than 100 kb of DNA. Due to the large size of this gene, the ensuing narrative is historical in nature and emphasizes the reasoning and strategies used to characterize this transcription unit.

Preliminary Southern analysis of total genomic DNA permitted an estimate of the minimal size of the malic enzyme gene. Two single-stranded probes were used to probe DNA isolated from the liver of either an individual duck or chicken. The first probe, ssDME1, represented ~95% of duck hepatic malic enzyme mRNA. It hybridized to at least seven EcoRI fragments ranging in size from ~1.5 kb to >10 kb (Fig.5A). Some signals exhibited greater intensities than others, possibly reflecting heterogeneity in the length of the probe, binding of more than one molecule of probe per fragment due to the presence of multiple exons, or binding of probe to different fragments of the same length. Assuming that (a) the probe was specific for the malic enzyme locus, (b) all exons were represented by this probe, (c) the probe formed a stable hybrid with at least one exon per restriction fragment, and (d) no restriction fragment escaped detection because it lacked exon(s), the minimum length of the gene was estimated to be 18-38 kb.

The second probe, ssCME2, represented the 3' one-third
Fig. 5. Southern analyses of genomic DNA isolated from both duck liver (A) and chicken liver (B). Genomic DNA was isolated from hepatic nuclei and digested with the restriction enzymes indicated. Following electrophoresis through neutral agarose (10 μg per lane), the DNA was transferred to nitrocellulose and hybridized to single-stranded 32P-labeled probes derived from pDME1 (A) or pCMET2 (B). OR, origin; DF, dye front.
kilobase pairs
of chicken hepatic malic enzyme mRNA. ssCME2 hybridized to two genomic EcoRI fragments, of 5.9 kb and 7.6 kb (Fig.5B). pCME2 lacks an internal EcoRI site, indicating that these two fragments are separated by at least one EcoRI site in an intron. This prediction was confirmed by the isolation of a genomic clone that encodes the 3' end of malic enzyme mRNA.

Initial attempts to isolate the gene involved screening a library of chicken hepatic DNA kindly provided by Dr. J. Dodgson (78). Recombinants had been prepared in the bacteriophage lambda replacement vector, Charon 4A, in the following manner. Total hepatic, genomic DNA was partially digested with HaeIII and AluI. Fragments ranging in size from 14-22 kb were isolated. Synthetic EcoRI linkers were ligated to the fragments and these EcoRI fragments ligated into the purified "arms" of Charon 4A. The recombinant molecules were packaged into viable phage particles in vitro and subsequently amplified to establish a permanent library, hereafter referred to as Library I.

Five genome equivalents of Library I (325,000 recombinants) were screened with ssGME1, the only cloned malic enzyme cDNA then available. One bona fide genomic clone, λCME1, was obtained. A detailed restriction map (Fig.6) was constructed in collaboration with B. Wilson of our laboratory. Sequences in the 3'-most one-third of malic enzyme cDNA hybridized exclusively to two EcoRI fragments of
Fig. 6. Partial restriction map of λCME1, a recombinant bacteriophage that encodes the 3' end of chicken malic enzyme mRNA. λL, left arm of Charon 4A; λR, right arm of Charon 4A; A, AvaI; B, BamHI; E, EcoRI; H, HindIII; K, KpnI; P, PstI; P, PvuII; Sa, SacI; Sm, SmaI; X, XbaI; solid bars, flanking regions of cloning vectors; cross-hatched bars, approximate locations of exons in the 7.6 kb EcoRI fragment. The 7.6 and 2.6 kb EcoRI fragments were subcloned into pUC vectors for detailed mapping. Exons within the 2.6 kb EcoRI fragment do not encode sequences 5' of the XbaI site in pCME4 and pCME5.
λCME1. Probes representing the 5' end of malic enzyme mRNA, for example dsDME1pE5', do not hybridize to λCME1. The exon-containing fragments of λCME1 did not hybridize to sequences 5' of the XbaI site within the insert of pCME5. These data indicated that λCME1 contains ~15.6 kb of genomic DNA and encodes the 3' region, and possibly the 3' end, of malic enzyme mRNA.

The exon-containing EcoRI fragments within λCME1 are 2.6 kb and 7.6 kb in length, consistent with the Southern analyses of total genomic DNA using probes derived from pCME2. At this point in time, the cloned 2.6 kb EcoRI fragment seemed likely to be derived from the 5.9 kb EcoRI fragment detected in total genomic DNA because the 2.6 kb EcoRI fragment abuts the left arm of the cloning vector. Digestion of the genomic DNA with HaeIII and AluI prior to cloning produced subsets of the 5.9 kb genomic EcoRI fragment. This hypothesis was substantiated by the isolation of an overlapping genomic clone (λ15A, described below). The 3'-most cDNA probe, ssCME1tr, hybridized only to the 7.6 kb EcoRI fragment, establishing that the 2.6 kb fragment is 5' of the 7.6 kb fragment. T. Winter of our laboratory determined the nucleotide sequence of a 617 nt XbaI-HindIII fragment derived from the 7.6 kb EcoRI fragment. It contains a polyadenylation signal (AATAAAA) beginning at
nucleotide 354 and corresponds exactly to the nucleotide sequence of the 3' end of duck malic enzyme cDNA. This fragment likely encodes the 3' end of malic enzyme mRNA. The absence of additional, related tracts suggests that mature malic enzyme mRNA exhibits little, if any, heterogeneity at its 3'end. Confirmation of these hypotheses awaits the refined structural analysis of the 3' end of the RNA.

By this time pCME5 had been isolated. The insert of pCME5 appeared to represent virtually all of the exons of the malic enzyme gene. It was used to screen Library I for genomic clones containing the balance of the gene. Sequences comprising the entire insert of pCME5 and its 5' PstI· HindIII fragment (CME5PH5*) were used to screen Library I. I had anticipated that the 5' cDNA fragment would identify genomic clones containing the promoter and putative regulatory regions of the gene.

To facilitate analysis of the 5' end of the gene, additional Southern analyses of the malic enzyme locus in total hepatic genomic DNA were completed (Fig.7). ssCME5PH5 detected two EcoRI fragments (~15 kb, 7.5 kb) and two PstI fragments (1.5 kb, ~0.60 kb). Only one size class (2.8 kb) of fragment was detected in genomic DNA digested with either MspI alone or both MspI plus EcoRI.

