Isolation and Characterization of Two Distinct Rainbow Trout Albumin Genes

by

Zhi Quan Gong

A Thesis submitted with the requirements for the Degree of Doctor of Philosophy in the University of Toronto

© Copyright by Zhi Quan Gong 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-27655-4
To My Family
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>i</td>
</tr>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>viii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>x</td>
</tr>
<tr>
<td>Abstract</td>
<td>xi</td>
</tr>
</tbody>
</table>

## Chapter One: Introduction

1. Albumin structure and chemical properties........................................ 2
   1.1 Primary structure ................................................................. 2
   1.2 Tertiary structure ...................................................................... 6
   1.3 Physicochemical properties ........................................................ 6
2. Ligand binding by albumin..................................................................... 8
3. Albumin gene and the albumin supergene family.................................... 9
   3.1 Albumin gene .................................................................................. 9
   3.2 Albumin supergene family ............................................................. 10
   3.3 Evolution of albumin supergene family......................................... 13
4. Albumin gene expression...................................................................... 15
   4.1 Tissue-specific and developmental expression of albumin gene ....... 15
   4.2 Extracellular signals and albumin gene expression ....................... 17
5. Regulation of albumin gene expression .............................................. 19
   5.1 Albumin promoter and enhancer................................................... 19
   5.2 Transcription factors involved in albumin gene expression............ 26
   5.2.1 Transcription initiation and activation .................................... 26
5.2.2 Transcription factors involved in albumin gene transcription ................. 29
6. Objectives of the current study................................................................. 41

Chapter Two: Isolation and Characterization of Two Distinct Rainbow Trout
Albumin cDNAs

Introduction ............................................................................................. 45
Materials and Methods .......................................................................... 46
1. Tissue collections .................................................................................. 46
2. cDNA library construction and screening ......................................... 46
3. Cloning of the 5' fragment of rainbow trout albumin 1 cDNA by reverse
   PCR ....................................................................................................... 48
4. Restriction mapping and DNA sequencing ........................................ 49
5. RNA analysis ......................................................................................... 51
6. Nuclear run-on experiments ............................................................... 53

Results ....................................................................................................... 57
1. Northern hybridization with asALB-30mer probe............................... 57
2. Isolation of two distinct rainbow trout albumin cDNA clones ............ 57
3. Characterization of rainbow trout albumin 2 cDNA ........................... 59
4. Characterization of rainbow trout albumin 1 cDNA ............................ 63
5. The long 3' untranslated region in the rtALB1 cDNA ......................... 65
6. Comparison of fish albumins ............................................................... 69
7. Tissue-specific expression of rainbow trout albumin gene .................. 73
8. Developmental expression of rainbow trout albumin gene ................ 75
9. Different levels of rtALB1 and rtALB2 mRNA ................................. 78
10. Different transcriptional rates of rtALB1 and rtALB2 genes ................. 78

Discussion .............................................................................................. 83
Chapter Three: Isolation and Characterization of Two Rainbow Trout Albumin Promoters ........................................................................................................... 89

Introduction ................................................................................................................................. 90

Materials and Methods ......................................................................................................................... 92

1. Preparation of DNA from rainbow trout testis ........................................................................... 92
2. Restriction digestion and fractionation of high molecular weight DNA .................................... 92
3. Construction and screening of genomic DNA library ................................................................. 93
4. Purification of bacteriophage DNA .......................................................................................... 94
5. Restriction mapping of genomic DNA clone ............................................................................ 95
6. Subcloning of albumin genomic clones ..................................................................................... 95
7. DNA sequencing of albumin gene ............................................................................................ 95
8. Southern hybridization .............................................................................................................. 99
9. Primer extension .......................................................................................................................... 101
10. CAT constructs .......................................................................................................................... 102
11. Cell culture ................................................................................................................................ 104
12. Transfection experiments .......................................................................................................... 104
13. PCR cloning of albumin I promoter ......................................................................................... 105
14. Site-directed mutagenesis ......................................................................................................... 106

Results ............................................................................................................................................ 108

1. Isolation of rainbow trout albumin genomic clones .................................................................. 108
2. Southern analysis of positive genomic clones ............................................................................ 108
3. Determination of the 5' ends of clone 4 and clone 16 ................................................................. 117
4. DNA sequencing .......................................................................................................................... 120
5. Determination of transcription initiation site .............................................................................. 124
6. Southern hybridization .............................................................................................................. 124
7. Sequence of the 5'-regulatory region ......................................................................................... 127
8. Characterization of the 5' regulatory region .............................................................................. 131
9. PCR cloning of rainbow trout albumin 1 promoter ........................................ 133
10. Functional analysis of the two albumin promoters ........................................ 135
11. Mutational analysis of the rtALB2 promoter .................................................. 137
Discussion ............................................................................................................ 142

Chapter Four: Summary and Future directions...................................................... 149

References ............................................................................................................ 154
## List of Figures

<table>
<thead>
<tr>
<th>Figures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1. Amino acid sequence of the human serum albumin</td>
<td>5</td>
</tr>
<tr>
<td>1-2. Tertiary structure of human serum albumin</td>
<td>7</td>
</tr>
<tr>
<td>1-3. Map of the human serum albumin gene</td>
<td>10</td>
</tr>
<tr>
<td>1-4. Comparison of the primary structure of vitamin D binding protein and albumin</td>
<td>14</td>
</tr>
<tr>
<td>1-5. Evolutionary tree of the Alb/AFP/DBP gene family</td>
<td>16</td>
</tr>
<tr>
<td>1-6. Ontogeny of mouse AFP and albumin mRNA levels</td>
<td>18</td>
</tr>
<tr>
<td>1-7. Schematic representation of the cis elements of albumin promoter and trans factors interacting with these cis elements</td>
<td>22</td>
</tr>
<tr>
<td>1-8. Schematic representation of transcription factors interacting with mouse albumin enhancer</td>
<td>24</td>
</tr>
<tr>
<td>1-9. Schematic model of transcription initiation complex assembly</td>
<td>27</td>
</tr>
<tr>
<td>1-10. Conserved domains of the HNF1 transcription factor family</td>
<td>30</td>
</tr>
<tr>
<td>1-11. Functional domains of C/EBPa and model of the C/EBP dimmer via leucine zipper</td>
<td>35</td>
</tr>
<tr>
<td>2-1. Northern hybridization of rainbow trout liver total RNA with asAlb-30mer</td>
<td>58</td>
</tr>
<tr>
<td>2-2. Restriction maps of rtALB1 and rtALB2</td>
<td>60</td>
</tr>
<tr>
<td>2-3. rtALB2 cDNA and derived amino acid sequence</td>
<td>61</td>
</tr>
<tr>
<td>2-4. Reverse PCR amplification of the 5' fragment of rtALB1</td>
<td>64</td>
</tr>
<tr>
<td>2-5. rtALB1 cDNA and derived amino acid sequence</td>
<td>66</td>
</tr>
<tr>
<td>2-6. Comparison of rtALB1 and Atlantic salmon albumin 3' untranslated region</td>
<td>68</td>
</tr>
<tr>
<td>2-7. Comparison of three fish albumin amino acid sequences</td>
<td>70</td>
</tr>
<tr>
<td>2-8. Identities between the three fish albumins</td>
<td>71</td>
</tr>
<tr>
<td>2-9. Schematic representation of the three-domain structure of rtALB2</td>
<td>72</td>
</tr>
<tr>
<td>2-10. Northern blot of rainbow trout total RNA from different tissues</td>
<td>74</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2-11.</td>
<td>Northern blot of rainbow trout total RNA from different developmental stages</td>
</tr>
<tr>
<td>2-12.</td>
<td>In situ hybridization on rainbow trout embryos with rtALB2 cDNA</td>
</tr>
<tr>
<td>2-13.</td>
<td>Different levels of the rtALB1 and rtALB2 mRNA</td>
</tr>
<tr>
<td>2-14.</td>
<td>The probes used in nuclear run-on experiment</td>
</tr>
<tr>
<td>2-15.</td>
<td>Comparison of the transcriptional rates of rtALB1 and rtALB2 gene</td>
</tr>
<tr>
<td></td>
<td>by nuclear run-on assay</td>
</tr>
<tr>
<td>2-16.</td>
<td>Two hypothesis of rainbow trout albumin gene evolution</td>
</tr>
<tr>
<td>3-1.</td>
<td>Strategy of restriction mapping of genomic DNA clones</td>
</tr>
<tr>
<td>3-2.</td>
<td>Restriction maps of genomic clone 4 and clone 16</td>
</tr>
<tr>
<td>3-3.</td>
<td>Subclone strategy for clone g16-5</td>
</tr>
<tr>
<td>3-4.</td>
<td>Nested deletions of clone g16-5H2k</td>
</tr>
<tr>
<td>3-5.</td>
<td>Strategy for rainbow trout albumin2 promoter-CAT construction</td>
</tr>
<tr>
<td>3-6.</td>
<td>Southern analysis of positive genomic clones for trout albumin gene</td>
</tr>
<tr>
<td>3-7.</td>
<td>Restriction mapping of genomic clones g4 and g16</td>
</tr>
<tr>
<td>3-8.</td>
<td>PCR searching the location of the transcription start site</td>
</tr>
<tr>
<td>3-9.</td>
<td>Partial DNA sequence of the rtALB2 gene</td>
</tr>
<tr>
<td>3-10.</td>
<td>Primer extension mapping of the transcription start site</td>
</tr>
<tr>
<td>3-11.</td>
<td>Southern hybridization of genomic DNA with rtALB2 cDNA probe</td>
</tr>
<tr>
<td>3-12.</td>
<td>Southern hybridization of genomic DNA with the rtALB2 5' flanking probe</td>
</tr>
<tr>
<td>3-13.</td>
<td>Potential DNA binding sites for some transcription factors</td>
</tr>
<tr>
<td></td>
<td>in the 5' flanking region of rtALB2 gene</td>
</tr>
<tr>
<td>3-14.</td>
<td>CAT activities of albumin-CAT constructs</td>
</tr>
<tr>
<td>3-15.</td>
<td>PCR cloning of the rtALB1 promoter</td>
</tr>
<tr>
<td>3-16.</td>
<td>Comparison of albumin promoters</td>
</tr>
<tr>
<td>3-17.</td>
<td>Functional comparison of the two rainbow trout albumin gene promoters</td>
</tr>
<tr>
<td>3-18.</td>
<td>Site-directed mutations on the rtALB2 promoter</td>
</tr>
<tr>
<td>3-19.</td>
<td>CAT activities of mutants</td>
</tr>
</tbody>
</table>
List of Tables

1-1. Amino acid composition of serum albumin 4

2-1. Comparison of silent mutation and replacement mutation of some genes in salmon 85

3-1. Summary of southern results of rainbow trout albumin genomic clones 109

3-2. Comparison of cis elements in proximal promoter regions of ALB, AFP, and DBP genes 129
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>α-fetoprotein</td>
</tr>
<tr>
<td>ANF</td>
<td>albumin negative factor</td>
</tr>
<tr>
<td>BST</td>
<td>Bluescript vector</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyl transferase</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CAAT/enhancer binding protein</td>
</tr>
<tr>
<td>CMV</td>
<td>cytomegalovirus</td>
</tr>
<tr>
<td>cpm</td>
<td>count per minute</td>
</tr>
<tr>
<td>DBP</td>
<td>vitamin D-binding protein, or site D binding protein</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>GTFs</td>
<td>general transcription factors</td>
</tr>
<tr>
<td>HBSS</td>
<td>Hanks' Balanced Salt Solution</td>
</tr>
<tr>
<td>HNF1</td>
<td>Hepatocyte nuclear factor 1</td>
</tr>
<tr>
<td>HSA</td>
<td>human serum albumin</td>
</tr>
<tr>
<td>Inr</td>
<td>initiator</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilobases</td>
</tr>
<tr>
<td>kD</td>
<td>kiloDalton</td>
</tr>
<tr>
<td>MEM</td>
<td>minimum essential medium</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-[N-morpholino]propanesulfonic acid</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>Mys</td>
<td>million years</td>
</tr>
<tr>
<td>NF-1</td>
<td>nuclear factor 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>RNA pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>rtALB1</td>
<td>rainbow trout albumin 1</td>
</tr>
<tr>
<td>rtALB2</td>
<td>rainbow trout albumin 2</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TAF</td>
<td>TBP-associated factor</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-binding protein</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactoside</td>
</tr>
</tbody>
</table>
Acknowledgments

I would like to thank my supervisor Dr. Choy L. Hew for giving me the opportunity to work in his laboratory, and for all his guidance and support over the years. I wish to especially thank him for his encouragement during difficult times to make me more independent. I would also like to thank Dr. H. Elsholtz and Dr. F. Keeley for being my co-supervisors and for their valuable advice.

I would like to thank all members, past and present, in Dr. Hew's laboratory and the HSC/Pharmacia Biotechnology Service Center for their friendship and helpful discussions. Special thanks are given to Dr. Zhiyuan Gong, my brother, for helping me to establish the molecular biology techniques and for monitoring the project progress. I also thank Mr. Augustin Nguyen for his assistance in DNA sequencing.

To my family, I am deeply grateful for their love and support which contribute to the completion of this project. To my husband, Dr. Yau Hui, I thank you from the bottom of my heart for being waiting for me overseas and encouraging me to carry on. To my son, Joshua. I thank you for bringing me the most joyful time in my life.

The financial support from the Research Institute of the Hospital for Sick Children and the Department of Fishery and Oceans, Government of Canada is gratefully acknowledged.
Isolation and Characterization of Two Distinct Rainbow Trout Albumin Genes

Zhi Quan Gong
Department of Clinical Biochemistry, University of Toronto

Abstract

Two distinct albumin cDNAs (rtALB1 and rtALB2) were isolated from the rainbow trout (Oncorhynchus mykiss) liver cDNA library. The rtALB1 cDNA had 2761 bp in total which included 69 bp in the 5' untranslated region, 1821 bp in the reading frame to encode 607 amino acids, and a long 3' untranslated region of 872 bp. The rtALB2 cDNA was 2250 bp in length and contained 78 bp in the 5' untranslated region, 1824 bp in the reading frame to encode 608 amino acids, and 348 bp in the 3' untranslated region. The two albumins shared 81.5% and 77.5% identities on the nucleotides and the primary protein sequences respectively. Both rtALB1 and rtALB2 genes were expressed only in adult liver but not in other tissues as tested by Northern blotting. Nuclear run-on assay showed that the rtALB1 gene was transcribed three times higher than the rtALB2 gene. During embryonic development, the albumin mRNA is first detected in the yolk sac membrane and the intestine before the liver is formed. After the liver appears at the 6th week, the albumin gene is highly expressed in this organ.

A genomic clone encoding the rtALB2 was isolated from a rainbow trout genomic DNA library. Approximately 1.2 kb of the 5' flanking region was characterized in a transient expression system. The rtALB2 promoter could direct the expression of a reporter gene (CAT gene) in human hepatoma cells HepG2 but not in Hela cells. Three functional regions were identified in the 5' flanking region: a negative region between bp -1224 to -906, a positive region between bp -487 to -104, and a promoter region within the first 100 bp immediately upstream of the transcription start site. The rtALB2 gene did not have a
typical TATA box in its proximal promoter region and site-directed mutagenesis suggested that it was a TATA-less promoter.

The rtALB1 promoter was amplified by PCR from genomic DNA and sequenced. It contained a canonical TATA box in the -30 to -25 region and a consensus HNF1 binding site upstream of the TATA box, a structure representative of mammalian albumin promoters. Functional analysis showed that the rtALB1 promoter was 3-4 folds more active than the rtALB2 promoter in HepG2 cells, which was consistent with the nuclear run-on results. It was concluded that the differential expression of the two rainbow trout albumin genes is regulated by their different promoter structures.
Chapter One: Introduction

Literature Review on Albumin and Research Objective
Serum albumin is the most abundant protein in plasma. The concentration of serum albumin in mammalian plasma is 33-40 g/L, which accounts for about half of the total serum proteins (Peters and Sjoholm, 1977; Andersson, 1979; Putnam, 1984). The principal role of albumin is to transport fatty acids, bilirubin, and a myriad of other molecules in the blood. Albumin is also a major contributor to the osmotic pressure of the plasma. As a result of its ligand-binding properties and high concentration, albumin provides a stabilizing effect on plasma solute levels. Albumin is one of the earliest proteins to be studied. In the 1800s, albumin was described as a water-soluble protein in the blood (Peters, 1995). Early protein chemists used salt precipitation to fractionate serum proteins and albumin was obtained by the precipitation at 50% saturation of ammonium sulfate followed by lowering the pH to 4.4 and dialyzing the precipitate against water (Andersson, 1979). Since albumin can be easily prepared in pure form and in large amounts, it is one of the most intensively studied proteins. Albumin has been used as a general model in the study of protein conformation, denaturation processes and ligand binding properties. In the past 10-15 years, the albumin messenger RNA and gene has been isolated and characterized. Since the albumin gene is expressed exclusively in liver, it is of special interest in the study of tissue-specific gene transcription.

1. Albumin structure and chemical properties

1.1 Primary structure

The amino acid sequence of human albumin has been determined by both chemical methods and by deduction through cDNA sequencing (Lawn et al., 1981; Dugaiczyk et al., 1982; Minghetti et al., 1986). Human albumin consists of one single polypeptide chain. It is synthesized as a 609-amino acid preproalbumin which contains a signal peptide of 18 residues on the N-terminal followed by a propeptide of 6 residues. The signal peptide is hydrophobic and its sequence has been determined as follows: Met-Lys-Trp-Val-Thr-Phe-
Leu-Leu-Leu-Leu-Phe-Ile-Ser-Gly-Ser-Ala-Phe-Ser. The propeptide of human albumin is a basic hexapeptide: Arg-Gly-Val-Phe-Arg-Arg, the function of which is not clear (Brown et al., 1979; Putnam, 1984). The signal peptide and the propeptide are cleaved in the lumen of rough endoplasmic reticulum and during secretion from the cell, respectively, which leaves the mature albumin with 585 amino acids. The amino acid composition is shown in Table 1-1. A unique feature of the amino acid composition of albumin is the low content of tryptophan, only one or two molecules in the polypeptide. Contents of methionine, glycine and isoleucine are also low. The albumin polypeptide contains a large number of polar amino acid residues which account for its high solubility.

The most unique structural feature of the albumin is their disulfide bonding pattern. Albumin contains 35 cysteine residues in total and 34 of these are involved in 17 disulfide bonds, the single free cysteine sulphydryl is at position 34 in human albumin. Some cysteines appear as adjacent Cys-Cys pairs, each cysteine residue in a pair forms a disulfide bond with the nearest residue in the chain before and after the pair. Thus the polypeptide is cross linked to nine double loops (Fig. 1-1). The first loop at the N-terminus contains a single loop in mammalian albumin, because there is no Cys-Cys pair. The other feature of the albumin primary structure is the occurrence of three repeating homologous domains. The three domains, numbered I, II, III from the amino terminus, correspond to amino acids 1-191, 192-384, and 385-585. Sequence identity among the domains are 18-25%. Each domain contains three double loops (long-short-long). Proline residues are usually found at the tip of the loop. The single tryptophan residue invariably occurs in loop 4 in the second domain.

The amino acid sequence of albumin has been derived from cDNA sequences in many other organisms, such as mouse (Brown and Papaconstantinou, 1979), rat (Sargent et al., 1981), pig (Weinstock and Baldwin, 1988), sheep (Brown et al., 1989), Xenopus laevis (Haefliger et al., 1989; Moskaitis et al., 1989), Atlantic salmon (Byrnes and Gannon, 1990), cow (Holowachuk, 1991), chicken (Cassady, 1991), sea lamprey (Gray and
Table 1-1 Amino Acid Composition of Serum Albumin (from Peters, 1995)

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Human</th>
<th>Bovine</th>
<th>Rat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartic acid</td>
<td>36</td>
<td>40</td>
<td>32</td>
</tr>
<tr>
<td>Asparagine</td>
<td>17</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>Threonine</td>
<td>28</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>Serine</td>
<td>24</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>62</td>
<td>59</td>
<td>57</td>
</tr>
<tr>
<td>Glutamine</td>
<td>20</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Proline</td>
<td>24</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>Glycine</td>
<td>12</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Alanine</td>
<td>62</td>
<td>46</td>
<td>61</td>
</tr>
<tr>
<td>Valine</td>
<td>41</td>
<td>36</td>
<td>35</td>
</tr>
<tr>
<td>Cystine/2</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>Methionine</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>8</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Leucine</td>
<td>61</td>
<td>61</td>
<td>56</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>18</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>31</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Lysine</td>
<td>59</td>
<td>59</td>
<td>53</td>
</tr>
<tr>
<td>Histidine</td>
<td>16</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Arginine</td>
<td>24</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>585</td>
<td>583</td>
<td>584</td>
</tr>
</tbody>
</table>
Fig. 1-1 Amino acid sequence of human serum albumin

The numbered vertical bars indicate the positions of introns in the pre-mRNA.
(from Peters, 1995)
Doolittle, 1992), macaque (Watkins et al., 1993), horse (Ho et al., 1993), and cobra (Havsteen et al., 1994). All of them have similar three domain structures as human albumin except for lamprey albumin which has seven repeating domains.

1.2 Tertiary structure

The three-dimensional structure of human serum albumin (HSA) has been determined at a resolution of 2.8 Å (He and Carter, 1992, Carter and Ho, 1994). The X-ray crystallography of HSA shows that it is a heart-shaped or equilateral triangular molecule 80 Å on a side, with average thickness of about 30 Å, and a calculated molecular volume of about 88,249 Å³ (Fig.1-2). The tertiary structure confirms the three homologous domains predicted by the primary structure. Each domain comprises two subdomains (A and B) that possess common structural motifs. The albumin molecule is highly helical, with 67% of the residues involved in a total of 28 α-helices. There is no β-sheet in the molecule. The remaining residues are either in β-turns (10%) or as an extended peptide (23%). There are 10 α-helices in each domain (h1-h10). The albumin tertiary structure is unusual in that each of its adjacent Cys-Cys pairs is included in a helical segment, where their bonds always join helices. Of other amino acid residues that are often classified as helix breakers, 7 of the 24 prolines, 11 of the 17 asparagines, and 6 of the 12 glycines occur within the helices. It is found that the helical bundles of subdomains IIA and IIIA form hydrophobic cavities as principal ligand binding regions.

1.3 Physicochemical properties

The molecular mass of human serum albumin is calculated as 66 kD from its amino acid composition. Albumin is one of the most soluble plasma proteins. Solutions containing up to 30% of albumin can be prepared. The high solubility of albumin is probably connected with the large number of ionizable groups present in the molecule. At neutral pH the number of charged groups is roughly 200. The isoelectric point is the pH at
Fig. 1-2 Tertiary structure of human serum albumin

(A) space-filling model; (B) stereoview; (C) topological illustration of a typical domain. The N-terminus is at the right and the C-terminus is at the left.
(from Peters, 1995)
which the net charge of a molecule is zero. This is the pH at which a protein will not migrate in an electric field. Since the number of carboxyl groups is somewhat larger than the number of basic groups, the isoelectric point of albumin is lower than 7. The isoelectric point of native human serum albumin was found to be between 4.7-5.5 by isoelectric focusing. This distribution of isoelectric points is related to the fact that different albumin molecules contain different amounts of bound fatty acids. The albumin molecule does not absorb visible light (400-800 nm). The residual yellow color of concentrated albumin solutions is caused by the presence of colored substances such as bilirubin and laemin which bind to the protein. The ultraviolet absorption spectrum of albumin is of the normal protein type showing a maximum at 279 nm and a minimum around 255 nm.