Four explanations were considered to account for the failure of the EcoRI/MspI double digest to produce two
Fig. 7. Southern analysis of total genomic DNA from chicken liver using an homologous 5'cDNA probe. Genomic DNA was isolated from hepatic nuclei and digested with the following restriction enzymes: Lane A, MspI; Lane B, EcoRI; Lane C, MspI and EcoRI; Lane D, PstI. Following electrophoresis through neutral agarose (10 µg per lane), the DNA was transferred to nitrocellulose and hybridized to single-stranded $^{32}$P-labeled probe representing the 0.3 kb, 5' PstI·HindIII fragment of pCMF5.
fragments. (1) The EcoRI did not function in the double digest. This is unlikely because the control double digest of pBR322 performed in parallel was successful. (2) The signal from the MspI product(s) and the MspI/EcoRI product(s) resulted from hybridization to two or more fragments of approximately equal size. This seemed unlikely because the intensities of the signals were about equal to those of each of the products detected in EcoRI and PstI single digests. (3) The MspI product and MspI/EcoRI product were identical, derived from the same genomic address as one of the EcoRI and one of the Pst I fragments. The other MspI fragment, corresponding to the other EcoRI and PstI fragments, was not detected because a high linear density of MspI sites therein rendered it refractory to hybridization. (4) The two EcoRI and two PstI fragments were due to allelic polymorphism. Neither hypothesis (3) nor (4) could be reasonably excluded, thus unequivocal explanation of these observations awaited the characterization of the relevant sequences.

The insert of pCMES was isolated, nick-translated, and used to screen five genome equivalents from Library I. After plaque purification, preliminary restriction and Southern analyses demonstrated that five clones were non-identical. One was identical to λCMES1 and was not considered further.
The remaining four were designated λCME15, 2A, 2B, and 20B, respectively. The linear order of the cloned DNAs along the malic enzyme gene was determined by probing Southern blots of RsaI-digested malic enzyme cDNA with nick-translated, exon-containing genomic EcoRI fragments (Fig.8).

The inserts of only two clones were shown to overlap. Clone 15 overlaps the 5' end of λCME1 and contains a 5.9 kb, exon-containing EcoRI fragment as predicted by previous analyses. Subsequent Southern analyses revealed no overlap between clones 15A, 2A, 2B, or 20B. Most significantly, exons in clones 15A, 2A, and 2B hybridize only to sequences 3' of the unique HindIII site in the malic enzyme cDNA, whereas exon(s) in λCME20B reside(s) exclusively within a 3.4 kb EcoRI fragment and hybridize(s) only to sequences 5' of the unique HindIII site in the malic enzyme cDNA. This 3.4 kb fragment was derived from either the 7.5 kb or the ~15 kb EcoRI fragments previously detected in total genomic DNA by ssCME1PH5'. It was located at one end of the insert and truncated by partial digestion used in the preparation of Library I. Recent sequence analyses of exon-containing fragments of λCME2B have clarified its relationships to pCME5 (Fig.8.)

The relationship between the exon(s) of λCME20B and the fragments of genomic DNA to which the 5' cDNA probe
Fig. 8. Partial restriction maps of genomic clones encoding portions of chicken malic enzyme mRNA. Inserts were cloned in the lambda vectors EMBL4 (#245) and Charon 4A (all others). The order of the inserts was determined by hybridization of exon-containing fragments to specific fragments of the chicken cDNA and/or analyses of the nucleotide sequences. In those cases where the location of an exon has been confirmed by sequencing, a line has been drawn to connect the address with the corresponding segment of the cDNA. The absence of a connection does not necessarily reflect the lack of exons from a genomic clone; hybridization analyses suggest the presence of exons in λCME2A and the 5'-most EcoRI fragment of λCME2B (not shown). Short vertical segments represent gaps that exist between clones. A, AvaI; B, BamHI; E, EcoRI; H, HindIII; Ps, PstI; Pv, PvuII; R, Rsal; S, Sphi; X, XbaI. Restriction sites are indicated only when they were verified for that specific clone; it does not mean that they are absent from overlapping regions.
hybridized was defined by performing Southern analyses using erythrocyte genomic DNA from three chickens, each from a different inbred strain. Blots were hybridized with the 5' cDNA probe or the exon-containing, 0.8 kb PstI·EcoRI fragment of λCME20B (ss20BPtex\(^+\)). For each restriction enzyme used, the genomic probe detected one of the two fragments detected by the cDNA probe (Representative blots are shown in Fig.9). These results and restriction analysis of the corresponding region of λCME 20B indicate the following: (1) The region lacks a high linear density of sites for MspI, hence the exon-containing MspI fragment should not be refractory to hybridization in Southern analysis. (2) This region of λCME20B shares the genomic address of the 7.5 kb EcoRI and 1.5 kb PstI fragments previously detected in total genomic DNA. Hence polymorphism cannot account for results of previous Southern analyses of total genomic DNA. (3) The balance of the 5' end of the cDNA is encoded by exon(s) residing within fragments liberated by PstI (0.60 kb), HindIII (8.4 kb), BamHI (~15 kb), and EcoRI (~15 kb). These conclusions are supported by the sequence analysis of the exon-containing region of λCME20B and isolation of additional, overlapping clones.

The nucleotide sequence of the exon-containing region of λCME20B was determined by M. Glynias and S. Klautky of our laboratory. Comparison of the sequence to that of the
Fig. 9. Evaluation of potential restriction site polymorphisms in the malic enzyme locus of the chicken. Genomic DNA was isolated from the erythrocyte nuclei of a single chicken from the inbred line "15". Following digestion with various restriction enzymes and electrophoresis through neutral agarose (10 µg per lane), the DNA was blot-transferred to nitrocellulose and hybridized to single-stranded $^{32}$P-labeled probes representing the 5' PstI·HindIII fragment of pCME5 (A) or the 0.8 kb, exon-containing PstI·EcoRI fragment of $\lambda$CME20B (B). In both cases, Lanes A–E correspond to digestion with PstI, EcoRI, PstI plus EcoRI, BamHI, and HindIII, respectively. Horizontal arrows in (A) designate two apparent signals in Lanes A and C. Qualitatively similar results were obtained with DNA isolated from chickens from inbred lines "72" and "P". OR, origin.
5' PstI·HindIII fragment of malic enzyme cDNA revealed that a single exon encodes a sequence of 132 nt whose 5' end maps 171 nt upstream of the HindIII site in the cDNA. We predicted that the remaining 39 nt were encoded by exon(s) residing within the 3' part of the 7.5 kb genomic EcoRI fragment not contained in λCME20B. The 5' part of the 5' PstI·HindIII fragment of the cDNA most likely was encoded by exon(s) within the ~15 kb EcoRI, hence the ~0.60 kb PstI fragment.

Library I was screened repeatedly to isolate the 5' end and regions of the gene downstream from λCME20B. Fifteen genome equivalents were probed independently with (a) CMEPh5', (b) 5'-most EcoRI fragment of λCME20B ("5' walk"), and (c) 3'-most EcoRI fragment of λCME20B ("3' walk"). The following conclusions are based on plaque hybridizations, restriction mapping, and Southern analyses of the genomic clones.

Library I was screened with randomly primed (rp) CMEPh5'. This strategy was appropriate because (a) assuming the 5' end of the gene resided in a clone that was under-represented in the library, screening a significantly larger number of recombinants would increase the probability of detecting the clone; (b) given a limit of $10^6$ counts per ml of hybridization solution, probing with only the 5' fragment would maximize the concentration of those sequenc-
es in the screening; and (c) a probe prepared by random priming is preferred to one that is "nick-translated" because the probe population, undoubtly heterogenous, should not contain extremely short molecules that will hybridize inefficiently, or not at all, to the target sequences. The potential technical problem outlined in (c) above could arise if a particular DNA fragment is especially sensitive to DNase. This would compromise the ability to detect the target, particularly when, as in this instance, it consists of one or more exons of unknown length and base composition. No clones encoding the 5' end of the cDNA were detected. Two overlapping, exon-containing clones were obtained (λCME1A, 14A), both of which were linked to λCME20B (Fig.8; only 14A is illustrated because it extends further in the 3' direction than the other). If the library contained the 5' end of the gene, the region was either drastically under-represented or packaged in recombinants that could not be detected with cDNA probes.

The 5' walk was of great interest because it might result in the isolation of the 5' end of the gene. Prior to screening the library, randomly primed EcoRI fragment was used to probe a Southern blot of erythrocyte genomic DNA to determine if repeated sequences were present in the fragment. The probe hybridized to a single EcoRI fragment of ~12 kb and two BamHI fragments of ~13 and ~15 kb
(Fig.10). Since the probe contained BamHI sites separated by 0.7 kb, these data indicated that it was a unique fragment in the chicken genome and validated its use in screening. Since the 4.8 kb probe abuts the arm of λCME20B, it was not surprising that it was derived from a larger genomic EcoRI fragment. The sizes of the two BamHI fragments were too large to estimate accurately; however, the BamHI maps of genomic clones λCME 1A, 14A, and 20B predicted a BamHI product of ~13 kb in length. The smaller of the two fragments detected by the 4.8 kb probe yielded the stronger autoradiographic signal as expected, because, based on length, it should bind ~68% of the probe. Curiously, the 0.7 kb BamHI fragment spanned by the probe was not detected; this could result from a restriction site polymorphism. In any event the genomic probe detected neither the large EcoRI fragment (~15 kb) nor the large HindIII fragment (8.4 kb) detected by the 5'cDNA probe, suggesting that this region of the malic enzyme gene locus might be organized as shown in Fig.10.

I planned to employ the 5'-most genomic fragment of λCME20B to maximize the probability of detecting an overlapping clone which extended 5'-ward, and then test for the presence of the desired exon(s) using the 5' cDNA probe. At the least, I would isolate an overlapping clone(s). The 5' walk produced nine clones, four of which did not hybridize
Fig. 10. Southern analysis of total genomic DNA from chicken liver using the 5'-most EcoRI fragment of λCME20B. Genomic DNA was isolated from hepatic nuclei, digested with BamHI (Lane A) or EcoRI (Lane B), and subjected to electrophoresis through neutral agarose (10 μg per lane). Following transfer to nitrocellulose, the DNA was hybridized to randomly primed, $^{32}$P-labeled 5'-most EcoRI fragment of λCME20B. The resulting autoradiogram is shown at the top. A schematic diagram proposed to account for the result is shown at the bottom. The 5'-most EcoRI fragment of λCME20B used as probe is designated with asterisks. A hypothetical restriction map of the overlapping region of the malic enzyme locus is shown in parallel. A vertical arrowhead points to the BamHI·EcoRI fragment proposed to encode the 5' end of pCME5. Or, origin.
to the 3' end of λCME20B. Two of these (λCME42, λCME55) were unique with respect to length and provided an additional 12 kb of the gene (Fig.8). Neither clone hybridized to the 5' cDNA probe, thus both appear to consist exclusively of intronic sequence.

The 3' walk yielded the balance of the 7.5 kb EcoRI fragment detected in total genomic DNA by the 5'cDNA probe, as well as additional sequences downstream of λCME20B. λCME13 contains a 7.5 kb EcoRI fragment and a 1.5 kb, exon-containing PstI fragment (Fig.8). A 38 nt oligodeoxynucleotide specific for the 39 nt of the cDNA unaccounted for in λCME20B did not hybridize to the PstI fragment. The clones described to this point constitute ~85 kb of the malic enzyme locus.

The inability to isolate genomic sequence encoding the balance of the 5' end of the cDNA suggested that it might be missing from Library I. All clones discussed thus far were isolated from genomic Library I constructed in Charon 4A. Ideally, its construction involved partial HaeIII and AluI digestion products. Thus, certain sequences will be under-represented in the library if they contain abnormally high or low numbers of sites for HaeIII or AluI, preventing their inclusion in the library. Alternatively, the desired recombinant may have been constructed in vitro, but was either never established or inefficiently propagated in in-
tact bacteria (105-108). Its insert may have been of a length incompatible with packaging or it may have been a substrate for nucleases in the host. Genomic DNA from plants and animals contains significant quantities of methylcytosine (109). Methylated nucleotides may be substrates for site-specific restriction systems that recognize the modified sequence. If established, such recombinants may grow poorly, as have recombinants whose inserts approach the maximum length tolerated by the bacteriophage (108). We sought an alternative collection of recombinants to enhance the probability of isolating the balance of the gene.