2. Ligand binding by albumin

A characteristic feature of serum albumin is its ability to bind a large number of different substances, such as fatty acids, amino acids (notably tryptophan and cysteine), steroids, metals (calcium, copper and zinc), and numerous pharmaceuticals. The albumin molecule is flexible in solution, which makes it readily adapt to different ligands. Also the three-domain structure of albumin provides a variety of binding sites for ligands.

The binding and transport of long-chain fatty acids is probably one of the main physiological functions of serum albumin. Long-chain fatty acids are those that have 16-20 carbon atoms in the chains. They are highly insoluble and would not dissolve in the blood if they were not solubilized by binding to protein. Among them are oleic acid, palmitic acid, linoleic acid, stearic acid, arachidonic acid, and palmitoleic acid. These substances are crucial intermediates in lipid metabolism. Typically they circulate in plasma at a total concentration just under 1 mM. Yet less than 0.1% of them are really "free fatty acids". Over 99.9% of the total is transported on albumin and loaded and off-loaded with amazing speed. Many binding sites for long-chain fatty acids have been determined, the three strongest ones are located at: (1) loops 8-9 in domain III, involving Lys-475; (2) loop 6 in
domain II, involving Lys-351; and (3) loop 3 in domain I, involving Arg-117. The long-chain fatty acids generally bind in a hydrophobic pocket with its negative charge interacting with the positive charge of a nearby lysyl or arginyl residue. Upon the binding of long-chain fatty acids, the albumin structure is changed to a more compact and more rounded shape, with an increase in stability of the albumin molecule.

Bilirubin is another molecule transported by albumin in the blood. The binding of bilirubin to serum albumin is important. This substance is toxic when present in the free state but binding to albumin renders it non-toxic. Bilirubin is transported by albumin to the liver where it is conjugated and excreted. Unlike long-chain fatty acids, bilirubin has one single strong binding site in albumin which is in loop 4, domain II, involving residues 198-308.

There are two sites in albumin (site I and site II) which are actively involved in binding of various small molecules (Sudlow et al., 1975, 1976). Site I is in subdomain IIA, loops 4-5, involving Lys-199 and Arg-222. Salicylates, some sulfoamides, and other drugs bind to this site. Site II is in subdomain IIIA, loops 7-8, centered around Tyr-411. Ligands for this site include tryptophan, thyroxine, octanoate, and some drugs. The free cysteine residue at position 34 provides a binding site for sulfhydryl compounds and certain oxidants. Two heavy metals, copper (II) and nickel (II) are bound by albumin specifically on the N-terminus with a histidine at position three. Calcium and magnesium are bound by albumin non-specifically. Chloride ions bind to paired basic residues such as Lys-Lys or Arg-Arg (Peters and Sjoholm, 1977; Hamilton et al., 1991; He and Carter, 1992; Peters, 1995).

3. Albumin gene and the albumin supergene family

3.1 Albumin gene

The human albumin gene lies on the long arm of chromosome 4, near the centromere at position q11-13 (Fig. 1-3) (Minghetti et al., 1986; Nishio et al. 1996). It
Fig. 1-3 Map of the human serum albumin gene

The map shows major restriction endonuclease sites and the location of exons (boxes) and introns (heavy lines). Their lengths in nucleotides are indicated by numbers above and below these elements. The first and the penultimate (14th) exons are partly untranslated as indicated by open boxes. The last (15th) exon is completely untranslated. The approximate location of the gene on chromosome 4 is indicated, (from Minghetti et al., 1986)
spans 16,961 nucleotides from the Cap site to the first poly(A) signal. There are 15 exons and 14 introns in the albumin gene. Consensus splice sequence GT(A,G)AGT at 5' ends and (C,T)AG/ at 3' ends are found in the 14 introns. The mature mRNA contains 2250 nucleotides including 1827 nucleotides in the translated portion and about 400 nucleotides in the untranslated regions. The first exon begins with the Cap site, AGC, where transcription starts with a Gppp nucleotide. Exon 1 is only partially translated, beginning with the initial methionine at base pair (bp) 40. The translation termination codon TAA is located in exon 14. Exon 15 is completely untranslated, which contains poly(A) signals, AATAAA, for polyadenylation of the mature mRNA, three in human and one in rat. Poly(A) tails are added to the mRNA about 15 nucleotides after the poly(A) signal. The middle 12 exons from number 2-13 can be grouped to three parts based on their lengths. Each group consists of four exons: exon 2-5, 6-9, and 10-13. The corresponding exons in each group have similar lengths. The three groups of exons encode the three albumin domains respectively (Minghetti et al., 1986; Fig. 1-1 and Fig. 1-3).

3.2 Albumin supergene family

Albumin gene belongs to a supergene family composed of albumin, α-fetoprotein (AFP), vitamin D-binding protein (DBP) and α-albumin (or afamin) genes. The structures of all four proteins are similar and genes for albumin, AFP and DBP have similar organizations, suggesting that they are related, possibly derived from a common ancestor (Gorin et al., 1981a; Cook, 1986; Ray et al., 1991; Minghetti et al., 1986). The four genes are located on the same chromosome, such as chromosome 4 of human (Harper and Dugaiczyk, 1983), chromosome 14 of the rat, chromosome 5 of the mouse, chromosome 8 of the pig, and chromosome 6 of the chick (Peters, 1995). In human, the four genes are linked tandemly on the chromosome: the AFP gene is 15 kb downstream of the albumin gene, the α-albumin gene is 10 kb downstream of the AFP gene, and the DBP gene is about 40 kb downstream of the α-albumin gene (Nishio et al., 1996).
**α-fetoprotein:**

α-fetoprotein (AFP) is the major plasma protein of the fetus and is considered as a fetal counterpart of albumin. It is a 590-amino acid polypeptide with a molecular mass of 70 kD. Its amino acid sequence shows 40% identity to the albumin amino acid sequence. The primary structure of human AFP is strikingly similar to the triple domain structure of albumin (Gorin et al., 1981). Most of the Cys-Cys disulfide bonds, 15 out of 17, are conserved in AFP as in albumin. The missing two disulfide bonds are in loop 6 and loop 9. Identity among the three internal domains of AFP is 18-25%, as it is in albumin. Unlike albumin, AFP is a glycoprotein. Human AFP has a single N-glycosylation site. Asn-Phe-Thr, at residues 232-234 in loop 4, mouse and rat AFPS each have two additional sites. Like albumin, AFP is highly helical, 67% of the residues are in α-helices as determined by CD. Compared to albumin, AFP binds saturated and monounsaturated long-chain fatty acids poorly, but binds arachidonate and other polyunsaturated fatty acids strongly.

The synthesis of AFP and albumin starts at the same time during development parallel to the appearance of the fetal liver. The circulating AFP levels climb to 3 g/L during human fetal life and fall to less than 5 μg/L by 1 year of life, whereas albumin levels rise only during the last months of gestation, reaching nearly the adult level of 40 g/L at birth. The AFP gene has been mapped and sequenced for human (Sakai et al., 1985; Gibbs et al., 1987), mouse (Tilghman, 1985); and rat (Nahon et al. 1987). In human, it lies 15 kb downstream from the albumin gene in chromosome 4 (Urano et al., 1984). The structure of AFP gene is similar to that of albumin gene. It also contains 15 exons in which three internal homologous groups like the albumin gene can be found. The nucleotides of the AFP mRNA are 52% identical to those of the human serum albumin mRNA, although the amino acid residues of the proteins are only 40% identical (Morinaga et al., 1983).
Vitamin D binding protein:

Vitamin D binding protein (DBP) is the major plasma carrier protein for vitamin D and its metabolites. It is a glycoprotein of 52 kD. The plasma concentration of DBP is about 0.5 g/L. The primary structure of DBP shows parallel locations of cysteine residues as that in the albumin molecule, suggesting similar disulfide-bonded loops in the DBP molecule. The striking difference of DBP from other members of this family is the absence of the last two loops, loop 8 and 9 of domain III (Fig.1-4). Amino acid identity between DBP and albumin is 19%, that between DBP and AFP is 16%; these identities are much weaker than the albumin-AFP identity of about 40% (Cooke, 1986). The human DBP gene is located on chromosome 4, at 4q-13 (Bowman and Yang, 1987). It contains 13 exons compared to the 15 in the albumin and AFP gene. Exon size and sequence data suggest that the loss corresponded to exons 12 and 13 of albumin gene (Ray et al., 1991; Witke et al., 1993).

α-albumin:

Recently, a new member of this supergene family has been found (Belanger et al., 1994; Lichenstein et al., 1994). This gene product is termed α-albumin (Belanger et al., 1994) or afamin (Lichenstein et al., 1994). Human α-albumin is a plasma protein with a concentration of 30 mg/L. It is glycosylated and the molecular mass determined by SDS-PAGE is 87 kD. The mature protein has 578 residues and its amino acid sequence shares 36%, 40%, and 21% identity with the human albumin, AFP, and DBP, respectively. The homology is strongly evident in the location of cysteines, implying a three-domain structure similar to other members of the supergene family. Northern hybridization showed that the rat α-albumin gene is selectively expressed in adult rather than fetal liver, and not in yolk sac, brain, or kidney. The rat α-albumin gene has been isolated and partially sequenced by Belanger et al. (1994).
Fig. 1-4 Comparison of the primary structures of vitamin D binding protein and albumin (from Coode, 1986)
3.3 Evolution of albumin supergene family

Fig. 1-5 shows the current evolutionary tree for the albumin/AFP/DBP supergene family. It is hypothesized that this supergene family originated from a protein of about 60 amino acids. Duplications of this small protein gave rise to a precursor of 190-amino acid, which duplicated again to form a three-domain structure, and the gene family was evolved from this ancestor protein. The divergence of albumin and DBP from the ancestor is calculated to have occurred at 560-600 million years (Myr) ago and the albumin/AFP divergence 280 Myr ago. DBP gene lost two exons in the long evolution progress, while the three-domain albumin persisted thereafter. According to this theory, the appearance of AFP was after the amphibian/reptile separation (350 Myr ago). Consequently, amphibians and fishes should lack AFP in their serum at larval or fetal stages. In supporting this theory, albumin has been found in mammals, birds, reptiles, amphibians, teleosts, and lamprey; DBP in teleosts and above but not in lamprey or elasmobranchs: AFP in mammals but not in amphibians (Haefliger et al., 1989; Peters, 1995).

4. Albumin gene expression

4.1 Tissue-specific and developmental expression of the albumin gene

The liver is the major if not the exclusive site for albumin synthesis. In Northern hybridization and nuclear run-on experiments, albumin mRNA was readily detected in liver, but not in other tissues like kidney, brain, spleen, heart, lung, or intestine (Powell et al., 1984; Sell et al., 1985). Albumin protein can be visualized within the cytoplasm of liver cells by albumin antibodies carrying a fluorescent label (Guillouzo et al., 1982). Nahon et al. (1988) found minute quantities of albumin mRNA in nonhepatic tissues like kidney and pancreas in their Northern hybridization. In vitro transcription showed that rat albumin promoter could direct transcription significantly in liver nuclear extracts, to a much lower extent in brain nuclear extracts (1% compare to liver), but not in spleen nuclear
Fig. 1-5 Evolutionary tree of the ALB/AFP/DBP gene family
(from Haefliger, 1989)
extracts at all (Gorski et al., 1986). According to Peters (1995) there has been no evidence that the mRNA in nonhepatic tissues is translated.