A library of chicken erythrocyte DNA was generously provided by Dr. H.-J. Kung of Case Western Reserve University. Recombinants had been prepared in the bacteriophage lambda replacement vector, EMBL 4, in the following manner. Total erythrocyte DNA was partially digested with EcoRI. Fragments from 14-20 kb in length were isolated and ligated to the previously purified "arms" of EMBL 4. Recombinant molecules were then packaged into viable phage particles in vitro. This library, Library II, was particularly valuable because it was constructed using partial EcoRI digestion products and had not been amplified.

Fifteen genome equivalents of Library II were screened with the \( \text{rpCME}_{\text{PH5}} \). Twenty-four putative positives were detected, five of which contained a ~15 kb EcoRI fragment
that hybridized to the 5' cDNA probe. Each of these five clones also contained a ~0.60 kb PstI fragment that hybridized to the probe, consistent with the results of genomic Southern analyses. One of the four clones, λCME245, was selected for further analysis. A partial restriction map of CME245 was constructed with the assistance of X.-j. Ma and S. Klautky of our laboratory (Fig.11). Assuming that pCME5 hybridizes to the 5'-most exon of the malic enzyme gene, the data presented above strongly suggested that the site(s) of initiation of transcription reside(s) within the ~0.60 kb, exon-containing PstI fragment of λCME245.

D. Structural Analysis of the Putative Promoter

I used three distinct but related analyses to determine if λCME245 contained the start site(s) for transcription of the malic enzyme gene: (a) determination of the primary structure of the exon-containing PstI fragment, (b) primer extension analysis, and (c) S1 nuclease analysis. Considered in toto, these analyses indicate that the 5'-most exon within the fragment is the first exon.

The primary structure of the putative promoter region was determined using the chain termination method of Sanger (97). The exon-containing PstI fragment was subcloned into M13mp18. Single-stranded DNAs containing either the sense or anti-sense strands were used as templates in conjunction
Fig. 11. **Partial restriction map of λCME245, a recombinant bacteriophage that encodes the 5' end of chicken malic enzyme mRNA.** λL, left arm of EMBL 4; λR, right arm of EMBL 4; A, AvaI; B, BamHI; Bs, BstEII; E, EcoRI; H, HindIII; N, NarI; P, PstI; R, RsaI; S, SmaI; St, StuI; Xb, XbaI; Xh, XhoI. **Solid bars,** flanking regions of EMBL 4. The exon-containing, 0.6 kb PstI fragment was subcloned into pIBI31 for detailed mapping and isolation of fragments to be used in S1 nuclease mapping. The open box indicates the approximate location of the first exon.
with Sequenase, initially under standard conditions. Strong "stops" were occasionally encountered, indicative of significant secondary structures in the templates. Unambiguous data could not be obtained with Sequenase regardless of the use of (a) enzyme at twice the standard concentration, (b) dITP, or (c) temperatures as high as 45°C. Employing Taq polymerase at 70°C resolved the ambiguities. The consensus sequence, presented in Fig.12, consists of 609 nt, including 87 nt of an intron, at least 129 nt of an exon, and no more than 393 nt of potentially flanking sequence.

The location of the start site of transcription was estimated by primer extension analysis. We synthesized an oligodeoxynucleotide primer (35 nt) in which the 3' end corresponded to the 3' A residue within the second "Kozak box" (Fig.12). It was used to prime "reverse" transcription of polyadenylated RNA isolated from chick-embryo hepatocytes treated with insulin and T3 to increase maximally the relative abundance of malic enzyme mRNA. The two major products were approximately 124 and 191 nucleotides in length, respectively (data not shown). Two potential sites for initiation of transcription were assigned accordingly (Fig.12). A second primer extension experiment was performed with polyadenylated RNA isolated from the liver of a chicken which was starved for 48 hr and subsequently fed for 24 hr. The primer was a second oligodeoxynucleotide (22 nt) whose 3' end corresponded to the first A residue 3' of the RsaI
Fig. 12. **Consensus nucleotide sequence of the coding strand of the promoter region of the chicken malic enzyme gene.**

The sequence of each strand was determined independently using M13mp18 recombinants, the universal primer, and two internal primers each specific for one of the strands. Recognition sequences for PstI which flank the insert are underlined twice. The recognition site for Rsal, a convenient reference for the 5' end of the cDNA, is underlined once. Putative Kozak boxes are designated K1 and K2, respectively. The proposed exon/intron junction is underscored with dots. A pyrimidine-rich sequence is italicized. Potential sites for the initiation of transcription based on primer-extension analyses are designated as follows: two sites mapped by priming hepatocyte RNA, vertical arrows above the sequence; multiple sites mapped by priming chicken liver RNA, vertical arrows beneath the sequence. In the latter case, the solid arrows correspond to the most abundant primer extension products. Sites for the initiation of transcription mapped by S1 analyses are indicated with squares above the sequence. Closed squares correspond to the most abundant S1-resistant products. Potential regulatory elements are underlined. The arrowhead indicates the orientation. Designations are as follows: rS14 T3RE, sequence similar to the thyroid hormone response element in the rat S14 gene; rGH, sequences conforming to the consensus for T3 response in the rat growth hormone gene; hP CRE, an identity to the CRE of the human proenkephalin gene; hP AP-2, identity to the AP-2 element of the human proenkephalin gene; rP/T AP-2, sequence similar to the AP-2 elements of the genes encoding rat prolactin and tyrosine aminotransferase. In addition, asterisks underscore a sequence common to the rat and chicken malic enzyme promoters which includes a core AP-1 element. A consensus Sp1 site is underscored with a wavy line.
5' CTGCAGGACT GACGGGCTTC GTCCCCGGGC AGCGCGCTTT ATCTTCGGCA
   CAARAAAATAAG CGTGGAGGAG CAGGAGCCCT GTCGAATTC CTCCCTAGCA
   GGATTAGGGA GCAGCCTGTG GCCGGTAGCG CTCAGGGCTT AGACCGCAGGG
   - 514 TgRE - 1
   AGGCAGGGGAG CCGCCTCAGC ACCGAGGCCT TCCGTCGGGA GGGTTTCGCG
   CCGGCGGGGG TGGCCCCGGTG GGTGACTCGAG CGCGCGCCGG GTGACCCCC
   hP AP-2
   GCCTGGCGGT GGGGCCTCCTG CGCGGCTCCTG CCCCCCTCCTT CGTTCCTCCC
   K
   CACTCTCTCT CTCTCTTCC CGCAGATCAT CCACCAGCCG GCGTCACCTC
   hP AP-2
   AGCAGGAGGAG GTCGCGCGCGG TTTGCTCCTG TCCCGCAGGA CGTACGGCAGT
   K1
   ACGCCTCGCC GCCGGAGTCC GTCGCCATGC AAGCAGGGAA GCCAGCAGGCC
   K2
   AGCCCGGGAG GCGCGAGGCC CTTACAGCTC ACCATGAAGA GGGGCTACGA
   ~~~~~~~
   GGTGCTCCCAG ACCGRCACCT CAACAAGGTA GGCAGGTGC CAGCCTCTTG
   hP AP-2
   CCCTCGGCGC CTTATCTGCAG GCCGGGGCGT GCGTGCCGGG CTCTTCTCTT
   CHATAGGACC TGCAG
recognition sequence (Fig.12). The two major products were 40 and 41 nucleotides in length (Fig.13), corroborating the previous result. These two major products correspond to the ~124 nt product detected in the first experiment, the resolution of which was apparently inadequate to resolve the two fragments. An identical relationship exists between the two largest products visible in Figure 13 and the ~191 nt product detected in the first experiment. The enhanced resolution of the second analysis, coupled with a prolonged exposure of the gel (not shown) allowed assignment of multiple putative start sites, including eight apparently preferred (Fig.12). Barring substantial physical differences between the poly(A)+ mRNAs isolated from liver or hepatocytes, the data suggest that transcription is initiated from similar, if not identical, sites in vivo and in hepatocytes.

This assay may lead to erroneous conclusions. Products might result from (a) premature termination of transcription of a full-length template (b) complete transcription of a minimally truncated, stable mRNA, or (c) transcription spanning regions of the mRNA formed by splicing of exons. Analysis using S1 nuclease allows one to estimate the length(s) of the exonic sequences, hence the approximate locations of intron/exon junctions, within a particular genomic fragment. In conjunction with the primer extension experiment, the results reveal if the 5'-most exon within
Fig 13. Primer-extension analysis of malic enzyme mRNA.

Twenty μg of chicken liver poly(A)+ RNA (Lane A) or yeast tRNA (Lane B) were hybridized with end-labeled MEOL 100 and the primer was extended with reverse transcriptase. The locations of products are indicated by horizontal arrows. The last four lanes (G, A, T, and C) contain products of dideoxy sequencing reactions performed with M13mp18 and universal primer. Not shown are the end-labeled primer, which appeared as a single band, and the products of extending malic enzyme RNA that was synthesized in vitro. The lengths of those products were consistent with the site of initiation of transcription in the pGEM recombinant used to prepare the synthetic malic enzyme RNA.
the fragment is the first exon and determine if the length of the primer-extended products reflect complete transcription of the mRNA.

The first S1 nuclease experiment utilized a 264 bp HaeIII fragment spanning all potential sites of initiation. Its 3' end is 68 nt downstream of the RsaI site in the cDNA, hence its 5' end is 196 nt upstream of the RsaI site. It was 5'end-labeled and subjected to two rounds of preparative electrophoresis through 6% native polyacrylamide gels. Autoradiography following the second electrophoretic separation revealed no other radioactive species in the gel. Following hybridization to polyadenylated hepatocyte RNA and digestion with S1 nuclease, S1-resistant radioactive species were resolved by electrophoresis through a denaturing polyacrylamide gel and visualized by autoradiography (data not shown). The presence of poly(A)+ RNA protected a segment of the probe from digestion. A fragment of similar size (72 nt) was protected by hybridization with malic enzyme "sense" RNA synthesized in vitro. The length of the fragment is consistent with the location of the major start site predicted by the primer-extension experiments. The experiment was repeated with a strand-specific probe not only to verify the result, but also to improve its resolution.

The second S1 nuclease experiment utilized a probe derived from the 264 bp HaeIII fragment described previ-
ously. The 5'end-labeled fragment was purified by preparative gel electrophoresis and digested with BstNI. This yielded a 49 bp HaeIII·BstNI fragment labeled at the 5' end of the 'm' strand and a 215 bp BstNI·HaeIII fragment labeled at the 5' end of the 'c' strand. The latter fragment was isolated following electrophoresis of the reaction mixture through a 6% native polyacrylamide gel. The 215 bp fragment was clearly separated from both the substrate, which was present in trace amounts, and the 49 bp fragment. Following its hybridization to polyadenylated RNA from chicken liver or chick-embryo hepatocytes and digestion with S1 nuclease, S1-resistant species were examined as described above (Fig.14). The presence of poly(A)+ RNA isolated from hepatocytes or chicken liver protected various segments of the probe from digestion. In both cases, the lengths of the most abundant products were 70 ± 2 nt, equal to those resulting from protection by synthetic malic enzyme RNA and consistent with the results of the primer-extension experiments. Thus, transcription of the gene is initiated from multiple sites within the PstI fragment (Fig.12).

This DNA fragment exhibits a variety of salient features (Fig. 12). The 5' flanking region lacks consensus sequences corresponding to the CCAAT and TATA "boxes." It thus differs from promoters typical of genes which are expressed in a tissue-specific fashion. Lack of the TATA