During development, the albumin gene and the related AFP gene are activated simultaneously. Both albumin and AFP genes are activated in fetal mouse liver ontogeny and their mRNAs are accumulated similarly in fetal development (Fig. 1-6). mRNAs for both AFP and albumin can be detected between 12 and 14 days in rat and mouse. The AFP mRNA level peaks just before birth; however, after birth the AFP transcription undergoes a dramatic decline, resulting in an approximately 10,000-fold reduction in AFP mRNA levels. Conversely, albumin mRNA increases steadily until after birth and maintains a plateau in adulthood (Tamaoki et al., 1974; Tilghman and Belayew, 1982; Muglia and Locker, 1984; Camper et al., 1989). mRNAs for AFP and albumin are also detectable in the yolk sac and the intestine in early fetal life (Sell et al., 1985; Cooke et al., 1991; Naval et al., 1992). In human, circulating AFP is detectable at 67 mg/L at 6.5 weeks of gestation, rising rapidly to 3 g/L by 10-13 weeks and falling to 10-100 mg/L at term. Circulating albumin is about 1.5 g/L at 6 weeks and reaches adult levels (35 g/L) by 40 weeks.

4.2 Extracellular signals and albumin gene expression

The albumin gene is constitutively expressed in normal adult liver. However, when tested in a cell culture system, the albumin gene is subjected to the regulation of various extracellular signals. In primary cultures of adult rat hepatocytes, transcription of the albumin gene decreased dramatically during culture without the addition of serum and hormones, becoming almost negligible 10 hours after plating. Dexamethasone, a glucocorticoid hormone, prevented this decrease and restored the transcription within 2 hours to the same level as that before culture (Nawa et al., 1986). Retinoic acid (RA) exhibits various potent effects on cell growth and differentiation. The treatment of human hepatoma Hep3B cells with retinoic acid reduced the transcription rate of albumin gene (Hsu et al., 1992). In another work, retinoic acid increased liver-specific AFP and albumin
Fig. 1-6 Ontogeny of mouse AFP and albumin mRNA levels

Concentration of AFP (open circle) and albumin (solid circle) mRNA were determined by dot blot hybridization at different stages before and after birth (B on the horizontal axis). (from Tilghman and Belayer, 1982)
gene expression in McA-RH 8994 cells, a Morris rat hepatoma cell line. Transforming growth factor-beta 1 (TGF-β 1) caused a significant reduction in the level of α-fetoprotein mRNA but not that of albumin mRNA in human hepatoma HuH-7 cells (Nakao et al, 1991). While in the mouse hepatoma cells BWTG3, TGF-β 1 could decrease the steady state mRNA levels of both albumin and α-fetoprotein and this effect was shown to be on the posttranscriptional level and dependent on de novo RNA and protein synthesis (Beauchamp et al., 1992). It has been noticed for many transcriptional regulators that their actions can be different depending on the cellular or genetic contexts (Wan and Wu, 1992).

The synthesis of albumin and albumin mRNA is decreased in diabetic animals. In vitro transcription assay demonstrated that the activity of albumin promoter was lower in diabetic liver nuclear extracts than in controls. Liver nuclear extracts from insulin-treated animals restored the albumin promoter activity. This effect of insulin on albumin promoter might be indirect as suggested by the fact that a prolonged period of treatment (i.e. days) is required for the observed changes in albumin promoter activity. It is possible that insulin acts to correct a metabolic abnormality that in turn regulates expression of the gene (Wanket and Wong, 1991). Using hepatoma cell culture, Messina (1992) has examined the direct effect of insulin on albumin gene expression. It is found that insulin inhibited transcription of the albumin gene in a rapid, dose-dependent manner. Within 15 minutes, albumin transcription was reduced by approximately 80% in rat H4IIE cells. Many other factors such as growth hormone (Johnson et al., 1991), estrogen (Hammer et al., 1986), vasopressin (Chojkier et al., 1989), phenobarbital (Bertaux et al., 1992), and extracellular matrix (Caron 1990) have also been reported to regulate albumin gene expression. The molecular mechanisms by which these different signals may act on albumin transcription are still largely unknown. In a recent study, it is found that the albumin enhancer is not functional in ras- transformed hepatocyte lines. Furthermore, transient overexpression of an activated or normal c-Ha-ras was able to inhibit the activity of the cotransfected albumin enhancer in an immortalized hepatocyte cell line. These results suggested that the ras signaling pathway may be one of
the pathways to transduce the extracellular signals to the albumin gene expression (Hu and Isom, 1994).

5. Regulation of albumin gene expression

5.1 Albumin promoter and enhancer

The expression of the albumin gene is mainly regulated at the transcriptional level. The albumin gene is transcribed by RNA polymerase II (RNA pol II). Promoters of protein-coding genes in eukaryotic cells generally consist of three distinct cis elements: core or basal promoter elements, promoter proximal elements, and distal enhancer elements (Burley and Roeder, 1996). The core promoter elements are defined as "minimal DNA elements that are necessary and sufficient for accurate transcription initiation by RNA pol II in reconstituted cell-free systems" (Roeder, 1996). The most common of these elements are the TATA box (consensus TATAa/tAa/t), located near position -30 to -25 (the transcription start site is defined as position +1), and a pyrimidine-rich initiator (Inr, consensus YYANt/aYY, Y: pyrimidines) located around the transcription start site. Individual core promoters may contain both, one or neither of these elements and are respectively denoted as TATA+Inr+, TATA+Inr-, TATA-Inr+, or TATA-Inr-. Promoter proximal elements are elements located between 50 and 200 bp upstream of the cap site, which can interact with DNA-binding regulatory proteins to modulate transcription. Finally, distal enhancer elements, which can be found far from the transcription start site in either direction and orientation, constitute another group of DNA targets for factors modulating pol II activity. The core promoter elements are conserved in almost all promoters, in almost all organisms, whereas the proximal and distal regulatory elements are highly varied and gene-specific (Roeder, 1996).
Albumin promoter:

The albumin promoter has been studied both in vivo and in vitro and similar results were obtained. In mouse, rat, human and cow, the albumin promoter region is highly conserved within about 250 bp upstream of the transcription start site (Frain et al., 1990; Power et al., 1994). All albumin promoters isolated to date, ranging from human to Xenopus, contain a TATA box and a CCAAT box at -25 and -80 region, respectively. The albumin promoter can direct reporter gene expression in differentiated hepatoma cells which express endogenous albumin, but not in non-hepatic cells or dedifferentiated hepatoma cells (Ott et al., 1984; Tronche et al., 1990). Hepatoma cell lines commonly used in albumin promoter studies are rat H4II cells (Heard et al., 1987; Herbomel et al., 1989; Tronche et al., 1989), human HepG2 cells (Friedman et al., 1986; Babiss et al., 1986, 1987), human Hep3B cells (Frain et al., 1990), and human HuH-7 cells (Hayashi et al., 1992). Rat albumin 5' flanking region up to -2000 bp has been tested in a transient expression system (Heard et al., 1987). Progressive deletions from the 5' end showed that sequences between -2000bp to -150bp were dispensable for the tissue-specific promoter activity. The 150 bp region proceeding the transcription start site contained all the critical elements necessary for the tissue-specific expression of rat albumin promoter in hepatoma cells. In an in vitro transcription system, Gorski et al. (1986) obtained similar results from the mouse albumin promoter, which was efficiently used only in liver nuclear extracts. Transcriptions in other tissue or cell nuclear extracts were much reduced (100-fold lower in brain nuclear extracts) or barely detectable (in spleen and Hela nuclear extracts). The sequences between -170 and -55 of mouse albumin gene were required for this liver-specific in vitro transcription.

The albumin promoters were further analyzed by footprinting and gel mobility shift. A salient feature of albumin promoters was that five (in rat) or six (in mouse) cis elements are compacted within 130 bp upstream of the TATA box (Cereghini et al., 1987; Lichtsteiner et al., 1987; Izban and Papaconstantinou, 1989). Functional analysis in vivo and in vitro with deleted and mutated promoters showed that all these cis elements were
Fig. 1-7 Schematic representation of the *cis* elements of albumin promoter and *trans* factors interacting with these *cis* elements.
positive regulatory elements (Herbomel et al., 1989; Tronche et al., 1989; Maire et al., 1989). The five cis elements in the rat albumin promoter were designated as PE (proximal element), CCAAT, DEI, DEII, DEIII (DE stands for distal element) from 3' to 5' (Fig. 1-7). The six elements in mouse albumin promoter were designated as A to F, and elements B to F corresponded to the five elements in rat. Trans-acting factors that interact with these cis elements include liver-enriched factors HNF1 (binds on PE), C/EBP (on DEI and DEIII), DBP (on DEI), and ubiquitous ones like NF-Y (on CCAAT box) and NF-1 (on DEII) (Babiss et al., 1987; Lichtsteiner et al. 1987; Cereghini et al., 1987, 1988; Raymondjean et al., 1988; Lichtsteiner and Schibler, 1989). Element A in mouse is a weak C/EBP binding site between the site B and the TATA box, which does not have a corresponding site in rat. Mutation on this site had little effect on the promoter activity (Maire et al., 1989). A hierarchy of the different cis elements for albumin promoter activity has been established by in vivo and in vitro analyses (Herbomel et al., 1989; Tronche et al., 1989; Maire et al., 1989). The HNF-1 binding site PE strongly activates transcription in a tissue-specific manner, next is the C/EBP binding site DEI. The NF-Y binding site CCAAT box has a lower activation potential and is less specific, being equally efficient in liver and spleen nuclear extracts. The remaining elements are relatively weak activator sites.

**Albumin enhancers**

The enhancer of mouse albumin gene is located far upstream (-10.5 kb to -8.5 kb) of the transcription start site as detected in transgenic mice (Pinkert et al., 1987). In transgenic mice, chimeric genes with mouse albumin 5'-flanking sequences fused to a human growth hormone (hGH) reporter gene were expressed only in liver. Up to 12 kb of the mouse albumin 5'-flanking region have been tested. The promoter alone (0.3 kb upstream of the cap site) directed marginal expression of the reporter gene. Remarkably, a region located at -10.4 kb to -8.5 kb was essential for high-level expression in transgenic mouse liver, which increased the albumin promoter activity by about 50 fold in a position-
Fig. 1-8 Schematic representation of transcription factors interacting with mouse albumin enhancer. Region I, II, and III are the regulatory, positive, and negative region respectively.

- C/EBP (binds on eE);
- HNF3 (binds on eG)
- HNF3 and NF1 (binds on eH)
- ANF (binds on negative region)
and orientation-independent manner. The sequences between -8.5 kb and -0.3 kb was dispensable. In transgenic mice, the mouse albumin enhancer did not function on a heterologous promoter, the human growth hormone promoter. The mouse albumin enhancer region was further analyzed in a transient expression system and it could stimulate the albumin gene expression by about 5 fold in hepatoma cells, but not in Hela cells. It was also functional on the heterologous mouse transthyretin promoter (Herbst et al., 1989, 1990). Three functional regions in the enhancer have been defined: a positive region (-10.03 kb to -9.94 kb), a negative region (-8.7 to -8.43), and a regulatory region (-10.20 kb to -10.11 kb) (Fig. 1-8). The positive region by itself could enhance the promoter activity in a cell-specific and orientation-independent manner. However, this positive effect was abolished by the downstream negative region. The negative region could also block the mouse transthyretin enhancer from activating its promoter, but it had no direct effect on the mouse albumin promoter by itself, suggesting that the mechanisms of action of this element was to prevent the action of distant positive acting sequences. The upstream regulatory region by itself was inactive but in conjunction with the positive region it could override the effect of the negative region (Herbst et al., 1989). DNase I footprint analysis of the regulatory and positive regions revealed 11 liver nuclear protein binding sites, designated eA to eK. Among them, site eE, eG, and eH were essential for full enhancer activity (Zaret et al., 1990; Liu et al., 1991). eE was in the regulatory region and binds C/EBP. eG and eH were in the positive region, eG bound the liver-enriched transcription factor HNF3 and eH bound both HNF3 and an ubiquitous transcription factor NF1 (Zaret et al., 1990; Jackson et al., 1993). The negative region contained two DNA-binding sites for an ubiquitous negative factor ANF (stands for albumin negative factor) (Herbst et al., 1990).