Fig. 14. *S1 nuclease analyses of malic enzyme mRNA isolated from both chick-embryo hepatocytes and chicken liver*. An end-labelled 215 bp BstNI·HaeIII fragment spanning all potential start sites for transcription was hybridized to samples containing 20 μg of the RNA indicated: yeast tRNA (A and B); yeast tRNA plus synthetic malic enzyme RNA (C); polyadenylated RNA isolated from either chick-embryo hepatocytes (D) or chicken liver (E). All samples, with the exception of A, were subsequently treated with S1 nuclease and the radioactive species in all samples were detected by autoradiography. The locations of the S1-resistant species in lanes C, D, and E are indicated by horizontal arrows.
"box" is consistent with the GC-rich character of the promoter and the initiation of transcription from multiple sites. As detailed in Chapter IV., the 5' flanking region contains distinct sequences that exhibit (a) similarity to T3-response elements found in other T3-responsive genes, (b) similarity to the AP-2 sites within the genes encoding rat prolactin and tyrosine aminotransferase, and (c) identity to the AP-2 site of human proenkephalin. Also identified were a putative binding site for the transcription factor AP-1 (110) and a poly(pyr·pur) sequence which may be involved in the regulation of transcription. Two additional features are noteworthy. One putative binding site for the transcription factor Sp1 resides between the two Kozak "boxes" (111). Also, as anticipated from previous data, the balance of the cDNA is encoded by this fragment: the exon1/intron 1 splice junction (AAG/GTAGGG) corresponds to the consensus nucleotide sequence at splice donor sites (112).

The restriction map of this DNA fragment predicted by its nucleotide sequence indicates that it would be cleaved by MspI into three exon-containing fragments of lengths 108, 7, and 119 nt, respectively. Fragments of such small size may transfer inefficiently during Southern blotting, hence escape detection by subsequent hybridization with a malic enzyme cDNA probe containing its 5' end. This adequately accounts for the apparent paradox described earli-
er: the linear density of MspI sites rendered sequences refractory to detection in a Southern analysis.

The AluI and HaeIII maps of the fragment also were examined. One site for AluI is present. Six sites for HaeIII exist. Significantly, either enzyme cuts only once within the exon such that both products should be detected by hybridization with cDNA probes. The possibility exists that the region was a preferred substrate for digestion by one or both of the enzymes during the preparation of Library I. Hence, the exon would not have been incorporated into recombinants.
CHAPTER IV
DISCUSSION

A. Introduction

The genomic clones described in the preceding chapter contain ~100 kb of non-overlapping genomic DNA derived from the malic enzyme locus in the chicken. The lengths of gaps existing between various clones have not been estimated, hence the absolute length of the transcription unit is unknown. Regardless of that value, the gene is one of a very few typified by both its overall great length and the high ratio of the combined length of its introns to that of its exons. Representative genes include those encoding malic enzyme in the rat (>100 kb; 113), c-abl (230 kb, including the first intron of at least 200 kb; 114), the receptor for epidermal growth factor (110 kb, including the first intron of 18 kb; 115), and dystrophin (two megabase pairs; 116). The latter case is extraordinary: the length of the cognate mRNA (14 kb, 117), while significant in its own right, is merely 0.7% of the length of the transcription unit.

The significance of the lengths of these genes, thus that of their introns, is unknown. The similarity of the lengths of the genes encoding malic enzyme may reflect the conservation of regions of substantial length which, in turn, conserves function(s). These could involve the higher order structure of the chromosome in fundamental nuclear
processes. Alternatively, these regions of the genome may be more complex than presently assumed. Their introns may be more than "raw" genetic material previously regarded as susceptible to high frequencies of evolutionary drift with regard to both primary sequence and size. Indeed, the Gart locus of *D. melanogaster* contains a gene within its first intron which is transcribed from the opposite DNA strand (118). Another *Drosophila* locus encodes dopa decarboxylase and shares an 88 bp region at its 3'end with the 3'end of a transcript from an unidentified gene on the opposite strand (119). Additionally, a locus in the rat contains overlapping genes, each occupying one strand of the same region of DNA: the mRNA encoding gonadotropin-releasing hormone shares a significant amount of exonic sequences with an RNA (SH) of undefined function encoded by the opposite strand (120). Evaluating such possibilities involving malic enzyme awaits the detailed analyses of the introns, including determination of their primary structures when warranted by preliminary data.

The primary structures of the proximal 5' flanking regions of the genes encoding both chicken (this text) and rat (121) malic enzyme have been determined. Generally, the regions are quite similar. Neither flank contains a sequence conforming to that of the canonical "CAAT box." The rat promoter contains a "TATA box" 622 nucleotides upstream of the major cap site, but it is apparently non-functional
because start sites mapping immediately downstream of that coordinate were not found. Such a sequence is absent from the chicken promoter. The lack of a functional TATA box is consistent with transcription initiating at multiple sites within both promoters. Both flanks share a putative binding site for the transcription factor AP-1 and are abundant in G and C residues. The rat promoter contains seven "GC boxes;" six are upstream from the major cap site and one is in the untranslated region of the mRNA. The latter feature is shared by the chicken sequence, which contains only one "GC" box located between two putative Kozak "boxes." The regions contrast significantly in that the chicken promoter lacks "GC boxes" and contains a pyrimidine-rich tract absent from the rat promoter. These differences raise interesting questions regarding how the two promoters are regulated similarly despite differences in primary structure.

Accordingly, the ensuing narrative considers specific modes of regulation of transcription and their potential relevance to the chicken malic enzyme gene. Pending thorough functional analyses of the promoter, this exercise is admittedly speculative in nature. Nevertheless, it is warranted in light of the documented effects of both T3 and glucagon upon transcription of the gene.
B. Transcription of Eukaryotic Genes and Its Regulation

Regulation of production of mRNA most commonly involves transcription, typically involving changes in rates of initiation at one or more sites. This is not surprising in prokaryotes because transcription is intimately coupled to translation. It is also conceptually appealing in eukaryotes because initiation represents the first committed step in a complex pathway. Revolutionary advances have been made in defining the molecular mechanisms of initiation, particularly in identifying cis-acting sequences and trans-acting factors which are required for either constitutive or regulated rates of transcription initiation.

The operative mechanisms in higher eukaryotes are complicated, involving both promoters and enhancers. Promoters reside very near the site(s) of initiation and function in a position dependent manner. A conserved sequence that establishes the accuracy of initiation (TATAAT, "TATA box") resides ~30 bp upstream of the cap site within the promoters of numerous genes transcribed by RNA polymerase II (122). It is typically preceded by a sequence of consensus GGGC TCAATCT ("CAAT box"; 122). Promoters lacking these elements typically contain multiple motifs which are GC-rich (GG TGGCG GGA GCT, 123), often located within methylation-free "islands" (124), and recog-