In the human albumin gene, there are three enhancers within 12.5 kb of the 5' flanking region. One enhancer is located in the proximal region between -486 bp and -221 bp, the other two are located in the far upstream region at -1.7 kb and -6 kb. The proximal enhancer provided a 5-fold stimulation of albumin promoter in human hepatoma cells.
Hep3B (Frain et al., 1990), whereas the two upstream enhancers together increased the reporter gene expression by 10-20 fold in another human hepatoma cells HuH-7 (Hayashi et al., 1992). Protein-binding studies demonstrated that the liver-enriched transcription factor HNF1 bound to the proximal enhancer and the upstream enhancer at -1.7 kb. The other upstream enhancer at -6 kb contained a GT-rich element TGGTTTG. Mutation of this GT-rich element resulted in the loss of transcriptional activity as well as binding to the HuH-7 nuclear protein. The same GT-rich element was also present in the mouse albumin enhancer, which bound to a liver-specific transcription factor eH-TF (Zerat et al., 1990), suggesting that similar mechanisms may be involved in the mouse albumin enhancer and the human albumin enhancer at -6 kb (Hayashi et al., 1992).

5.2 Transcription factors involved in albumin gene expression

5.2.1 Transcription initiation and activation

Eukaryotic cells contain three distinct RNA polymerases to transcribe different set of genes: RNA polymerase I (RNAPI), which transcribes ribosomal RNA (rRNA); RNA polymerase II, which synthesizes messenger RNA (mRNA); and RNA polymerase III, which synthesizes 5S rRNA and tRNA. The initiation of mRNA transcription needs the RNA polymerase II and a set of general transcription factors (GTFs) to assemble on the core promoter to form a pre-initiation complex (Fig. 1-9). The general transcription factors characterized so far include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIF. TFIID consists of a TATA-binding protein (TBP) and certain TBP-associated factors (TAFs), and it is the only general transcription factor with site-specific DNA binding ability. The first step of the formation of pre-initiation complexes is TBP binding to the TATA element, which is stabilized by TFIIA. The DNA-TFIID complex is then recognized and bound by TFIIB which serves as a bridge for the binding of RNA pol II and determines the distance between TATA element and the transcription start site. RNA pol II enters the complex in association with TFIIF. Finally, TFIIE and TFIIF enter the complex and the pre-initiation
Fig. 1-9 Schematic model of transcription initiation complex assembly

The bold line represents the promoter DNA while the bent arrow indicates the initiation site. Double-headed arrows indicate protein-protein interactions. Transcription factors are represented by their letter designations. pol, RNA polymerase II; B, D, E, F, H, general transcription factors TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. (from Buratowski, 1994)
complex assembly is finished. In the presence of ATP, the pre-initiation complex is activated and transcription is initiated (Conaway and Conaway, 1993; Corden, 1993; Maldonado and Reinberg, 1995). TFIIH has protein kinase and DNA helicase activities that may be required for the transition from assembly of the pre-initiation complex to the initiation and promoter-clearance events (Seroz et al., 1995).

The mRNA transcription is activated by DNA-binding transcription activator proteins that specifically recognize upstream promoter or enhancer elements. Eukaryotic transcription activators usually contain two functional domains: the DNA-binding domain and the transcription activation domain. The structure of many DNA-binding domains and the details of their interaction with the target site have been studied thoroughly in the past 10 years (reviewed in Nelson, 1995). Four structural motifs have been described for the DNA-binding domains of various transcriptional activators, they are: homeodomain, zinc finger, basic leucine zipper, and basic helix-loop-helix (Mitchell and Tjian, 1989; Struhl, 1995; Nelson, 1995). The activation domains are usually classified on the basis of their amino acid compositions, such as acidic activators, glutamine-rich, proline-rich, or serine- and threonine-rich activators (Mitchell and Tjian, 1989; Triezenberg, 1995). Many transcription factors contain domains that mediate the formation of homo- and heterodimers. Leucine zipper and helix-loop-helix motifs are typically found in dimerization domains (Lamb and McKnight, 1991). Transcriptional activators can interact with components in the preinitiation complex (such as TBP, TFIIIB, TFIIH, and TFIIA) or with some intermediary factors (such as TAFs and PC4) which can co-activate transcription or transduce the signal from the activation domain to a component of the basic machinery. By these protein-protein interactions, transcriptional activators can increase the recruitment of TGFs into the pre-initiation complex (Ptashne, 1988; Triezenberg, 1995; Bjorklund and Kim, 1996; Verrijzer and Tjian, 1996; Kaiser and Meisterernst, 1996). Another mechanism for activator's action is that binding of one factor to DNA can change local chromatin structure so that the access by additional factors is facilitated (Pina et al., 1990; Archer et
al., 1991; McPherson et al., 1993). The third mechanism involves the interaction between neighboring DNA binding regulatory factors. These types of interactions could modulate not only the activating potential of a factor, but also its DNA binding affinity and its interactions with non-DNA binding proteins (Milos and Zaret, 1992; Wright et al., 1994).

### 5.2.2 Transcription factors involved in albumin gene transcription

Albumin gene expression is mainly regulated at the transcriptional level. As described above, the albumin gene has a strong liver-specific promoter and enhancer, thus it provides a good model for the study of liver-specific transcription. Structural and functional studies indicated that the tissue specificity of albumin gene transcription is achieved by a combination of both liver-enriched and ubiquitous transcription factors binding on the proximal promoter and the upstream enhancer. Four families of liver-enriched transcription factors (HNF1, HNF3, HNF4, and C/EBP) have been identified so far to regulate the liver-specific transcription and three of them (HNF1, HNF3, and C/EBP) are involved in the control of albumin gene expression (Lichtsteiner and Schibler, 1989; Friedman et al., 1989; Jackson et al., 1993; Cereghini, 1996). Ubiquitous factors NFY and NF1 are also important in maintaining high levels of transcriptional activity of the albumin gene (Mantovani et al., 1992; Cereghini et al., 1987; Hu and Isom, 1994). These transcription factors work autonomously or synergistically to activate the transcription from the albumin gene promoter.

**HNF1**

HNF1 (also named as LFB1, HNF1α, or APF) is a liver-enriched trans-activator involved in the regulation of many liver-specific genes, such as albumin, β-fibrinogen, and α1-antitrypsin. HNF1 is a 628-amino acid glycoprotein with a molecular mass of 88 kD (Fig. 1-10) (Lichtsteiner and Schibler, 1989; Baumbueter et al., 1990). The cloning of HNF1 cDNA revealed that it is a homeodomain-containing factor (Frain et al., 1989). The
Fig. 1-10 Conserved domains of the HNF1 transcription factor family

Schematic representation of the structure of HNF1 and vHNF1-A and vHNF1-B proteins. Highly conserved sequences within the DNA binding domains of HNFα and vHNF1 proteins include the amino-terminal dimerization domain (domain A), POU domain (domain B), and homeodomain (domain C). (from Cereghini, 1996).
homeodomain was first identified in *Drosophila* homeotic proteins involved in the regulation of *Drosophila* embryogenesis. It was soon found in the DNA-binding domains of many transcription factors in vertebrates. Homeodomains consist of about 60 amino acids which form four α-helices with the helix II and III making a helix-turn-helix motif for binding DNA. HNF1 has a variant homeodomain which contains 80 amino acids and 21 residues loop out between the helix II and the helix III (Ceska et al., 1993; Gehring et al., 1994). Upstream of the homeodomain, a POU-like domain constitutes another part of the DNA-binding domain of HNF1. The 90-amino acid POU-like domain can increase the specificity of the DNA recognition without significantly changing the affinity of DNA-protein interaction (De Simone and Cortese, 1992). HNF1 binds DNA as a dimer and the dimerization domain is located on the N-terminal 34 amino acids to form a coiled-coil structure. The HNF1 dimer is stabilized by a cellular co-factor DCoH (Dimerization Cofactor of HNF1) (Nicosia et al., 1990; Mendel et al., 1991; Hansen and Crabtree, 1993). Three activation domains have been identified on the C-terminal half of the molecule. They are rich in proline, proline and glutamine, and serine, respectively (reviewed in Tronche and Yaniv, 1992).

The HNF1 binding site is commonly found in the proximal promoter region of genes it regulates. However, HNF1 binding sites that are important for transcriptional activity are also located in the enhancers of certain genes. The consensus HNF1 binding sequence derived from different HNF1 binding sites is: GGTAAATNATTAAC<sup>a/c</sup>. This sequence shows a palindromic structure, consistent with the fact that the HNF1 protein binds DNA as a dimer. The second half site of the palindrome is less conserved than the first half, which might confer different affinity of HNF1 for its targets (Tronche and Yaniv, 1992).

A variant HNF1 (vHNF1) that binds on the same DNA element as HNF1 has been identified. The vHNF1 (also known as HNF1β and LFB3) is a smaller protein and has a highly conserved DNA-binding domain and dimerization domain as HNF1 (about 93% and
75% identical to the responsive domains in HNF1 respectively). The carboxy-terminal region of vHNF1, which encompasses the activation domain, is shorter than HNF1 in the same region and does not show much similarity with HNF1. HNF1 and vHNF1 are able to form heterodimers with each other (Tronche and Yaniv, 1992; Cereghini, 1996). The expression of HNF1 and vHNF1 is not restricted to hepatocytes. HNF1 mRNA is equally abundant in liver, kidney, and intestine. It is also expressed in stomach and pancreas (Blumenfeld et al., 1991). vHNF1 is more abundant in kidney than in liver, intestine and stomach. vHNF1 is also expressed in tissues that lack HNF1, including the thymus, testis, ovary, and lung. During mammalian development, vHNF1 expression systematically precedes that of HNF1, suggesting that they play a different role in hepatocyte differentiation. Perhaps vHNF1 is involved in the onset and HNF1 in the maintenance of the expression of tissue-specific genes (Cereghini et al., 1992; Lazzaro et al., 1992).

HNF1 plays a dominant role in the albumin promoter activity. HNF1 binds on the PE (site B) of the albumin promoter (Fig.1-7) which contains a sequence of 5' TAGTTAA-TAATCTACA 3'. In a transfection assay, the PE site, along with the TATA box, is sufficient to drive transcription in a tissue-specific manner. When PE is impaired by methylation or mutation, the binding affinity for HNF1 is reduced and cooperation with factors binding on upstream elements is required to restore the promoter activity (Tronche et al., 1989). When tested in vitro, mutation on the PE site has stronger effect on the transcription than mutations in other elements, reducing the transcription efficiency ten-fold in liver but not in spleen nuclear extract. Multiple PEs strongly stimulate transcription in liver nuclear extracts, but not in spleen nuclear extracts (Maire et al., 1989). The PE can activate in vitro transcription in spleen nuclear extracts when purified HNF1 is added to the system (Lichtsteiner and Schibler, 1989). HNF1-binding sites are also found in human albumin enhancers, both in the proximal enhancer (-486 to -221) and in the upstream enhancer at -1.7 kb. These two binding sites can increase transcription from the albumin promoter in a dose-dependent manner.
The albumin gene is expressed in differentiated hepatoma cells but not in dedifferentiated hepatoma cells, which parallels the expression of HNF1. H4II and Fao are highly differentiated rat hepatoma cell lines. H5 and C2 are dedifferentiated derivatives of H4II and Fao, which no longer exhibit liver-specific functions such as albumin synthesis. In gel mobility shift assay, a complex formed between albumin PE and HNF1 is detected in H4II and Fao nuclear extracts, but not in H5 and C2 nuclear extracts. In H5 and C2 nuclear extracts, a smaller complex is formed on the albumin PE by binding of vHNF1. Interestingly, C2rev7, a differentiated revertant cell line derived from C2 cells, shows the disappearance of the PE-vHNF1 complex and the reappearance of the PE-HNF1 complex. This phenomenon is also seen in somatic hybrid cells. HF1 is a somatic hybrid between Fao and H5 cells, showing extinction of liver-specific function. HF1-5 is a subclone derived from HF1 in which the previously extinguished functions are re-expressed. The HF1 cells have the same DNA-protein complex in native polyacrylamide gels as H5 and C2 cells have, while the HF1-5 cells show the same pattern as H4II and Fao cells (Cereghini et al., 1988). Northern hybridization and nuclear run-on assays detect HNF1 mRNA in all differentiated hepatoma cells, but not in dedifferentiated hepatoma cells. Contrary to HNF1, the mRNAs coding for two other nuclear factors involved in albumin transcription, C/EBP and NF1, do not follow the distribution of albumin transcripts in these cell lines. These results indicate that the inactivity of the albumin promoter upon both dedifferentiation and extinction is related to the block of HNF1 mRNA synthesis (Cereghini et al., 1990).