nized by the factor Sp1 (111). Within the latter class of promoters, transcription often initiates at multiple sites, exemplified by the gene encoding HMG-CoA reductase (125). The functions of these promoters are often augmented by enhancers. These modules can substantially activate transcription over long distances independently of their orientations either up- or down-stream of the site(s) of initiation (122).

Both promoters and enhancers are often interspersed with elements which mediate transcriptional responses to environmental stimuli. Examples include motifs which confer the responsiveness of metallothionein to copper (126), HMG CoA reductase to sterols (127), bovine prolactin to glucocorticoids (128), rat growth hormone to thyroid hormone (129), and the human glycoprotein hormone α-subunit to cAMP (130). These cis-acting elements interact with protein factors to alter rates of transcription.

The various trans-acting factors identified to date generally consist of domains, usually separable, that confer ability to bind DNA or activate transcription (131). The DNA binding activity is typically limited to a sequence of 60 to 100 amino acids and is necessary, but not sufficient, for the activation of transcription. The structures involved in binding vary markedly, including "zinc fingers" in Sp1 (132) and the "leucine zipper" in CREB which is re-
quired for its dimerization (133). Regions capable of activation are also relatively small, ranging in size from 30 to 100 amino acids (131). These structural motifs are equally diverse in primary structure, containing regions that are highly acidic, or abundant in glutamine or proline. Presumably, the existence of various activation domains imparts the required flexibility to interact with other molecules, including the subunits of RNA polymerase.

1. **Regulation of transcription by thyroid hormone.**

Many effects of thyroid hormone are initiated by its binding to a non-histone, chromosomal protein (134-136). The interaction is proposed to affect the rates of accumulation of specific nuclear RNAs by changing the rates of transcription of the corresponding genes. This hypothesis is exemplified by the induction of transcription of the growth hormone gene by T₃ (137). The stimulation is dependent upon the specific interaction of the T₃ receptor with sequence elements flanking the gene (138).

Numerous promoters have been analyzed in an attempt to identify the cis-acting sequences conferring responsiveness to T₃. Induction is mediated by the 5' flanking regions of the genes encoding growth hormone (139,140) and prolactin (141) in rat. Negative regulation is mediated through the 5' flanks of the genes encoding rat prolactin (141) and rat glycoprotein hormone α-subunit (142). In each
case, binding of the T₃ receptor to the 5' flanking regions has been reported. Nevertheless, a consensus sequence for the T₃RE has not been identified. Lack of a common sequence in the α-subunit promoter is not necessarily surprising because its transcription is suppressed by T₃. The situation with the other genes has yet to be clarified. Mutational analyses of the prolactin promoter demonstrated that the sequence GGTCA is required for T₃ to exert stimulatory effects. Similar dissection of the growth hormone promoter revealed the existence of four response elements that each bind T₃ receptor. Alignment of the four binding sites suggested the involvement of two purine-rich regions. The 3' region (consensus: GGGÅ TCG/Å) is highly conserved. Four segments of the chicken malic enzyme 5' proximal promoter conform to this consensus; each is found on the opposite strand (Fig.12). In addition, the flanking region contains a sequence that is very similar to the T₃RE in the gene encoding rat S14 (143). A similar sequence resides (a) in the rat malic enzyme 5' flanking region and corresponds to a DNase I hypersensitivity site induced by T₃ (144) and (b) in the rat growth hormone gene. These sequences appear to be related to an idealized palindromic motif from which T₃ receptor binding sites have been derived ("Glass palindrome"):
chicken malic enzyme: C A G G G · C T T A G A C · G C A G G
rat S14: C A G G C · C C T T G A C C C C C A G
rat malic enzyme: C A G G C C C C T G G G G · G C A T G
rat growth hormone: C A G G G A C G T · G A C C G C A G G
Glass palindrome: T C A G G T · C A T · G A C C T G A
The significance of these similarities awaits additional functional analyses.

The mechanism by which T₃ regulates transcription of the chicken malic enzyme gene remains conjectural. Examination of the proximal promoter reveals some similarities to T₃REs identified in other promoters. Thus three possible scenarios can be envisioned. (a) Transcription may be regulated directly by interaction with the T₃ receptor and involve sequences that are functionally, but not structurally, identical to T₃REs of other genes responsive to T₃. (b) Despite the noted similarities, the T₃RE within the malic enzyme flanking region may reside 5' of the region subjected to sequence analysis. (c) Transcription of the chicken malic enzyme gene may be regulated indirectly by T₃. If so, the response element would likely bear little resemblance to other T₃REs because the malic enzyme gene would belong to a distinct regulatory network.

2. Regulation of transcription by glucagon.

The effects of glucagon are almost certainly mediated by cAMP. The stimulatory effects of cAMP upon transcription
in eukaryotes are well documented. Comparison of the promoter-regulatory regions of specific genes has resulted in the identification of two cis-acting elements mediating the positive transcriptional response (145). One motif ("CRE", T(G/T)ACGTCA) is common to genes encoding PEPCK, somatostatin, and tyrosine hydroxylase in rat. The second motif ("AP-2 element") is common to genes encoding human metallothionein II\textsubscript{A} and growth hormone, as well as those encoding rat prolactin and tyrosine aminotransferase.

Unlike the CRE, the AP-2 element confers responsiveness to both cAMP and phorbol esters. Similar sequences reside 5' of the major start site in the chicken malic enzyme promoter (Fig.12). A tract identical to the human proenkephalin CRE (CTGCCTCA, 146) resides on the opposite strand. Two tracts identical to the human proenkephalin AP-2 element (CCGCCGGC, 146) are on the sense strand. In addition, the sense strand contains a sequence very similar to the AP-2 elements in rat prolactin (CCCTCCC, 147) and rat tyrosine aminotransferase (TCCCTCCC, 148).

The functional significance of these elements with respect to malic enzyme is unknown. Reports of the inhibition of transcription by cAMP in eukaryotes are rare. Dibutyryl cAMP abolishes the dexamethasone-induced transcription of the glutamine synthase gene (149). Increasing the intracellular concentration of cAMP inhibits transcription
of the transin gene in Rat-1 cells (150). Nevertheless, insufficient data exist to justify speculation on the identity of putative specific cis-acting element(s).

3. **Transcription mediated through the AP-1 motif.**

An AP-1 element can act as a basal-level enhancer (151) and as a mediator of response to phorbol esters (152). Preparations of mammalian AP-1 contain proteins encoded by multiple genes, including c-jun, jun B, jun D, fos, and fra (131). Homo- and hetero-dimerization of these various proteins suggests that transcription mediated by this element may be complex (153). In this context, it is interesting to note that AP-1 proteins and CREB bind weakly to each other's recognition sequence (131). Whether the proteins form heterodimers that combine or, indeed, impart novel functions remains unknown.

Conservation of the motif within the chicken and rat promoters implies a functional significance. Significantly, deletion of a region containing the element from the rat malic enzyme promoter reduces maximal promoter activity by 62% (154). The activity of the element, particularly in regard to phorbol esters, remains to be established for either promoter.

4. **Regulation involving conformational changes in DNA.**

Additional complexity may result from the regulated alterations in the architecture of the promoter. The abun-
dance of the dinucleotide CpG and the presence of a pyrimidine-rich tract are two features of the avian malic enzyme promoter that could participate in this type of regulation. In vertebrates, the rate of transcription and methylation of the cytidine within the CpG dinucleotide are correlated inversely. Methylation may alter interactions between DNA and trans-acting factors either directly by steric hindrance or indirectly by altering chromatin structure (155). Activation of a repressed gene would require expression of a function(s) capable of overriding the methylation, perhaps at specific sites (156). Regulating methylation of the promoter could be integrated, directly or indirectly, with changes in the secondary structure of the pyrimidine-rich region. Similar sequences reside within the promoters of actively transcribed eukaryotic genes (157,158) and are sensitive to digestion by nucleases specific for single-strands (159). In these regards, two points warrant emphasis. The preferred start sites for transcription of the malic enzyme gene map downstream of the pyrimidine-rich tract. In addition, this tract exhibits hypersensitivity to single-strand specific nucleases in supercoiled plasmids in vitro. The 5' flanking region (through -700, mapped with respect to the major site of initiation) exhibits hypersensitivity to DNase in situ in nuclei isolated from nutritionally manipulated chicks (160). This hypersensitiv-
ity is induced by feeding a previously fasted animal. The extents to which covalent modification of the DNA per se and the interaction of trans -acting factors with the promoter cause hypersensitivity are currently foci of intensive investigation.

In summary, the regulation of transcription in higher eukaryotes appears more complex than in prokaryotes. The activity of a promoter may be regulated as functions of both its architecture and the biochemical states of its conate trans -acting factors. Its specific architecture is determined by the residence of particular elements of sequence, the biochemical states thereof (eg. possibly methylated), and the context in which they appear. The presence within regulatory regions of motifs responsive to different stimuli surely reflects the evolution of exquisite regulatory networks upon which multicellular physiology is based.

The foregoing discussion addressed the potential transcriptional regulation of the malic enzyme gene by specific cellular mediators. The mechanisms governing accumulation of malic enzyme mRNA might not operate exclusively by altering rates of initiation of transcription. The signals regulating expression of the gene may also alter the abundance of transcripts via mechanisms functioning distal to the initiation of transcription. Both diet and hormones
appear to alter the half-life of malic enzyme mRNA. Hence the balance of this discussion addresses the regulation of stability of mRNA.

C. Regulation of Stability of mRNA.

Altering the degradation rate constant of an mRNA will profoundly influence the concentration of the mRNA and the rate of synthesis of the encoded protein. Thus, changes in the stability of mRNA may dramatically determine the ability of a cell to replicate, differentiate, and respond to perturbation. Aggregation of *D. discoideum* exemplifies this mechanism (161,162): the formation of a multicellular structure involves induction of certain genes and the accumulation of transcripts of long half-lives. Upon disaggregation, transcription of the genes ceases and the cognate mRNAs are selectively and rapidly degraded. Thus, decreasing the stability of an mRNA is a rapid method for the cell to reduce the expression of a gene, particularly if its transcription is repressed.

The mechanisms by which the degradation of specific cytoplasmic mRNAs is regulated are beginning to emerge. Various insights have been gained by examining changes in the half-life of a particular mRNA as a function of either its length or the presence of putative regulatory regions in chimaeric molecules. The non-coding, coding, and poly(A) segments, acting alone or in concert, are important determ-
inants of stability of particular mRNAs. Presence of a poly(A) tract influences the rate of degradation of globin mRNA (163,164) and histone mRNA (165). The non-coding segments of c-myc mRNA appear to determine the stability of the mRNA, possibly by interacting with each other. The abundance of β-tubulin mRNA is regulated by the intracellular concentration of tubulin heterodimers (166). Stability of the mRNA is mediated by sequences residing within the first sixteen translated codons of β-tubulin mRNA. Reflecting the differences in regulatory characteristics, the sequences within malic enzyme mRNA that mediate changes in its half-life are probably not shared by the specific mRNAs discussed above.

We possess the reagents necessary to identify the structural determinants of malic enzyme mRNA stability. The 3'-most exon in λCME1 provides the balance of the mRNA not represented in pCME5. The 5' untranslated region is encoded by λCME245. Sequences derived from each can be used in combination with cDNAs for malic enzyme mRNA as well as other mRNAs to test half-lives as a function of nutritional state in vivo and hormonal treatment of cultured hepatocytes.

The foregoing considerations emphasize the potentially complex means by which the information encoded in a gene may be expressed. A general corollary to this statement is that the regulation of networks and interactions thereof
will prove to be considerably more complicated, involving the expression, modification, and localization of molecules exerting control. Implied in the latter statement is the expectation of competition between networks for the mediators.

D. Future Prospects

A combination of classical and novel techniques should elucidate the operative mechanisms of regulation of accumulation of malic enzyme mRNA. Coupled with a thorough examination of the kinetics of induction of previously identified regulatory processes in a transgenic animal model, in vitro analyses will exquisitely couple the physicochemical characterization of the relevant macromolecules with kinetic analyses of flux through a metabolic pathway. Clearly, the immediate challenge is to identify the cis -acting sequences and their cognate trans -acting factors which interact to regulate the rate of initiation of transcription of the chicken malic enzyme gene. In vitro transcription systems should be invaluable in this regard (167). Ultimately, it may be enlightening to employ the scanning tunneling microscope (STM, 168) to make high-resolution observations of nucleic acids in native specimens (169) as a function of both the nutritional state of the whole animal and the hormonal milieu of the cultured primary hepatocyte. The STM has been used to observe both the
alternation of major and minor grooves and helical periodicities in dsRNA and DNA and to quantitate dimensions and structural features of nucleic acids. It may permit viewing of variations in helical structure due to local sequence, transitions between helical forms, and perturbations resulting from the binding of regulatory molecules.
REFERENCES


66. Glyrias, M.G., doctoral dissertation, Department of Pharmacology, Case Western Reserve University,
manuscript in preparation.


Sci. USA 72, 3961-3965.


146. Comb, M., Birnberg, N.C., Seasholtz, A., Herbert, E.,