HepG2 is a human hepatoma cell line of fetal phenotype, which expresses both α-fetoprotein and serum albumin. Expression of reporter genes driven by albumin promoter in HepG2 cells is low compared to that in highly differentiated hepatoma cells. HNF1 and vHNF1 homodimers and heterodimers are both present in HepG2 cells. Cotransfection of an HNF1 expression vector stimulates transcription from exogenous albumin promoters significantly in HepG2 cells (Rollier et al., 1993). Albumin gene transcription is reduced in diabetes mellitus. It seems that the repressive effects of diabetes mellitus are mediated by
HNF1, since the abundance and binding activity of HNF1 on the albumin promoter is reduced (Barrera-Hernandez et al., 1996). In a recent study, the HNF1 gene was inactivated by homologous recombination and the HNF1-deficient mice fail to thrive and die around weaning period. The function of liver and kidney is impaired in mutant mice. As targets of HNF1, the transcription rate of albumin gene is reduced and the phenylalanine hydroxyase (PAH) gene is totally silent (Pontoglio et al., 1996).

HNF1 acts on albumin promoter autonomously or synergistically with other transcription factors. As shown above, PE alone is able to activate transcription in vivo and in vitro. C/EBP is another liver-enriched transcription factor which functions on the albumin proximal promoter. Cotransfection of HNF1 or C/EBPα expression vectors with the albumin promoter constructs stimulate transcription in Hela cells. When cotransfecting with both HNF1 and C/EBPα expression vectors, the activity of the albumin promoter is approximately 10 fold higher than the sum of the activities achieved with HNF1 and C/EBPα alone. Structural and functional analyses showed that this synergistic activation is mediated by an activation domain located at amino acids from 351 to 507 in HNF1, which is rich in proline and glutamine residues (Wu et al., 1994).

**C/EBP and DBP**

C/EBP is a family of transcription factors which recognize the same DNA element and can form homo- and heterodimers with each other (Sladek and Darnell, 1992; De Simone and Cortese, 1992; Xanthopoulos and Mirkovitch, 1993). Since it was originally shown to bind to two distinct sites: the CCAAT box of several promoters and the core motif TGTGG(A/T)(A/T)G of several viral enhancers, it was termed as CAAT/enhancer binding protein (C/EBP) (Landschulz et al., 1988). Now, the optimal C/EBP binding site is artificially designed as a palindromic sequence: GATTGC CGCAATC (Wedel and Ziegler-Herrbrock, 1995). However, natural C/EBP binding sites are very degenerate and usually homology to only one half-site is found. C/EBP proteins contain a highly conserved DNA
Fig. 1-11 Functional domains of C/EBPα (from Cereghini, 1996) and model of the C/EBP dimer via leucine zipper (from Wedel and Ziegler-Heitbrock, 1995).
binding domain located at the carboxy terminal of the polypeptide, whereas their activation
domains at the amino terminal are divergent among different members (Fig. 1-11). The
DNA binding domain is a basic leucine zipper (bZip) domain consisting of two regions: a
basic region of about 30 amino acids involved in DNA recognition and an immediately
adjacent leucine zipper region that mediates dimerization. The leucine zipper region contains
a heptad repeat of leucines within a 35-residue sequence which forms an amphipathic α-
helix with the leucines being located on one side every two turns. Leucines from two
monomers interdigitate like a zipper to form a Y-shaped dimer. The leucine zipper is on the
base of the Y, and the basic regions form two arms that make contacts with DNA in the
major groove (Alber, 1992). Dimerization is a widespread feature of transcriptional
regulators. Heterotopic dimerization contributes to the complexity of transcription
regulation by using limited transcription factors in a cell, since heterodimers can change
DNA binding affinity or specificity and can interact with more targets in the transcriptional
machinery (Lamb and McKnight, 1991).

To date, at least five distinct members of C/EBP family (C/EBPα, β, γ, δ, and
CHOP-10) have been identified and cloned. Most of the C/EBP proteins act as
transactivators. C/EBPα is a transactivator first isolated in rat liver nuclei (Landschulz et al.
1988). It is a 42 kD heat-stable protein which plays an important role in the transcriptional
regulation of several liver-specific and adipose-specific genes. C/EBPα transcripts are
present in a variety of cells, but the protein is detected only in differentiated hepatocytes,
adipocytes, intestinal epithelial cells, pregranulocyte, and myeloblastic cell lines
(Johnson and Williams, 1994). C/EBPα is speculated to play a role in establishing and
maintaining the differentiated, nonproliferative state of adult hepatocytes and adipocytes
(Umek et al., 1991; Lin and Lane, 1992). C/EBPβ has been isolated from rat (named as
NF-IL6, NF-M), human (IL-6DBP, CRP2), mouse (GPE-BP, AGP/EBP) and Chicken
(NF-M) (Wedel and Ziegler-Hettbrock, 1995). The C/EBPβ gene is expressed in most
tissues, however, its protein accumulates preferentially in liver. C/EBPβ can be
phosphorylated by different kinds of kinases coupling different signaling pathways, which results in inactivation of an inhibitory domain of the protein (Kowenz-Leutz et al., 1994). The active form of C/EBPβ has been implicated in the induction of genes involved in the acute-phase response, cytokine expression, and inflammation. C/EBPγ and C/EBPδ are other two members of the family. They are present at low concentrations in the normal liver and no significant contributions to liver-specific transcription have been observed for these two factors (Cereghini 1996). C/EBPδ is heavily induced during the acute-phase response (Flodby et al., 1993).

C/EBP proteins are also found to inhibit transcriptions. LIP (liver inhibitory protein) is a N-terminal truncated form of C/EBPβ, which loses the activation domain. It is able to heterodimerize with other bZip proteins and blocks the activation effects of C/EBPβ (Descombes and Schibler, 1991). The same effect is also reported for the N-terminal truncated form of C/EBPα (30 kD) (Ossipow et al., 1993). CHOP-10 (C/EBP-homologous protein) is another transcription inhibitor in the C/EBP family. This protein has a homologous bZip domain as other C/EBP proteins. Two substitutions occur at two highly conserved positions in its basic region, thus heterodimers between CHOP-10 and other C/EBP proteins are unable to bind DNA and the activation function is attenuated (Ron and Habener, 1992).

DBP (site D binding protein on albumin promoter) is another bZip transcription activator which binds to C/EBP site, but it can not form heterodimers with any of the known C/EBP proteins. DBP seems to bind in a mutually exclusive manner to the albumin promoter competing for binding with C/EBPα and C/EBPβ. DBP transcripts are detected at similar levels in most adult tissues, however, the levels of protein are highest in liver and moderate in the lung, kidney, brain, and pituitary. A remarkable property of the DBP expression is that it follows a stringent circadian rhythm: the level of DBP is almost undetectable in the morning and peaks at approximately 8 pm (Mueller et al., 1990; Wuarin and Schibler, 1990; Lavery and Schibler, 1993).
Binding sites for C/EBP have been found in both albumin promoter and enhancer. In the proximal promoter of the albumin gene, C/EBP bound to the DEI (site D) and activates transcription. Its effect is more obvious when the HNF1 binding site PE is impaired by methylation or mutation (Herbomel et al., 1989; Tronche et al., 1989). Mutation on DEI reduces the transcription by 3-fold in liver nuclear extracts. Multiple DEI sites highly stimulate in vitro transcription in liver nuclear extracts but not in spleen nuclear extracts, implying its involvement in the regulation of albumin gene tissue-specific expression (Maire et al., 1989). In the partially differentiated hepatoma HepG2 cells, the C/EBP level is about 20 times lower than liver cells as detected by Western blotting. Cotransfection of the C/EBPα expression vector and albumin promoter constructs increases transcription significantly in HepG2 cells (Friedman et al., 1989). The mouse albumin gene has an enhancer located about 10 kb upstream of the promoter, which contains three distinct functional regions: a positive and a negative region and a regulatory region. Footprinting and gel mobility shift assay have defined two C/EBP binding sites in the regulatory region (Herbst et al., 1989), while the positive region contains binding sites for transcription activators HNF3 and NF1 (Zaret et al., 1990; Liu et al., 1991; DiPersio et al., 1991; Jackson et al., 1993) and the negative region contains binding sites for ANF (albumin negative factor) (Herbst et al., 1990). It is proposed that the ubiquitous negative factor ANF (albumin negative factor) acts in nonhepatic cells as a general suppressor of the neighboring enhancer function. In hepatocytes, C/EBP assists factors (HNF3 and NF1) bound on the positive region to overcome the negative effects, but itself does not stimulate transcription directly (the regulatory region by itself is not capable of enhancement) (Herbst et al., 1989, 1990; Liu et al., 1991). Thus C/EBP can activate albumin gene expression by direct enhancement of the promoter activity or by suppression of the action of an ubiquitous negative factor ANF in hepatic cells.

C/EBPα contains three elements in its N-terminal trans-activation domain. These elements have little or no ability to activate transcription from albumin promoter by
themselves, but any two can cooperate to do so. C/EBPα also contains a negative regulatory region, the function of which is alleviated when C/EBPα is bound in the environment of the albumin promoter. It is speculated that cooperation between C/EBP and other factors bound to adjacent sites on the albumin promoter would alleviate the activity of the negative regulatory region (Friedman and McKnight, 1990; Nerlov and Ziff, 1994). Milos and Zaret (1992) showed that NF-Y was required for C/EBP to form stable complex on the albumin promoter and binding of C/EBP and NF-Y synergistically activate albumin promoter. Whether NF-Y functions to modulate the activity of the negative region of C/EBPα needs to be determined.

**NF-Y**

NF-Y (also referred to as CBF, CP1 and YEBP) is an ubiquitous transcription factor that was originally found to recognize the Y box of major histocompatibility complex (MHC) class II genes. It is a CCAAT box binding protein critical for the expression of diverse eukaryotic genes. NF-Y is composed of two subunits (NF-YA and NF-YB), both of which are required for DNA binding. Direct evidence for the involvement of NF-Y in the regulation of albumin promoter activity comes from the study of Mantovani et al. (1992), in which monoclonal antibodies to NF-Y can inhibit in vitro transcription from the albumin promoter. The antibodies also form supershift bands on the albumin CCAAT-box probe in a gel mobility shift assay. In transfected H4II hepatoma cells, the strength of the rat albumin promoter is mainly dependent on the integrity of the HNF1 binding site (PE or site B) and the TATA box. The presence of the CCAAT element located at position -80 increases promoter activity 2 to 3 fold. Sequences further upstream have no effect. However, when the PE is methylated, its affinity for HNF1 is reduced and the distal sequences (containing binding sites for NF-Y, C/EBP, and NF1) of the promoter become essential to restore the promoter activity and tissue specificity. Among the upstream elements, the CCAAT-box plays an important role since mutation on this site reduced the
promoter activity 50 to 100-fold (Herbomel et al., 1989; Tronche et al., 1989). Multiple NF-Y binding sites enhance in vitro transcription from the albumin promoter in both liver and spleen nuclear extracts, suggesting its action is not tissue-specific (Maire et al., 1989). NF-Y is postulated to be essential for albumin gene expression in early development when liver-enriched transcription factors such as HNF1 and C/EBP are limiting. During development, the NF-Y concentration in liver extract from 14-day mouse embryos, a stage of development when fetal hepatocytes efficiently transcribe the albumin gene, is similar as that in adult liver and in H4II and H5 hepatoma cells. By contrast, the relative concentration of HNF1 was lower in embryo liver extract than in adult. In 11.5-day embryos, the NF-Y binding activity on albumin CCAAT box is similar to that seen in the adult, whereas no HNF1 binding activity was detected. It is possible that the ubiquitous transcription factor NF-Y may potentiate the effect of the tissue-specific factor HNF1 which is present in low amount, resulting in a strong tissue-specific promoter (Tronche et al., 1991).

**NF1**

Nuclear factor 1 (NF1) represents a family of sequence-specific DNA-binding proteins which modulate both transcription and adenovirus DNA replication. NF1 proteins have highly conserved N-terminal DNA binding domain and variable C-terminal activation domain. The DNA-binding domain of NF1 does not show obvious features associated with well-characterized DNA-binding domains. NF1 binds to a palindromic site or a half palindromic site (TGGCA) as a dimer (Gounari et al., 1990). NF1 binding sites have been found in both the albumin promoter and the enhancer. It binds to the DEI1 (or site E) of the promoter (Cereghini et al., 1987). When the HNF1-binding site is impaired, DEI1 is required for full promoter activity, although its contribution to the strength and tissue specificity of albumin promoter is much less than DEI and CCAAT box (binding sites for C/EBP and NF-Y) described above (Herbomel et al., 1989; Tronche et al., 1989; Maire et al., 1989). In the mouse albumin enhancer, NF1 binds to a site close to the HNF3-binding
site and activates transcription cooperatively with other factors bound in the enhancer (Jackson et al., 1993).

**HNF3**

HNF3 is another liver-enriched transcription factor involved in albumin gene tissue-specific expression. HNF3 factor binds to the eG and eH sites of the mouse albumin enhancer and plays essential roles for the enhancer activity (DiPersio et al., 1991; Liu et al., 1991; Jackson et al., 1993). In the case of human albumin upstream enhancer at -6 kb, a GT-rich element (TGTTTGGC) closely related to the HNF3 binding sequence (TGTGTGC) is present, which forms DNA-protein complex with a component in the hepatoma cell HuH-7 nuclear extracts and is capable to enhance in vivo transcriptions (Hayashi et al., 1992).

HNF3 proteins (HNF3α, β, and γ) consist of a novel family of transcription factors that share a highly conserved DNA binding domain and bind DNA as monomers. The DNA binding motif displays a striking similarity to a region of the already known *Drosophila* homeotic gene forkhead (fkh), a gene required for the formation of terminal structures that give rise to the anterior and posterior gut (Cereghini, 1996). HNF3 proteins use a 100-amino-acid winged helix motif, a modified helix-turn-helix motif, for monomeric recognition of specific DNA target sites (Nelson, 1995). The consensus recognition sequence for HNF3 is: A(A/T)TRTT(G/T)RYTY (R: purines; Y: pyrimidines). HNF3 factors are primarily associated with liver- and lung-specific gene expression in the adult. Evidence also suggests that HNF3 factors participate in cellular differentiation and embryogenic pattern formation (Clevidence et al., 1994).

**6. Objectives of the current study**

Our knowledge about albumin mainly has come from the study of mammalian systems. In contrast, the study of albumin in fish, both at the protein level and at the gene
expression level, is very limited. For a long time, the existence of albumin in fish was
debatable since there is no protein present in fish serum to the extent observed in
mammalian serum (Perrier et al., 1973). In 1988, Davidson et al. (1988) detected fatty
acid-binding proteins in the serum of several salmonid species, which were thought to be
fish albumin. As a model gene, the study of regulation of albumin gene expression has
contributed significantly to our knowledge of the liver-specific gene expression in
mammals. Whereas the mechanisms governing the liver-specific gene expression in fish are
largely unknown. Thus it will be of special significance in the study of fish molecular
biology to clone a fish albumin gene and to investigate its expression and regulation. The
albumin gene promoter is also a useful tool in gene transfer study, which can direct a
foreign gene to express in the liver (Kuriyama et al., 1991). Our laboratory is interested in
antifreeze protein gene transfer into a salmonids to confer freeze tolerance. The expression
of the antifreeze protein transgene is very low in recipient fish due to its weak promoter
(Hew, 1989). The use of a strong fish liver promoter or enhancer is an attractive
alternative, which may increase the expression of antifreeze protein genes in transgenic fish
liver. The albumin gene is a good candidate for this purpose. Using rainbow trout which is
a tetraploid, we are also interested in whether both albumin genes are differentially
regulated. The current study is designed to clone the rainbow trout albumin cDNA and
gene, analyze the tissue specificity and developmental stages of its expression, and
characterize the promoter and enhancer which contribute to the tissue-specific expression.
The specific experimental approaches are as follows:

Part 1. Isolation and characterization of two albumin cDNAs in rainbow trout
1. Construction of rainbow trout liver cDNA library
2. Isolation of two trout albumin cDNAs by cDNA library screening and reverse PCR
3. DNA sequencing of the two trout albumin cDNAs
4. The tissue-specific and developmental expression of trout albumin gene

5. Comparison of the transcription rates of two trout albumin genes by nuclear run-on experiments

**Part 2. Isolation and characterization of two rainbow trout albumin gene promoters**

1. Construction of rainbow trout genomic DNA library and isolation of the albumin genes
2. Characterization of the albumin genes by genomic restriction mapping and DNA sequencing
3. Determination of the transcription start site by primer extension
4. Study of the expression pattern of trout albumin 2 promoter-CAT fusion genes by transient expression assay in human hepatoma cells
5. Isolation of the trout albumin 1 promoter by PCR and comparison of the activities of the two trout albumin promoters
6. Mutation analysis of the trout albumin 2 promoter
Chapter Two

Isolation and Characterization of Two Distinct Rainbow Trout Albumin cDNAs
INTRODUCTION

In mammals, albumin is identified as the most abundant serum protein with a molecular mass of about 68 kD. Due to its low level, the existence of albumin in rainbow trout was once questioned when the electrophoresis patterns of rainbow trout and human plasma proteins were compared (Perrier et al., 1973). Davidson et al. (1988) subsequently used the fatty acid binding properties of albumin to characterize the plasma proteins in salmonid fish and found they contained palmitate-binding proteins with a molecular mass similar to that of human serum albumin. These proteins were designated as albumin-like proteins. In 1990, Bynes and Gannon (1990) isolated the first fish albumin cDNA clone from the Atlantic salmon liver cDNA library. In order to isolate the rainbow trout albumin gene and characterize its transcription regulation, we set out first to clone the rainbow trout albumin cDNA. Rainbow trout is a tetraploid and has two genes for almost every protein. In this chapter, we report the isolation of two distinct rainbow trout albumin cDNA clones: rtALB1 and rtALB2. The tissue-specific and developmental expression of the trout albumin gene was examined by Northern blot. To understand whether these two albumin genes are regulated by different mechanisms, the transcriptional rates of the two trout albumin genes were compared by nuclear run-on experiments.
MATERIALS AND METHODS

1. Tissue collections

Rainbow trout (*Oncorhynchus mykiss*) (Coolwater Farm, Pickering, Ontario, Canada) were sacrificed and different tissues, i.e. eye, gill, heart, intestine, kidney, liver, muscle, spleen and stomach, were collected and immediately frozen in liquid nitrogen or dry ice and stored at -70°C. Fertilized rainbow trout embryos of 5 and 6 weeks were frozen at -70°C. The hatched fry's of 8 weeks were cut into three segments (Fig. 2-11): head, abdomen (with attached yolk sac) and tail, each segment was frozen separately until use. For nuclei preparation in nuclear run-on experiments, fresh liver was used.

2. cDNA library construction and screening
A. cDNA library construction

Rainbow trout liver poly(A)+RNA for cDNA library construction was purified by using the Micro-Fast Track mRNA Isolation Kit (Invitrogen). The cDNA library was constructed in Lambda ZAP II vector using the ZAP-cDNA Synthesis Kit (Stratagene). In brief, rainbow trout liver poly(A)+RNA was annealed to poly dT primers and reverse transcribed by M-MuLV reverse transcriptase to synthesize the first cDNA strands. For the second strand synthesis, RNase H was used to nick the RNA bound to the first strand cDNA to produce a multitude of fragments which serve as primers for DNA Polymerase I. The cDNA was then ligated to the Uni-ZAP XR vector arms and packaged into Gigapack II Gold packaging extract. The lambda cDNA library was amplified through the *E. coli* cell line PLK-F'.

B. Culture of bacteriophages on *Escherichia coli* cells

In cDNA library synthesis, 5-methyl dCTP was used in the first cDNA strand instead of dCTP, which would protect the cDNA from restriction digestion in subsequent
steps. The hemi-methylated bacteriophage DNA was introduced into an *E. coli* strain, PLK-F', which was mcrA-, mcrB- to protect the hemi-methylated cDNA from digestion by the mcrA and mcrB restriction system. After amplifying the cDNA library through PLK-F' cells, it was cultured in another *E. coli* strain, XL1-Blue, which produced blue plaques with nonrecombinants and white plaques with recombinants in the presence of IPTG and X-gal. After positive clones were isolated, the bacteriophages in the Uni-ZAP XR vector were in vivo excised as phagemids containing the double-stranded pBluescript and the cloned cDNA insert. The phagemid was mixed with XL1-Blue cells and spread on LB/Ampicillin plates to produce colonies. Plasmid DNA from minipreps of these colonies were used for DNA analysis such as restriction mapping, subcloning and sequencing.

C. cDNA library screening

The cDNA library was screened for albumin cDNA clones with a synthetic oligonucleotide: asALB-30mer, which corresponds to the published Atlantic salmon (*Salmo salar*) albumin cDNA sequence at base 1207-1226 (Byrnes and Gannon, 1990). The oligonucleotide and the corresponding cDNA sequence are:

- **cDNA:** 5' TGC TGC GAC ATG GAG GAC CAC GCA GAG TGT 3'
- **oligo:** 3' ACG ACG CTG TAC CTC CTG GTG CGT CTC ACA 5'

D. Hybridization with oligonucleotide probes

About 20 pmole of oligonucleotide was end labeled with [γ-32P]ATP (3000Ci/mmole, Amersham) by T4 nucleotide kinase (BRL) and purified by Nick-column chromatography (Pharmacia). About 1 x 10^6 CPM of labeled probe was used per milliliter of hybridization solution. Filters were prehybridized at 42°C for at least two hours in 5 x SSC (20 x SSC contains 175.3g NaCl, 88.2g Na-Citrate per liter, pH 7.0), 10 x Denhardt's [1 x Denhardt's = 0.02% BSA, 0.02% Ficoll (MW = 400,000), 0.02% polyvinyl-pyrrolidone (PVP-40, MW = 40,000)], 10% Dextran sulfate, 7% SDS, 20 mM PB, and 200 ug/ml of calf thymus DNA (Sambrook et al., 1989). Hybridization was
carried out in the same solution as prehybridization solution using 0.2% of sodium pyrophosphate instead of calf thymus DNA. Filters were incubated at 42°C for overnight and then washed in 5 x SSC-0.1% SDS. For cDNA library screening with the asALB-30mer probe, filters were washed at 42°C for 20 minutes.

3. Cloning of the 5' fragment of rainbow trout albumin 1 cDNA by reverse PCR

A. Reverse transcription reaction

To clone the 5' fragment of rainbow trout albumin 1 cDNA, 4 ug of rainbow trout total liver RNA were reverse transcribed by M-MuLV reverse transcriptase using random primers. The reverse transcription buffer contains 50 mM Tris-Cl (pH7.6), 60 mM KCl, 10 mM MgCl₂, 1 mM dNTP, and 1 mM DTT. 400U of M-MuLV reverse transcriptase (BRL), 0.5 ug of random primers (BRL) and 80 U of RNA guard (Pharmacia) were included in a 50 ul of reaction which was incubated at 37°C for two hours.

B. PCR conditions

For PCR amplification, 2 ul of reverse transcription reaction (equivalent to 0.2 ug of total RNA) was used in a 50 ul of PCR reaction. The 5' primer (5' AGATTCCGTC CTCCATCCAT C 3') corresponded to the first 21 bases of Atlantic salmon albumin cDNA and the 3' primer was the asALB-30mer used in cDNA library screening. The PCR buffer contained 50 mM KCl, 10 mM Tris-Cl (pH 8.4), 1.5 mM MgCl₂ and 0.1% Triton X-100. 0.2 mM of each dNTP, 0.5 uM of each primers and 5 unit of Taq polymerase was used in a 50 to 100 ul reaction. The PCR reaction was carried out for 30 cycles of 1 minute at 92°C for denaturing, 1 minute at 55°C for annealing, and 1 minute at 72°C for extension. From the PCR reaction, a 1.2 kb DNA fragment was amplified which was then purified from agarose gel and cloned into pT7 Blue vector (Novagen) and designated as rtALB1-p.
4. Restriction mapping and DNA sequencing

A. Restriction mapping

For restriction mapping, plasmid DNA of rtALB2 from the cDNA library and of rtALB1 from both the cDNA library and reverse PCR were cut by various restriction endonucleases (BRL or Pharmacia). DNA fragments were separated on 1% agarose gel in 1 x TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). rtALB2 is a full-length rainbow trout albumin cDNA clone. For rtALB1 restriction mapping, maps of the cDNA clone (rtALB1) including 2 kb at the 3' end and the PCR clone (rtALB1-p) including 1.2 kb at the 5' end were combined. Two clones overlapped for about 400 base pairs.

B. Subclone constructions

To sequence the inserts, subclones were constructed by restriction digestions or nested deletions. For rtALB2, digestions with EcoRI in combination with other restriction enzymes created a series of 5'-deletion subclones, i.e. rtALB2-5'NcoI(0.2kb), 5'PstI (0.5kb), 5'PstI(0.9kb), 5'StuI(1.2kb), 5'SstI(1.6kb), 5'SphI(1.8kb) and 5'AccI(2.0kb) (Numbers in brackets indicate the 5' starting position of each subclone). Similarly, digestion of HindIII in combination with other restriction enzymes created a series of 3' deletions: rtALB2-3'SphI(1.8kb), 3'StuI(0.8kb), PstI(0.5kb) and 3'NcoI(0.2kb). Digestion with ApaI made another subclone rtALB2-3'ApaI(1.6kb) (Numbers in brackets indicate the 3' ending position of each subclone).

For rtALB1, subclones were constructed from both clone rtALB1 and clone rtALB1-p. From clone rtALB1, EcoRI and HindIII were used in combination with other restriction enzymes to create deletions from the 5' end and the 3' end, respectively, as that in rtALB2. Subclones made in this way include rtALB1-5'StuI(1.2kb), 5'SstI(1.7kb), 5'SphI(1.8kb), 5'NcoI(2.0kb) and 3'ApaI(1.7kb). Nested deletions were used to obtain the other four subclones: rtALB1-5'd(2.15kb), 5'd(2.35kb), 5'd(2.5kb) and 5'd(2.6kb). Clone rtALB1-p is in pT7Blue vector. Digestion by HindIII released a 0.7 kb fragment.
from the 1.2 kb insert. Self-ligation cloned the first 0.5 kb fragment in pT7Blue vector, rtALB1-5' (0.5kb). The 0.7 kb fragment corresponded to the region 0.5 kb-1.2 kb and was cloned into pBluescript (KS) vector to make subclone rtALB1 (0.7-1.2kb).

C. Nested deletion

Nested deletions were made by using the Double-Stranded Nested Deletion Kit (Pharmacia). The subclone rtALB1-5'SphI (1.8kb) was cut by XbaI at its 5' end on the polycloning site. The created 5' protruding end was filled in with thionucleotides (dNTPαS) by Klenow fragment, which would prevent exonuclease III (ExoIII) digestion on the vector. The linearized DNA was then cut by BamHI to expose the 5' end of the insert for ExoIII digestion. During ExoIII treatment, aliquots were removed at time intervals to stop the digestions. Then the digested DNA molecules were blunt-ended by S1 nuclease and self-ligated to circular DNA which was then transformed into competent DH5α cells. Clones with inserts of desired sizes were selected by PCR.

D. DNA sequencing

Plasmid DNA were purified by alkaline lysis or a boiling method (Sambrook et al., 1989). About 2-5 ug of plasmid was used to set up one DNA sequencing reaction with Pharmacia T7 Sequencing Kit (Pharmacia) and [α-35S]dATP (1000Ci/m mole, Amersham). DNA inserts in pBluescript vector were sequenced by a T3 primer and an universal primer, and those in the pT7Blue vector were sequenced by T7 primer and universal primer. The universal primer was provided in the sequencing kit, the T3 and T7 primer were synthesized by the Biotechnology Service Center, Toronto Hospital For Sick Children. DNA sequences were analyzed by DNASIS for similarity comparison, restriction sites searching, and predicted amino acid sequence. The predicted amino acid sequences were analyzed by PROSIS.
5. RNA analysis

A. RNA purification

Total RNA were extracted from frozen tissues and embryos by a modified guanidinium thiocyanate method (Chomczynski and Sacchi, 1986). In brief, frozen tissues or embryos of about 0.1-0.5 g were homogenized in 1-5 ml of G.T.C. buffer (4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarcosyl, 0.5% mercaptoethanol). The homogenates were mixed with 0.1 volume of 2 M sodium acetate (pH 4.0) and extracted with an equal volume of water-saturated phenol/chloroform/isoamyl alcohol (25:24:1). After centrifugation at 4°C, the aqueous phase was transferred to a new tube and precipitated with isopropanol at -20°C for at least one hour. The precipitates were dissolved in RNA extraction buffer (0.1 M NaAc, pH 5.0, 0.5% SDS, 25 mM EDTA, 0.1% DEPC-treated and autoclaved) and extracted with acidic phenol/chloroform/isoamyl alcohol (phenol was saturated with 0.1 M NaCl, 1 mM MgCl2 and 10 mM NaAc, pH 5.1). The aqueous phase was precipitated by ethanol and the total RNA was dissolved in DEPC-treated water. RNA concentration was estimated by UV absorbency at 260 nm (1 OD260 = 40 ug/ml).

B. RNA separation and transfer

Total RNA was fractionated on a 1% agarose gel containing formaldehyde. The gel was cast in 1 x MOPS buffer [20 mM 3-(N-morpholino) propanesulfonic acid, 5 mM NaAc, 1 mM EDTA, pH 7.0] with 17% formaldehyde (v/v) (17 ml of formaldehyde in 100 ml). The running buffer was composed of 1 x MOPS and 9% formaldehyde (v/v). RNA samples were prepared according to Gong (1992), in which 2-5 ul of RNA samples (10-20 ug) were mixed with 8 ul of sample buffer (2 ul of formaldehyde, 1 ul of 10 x MOPS, 5 ul of formamide and 0.8 ug of ethidium bromide) and incubated at 65°C for 10 minutes before loading. After electrophoresis, RNA were transferred to Hybond-N hybridization transfer membranes (Amersham) in 25 mM PB (1 M PB contains 1 M Na2HPO4, 4 mM
EDTA, pH 7.2) by capillary elution for 20 hours or by vacuum transfer (2016 Vacugene Vacuum Blotting Unit and Pump, LKB) for 2-3 hours. Membranes were air dried and used for hybridization.

C. Northern hybridization

To label the probes, cDNA inserts were released from their vectors and purified by agarose gel. 0.1-0.3 ug of DNA fragments were labeled by [α-32P]dATP (10 mCi/ml, Amersham or Dupont) using Random Primers DNA Labeling System (BRL). The labelled probes were purified by Nick-column (Pharmacia) and counted on a scintillation counter (1219 Rackbeta Liquid Scintillation Counter, LKB). Before hybridization, filters were prehybridized for two hours at 52°C in hybridization solution (250 mM PB, 1% BSA, 5% SDS, 30% formamide, 5% dextran sulfate and 0.2 mg/ml of calf thymus DNA). Hybridization was carried out in the same solution with the addition of about 1 x 10^6 cpm/ml of labeled probe. After incubation at 52°C for overnight, membranes were washed in 2 x SET (1 x SET = 150 mM NaCl, 2 mM EDTA, 30 mM Tris-Cl, pH 8.0) supplemented with 0.2% sodium pyrophosphate (PpI, w/v) and 0.1% SDS for 1 hour at room temperature and 30 minutes at 65°C, then with 0.2 x SET-0.2% PpI-0.1% SDS at 65°C for another 30 minutes. Membranes were exposed to Kodak X-ray films with intensifying screen at -70°C.

D. cDNA fragments used in Northern hybridization

Full-length rtALB2 cDNA and DNA fragments specific to rtALB1 or rtALB2 were used as probes in Northern hybridization, respectively. The rtALB2-specific fragment, designated as rtALB2-3' probe, was the insert in subclone rtALB2-SacI(2.0kb). It contains 280 bp in its 3' untranslated region and shares a low identity (60%) with rtALB1 in the same region.
For rtALB1-specific probes, two fragments were purified from subclones rtALB1-5'SphI(1.8kb) and rtALB1-5"NcoI(2.0kb). From rtALB1-5'SphI(1.8kb), BamHI and NcoI digestion released a fragment of 170bp which was purified from agarose gel and subjected to AluI digestion to get a 120bp fragment. This fragment, designated as rtALB1-probe A, covered the sequence from 1919-2040 bases along the rtALB1 cDNA, including the last 12bp of the reading frame, 53bp between the translation stop codon and the first polyA signal (at base 1978-1983) and 57bp after the first polyA signal. It had 76% identity with the corresponding region of rtALB2. The other rtALB1-specific probe, rtALB1-probe B, was from subclone rtALB1-5"NcoI(2.0kb), which contains the last 800 bp of the 3' untranslated region after the first polyA signal.

6. Nuclear run-on experiments
   A. Isolation of nuclei from rainbow trout liver

   Nuclei were prepared by a modification of the method described by Gorski et al. (1986). Fresh rainbow trout liver (6-8 g) was homogenized in 100 ml of homogenization buffer (10 mM HEPES, pH 7.6, 15 mM KCl, 15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2M sucrose, and 10% glycerol) using a motor-driven 50 ml Teflon-glass homogenizer. 0.5 mM DTT, 0.5 mM PMSF (from 0.1 M stock in isopropanol alcohol) and protease inhibitors such as aprotinin (2 ug/ml), leupeptin (0.5 ug/ml) and benzamidine (1mM) were added to the buffer just before use. The homogenate was aliquoted to three to four parts and each part was layered onto 10 ml cushion of homogenization buffer in a Beckman SW27 tube. Cell lysates were centrifuged at 24,000 rpm for 60 minutes at -2°C in a Beckman SW27 rotor. After centrifugation, supernatants were carefully removed and nuclei at the bottom of each tube were resuspended in 300 ul of glycerol storage buffer (25% glycerol, 50 mM Tris-Cl, pH 8, 1 mM EGTA, 1 mM spermidine, 1 mM DTT and 0.1 mM PMSF). Nuclei in glycerol storage buffer was used for RNA chain extension reaction immediately. All steps were manipulated at 0°C.
B. RNA run-on reaction and isolation of run-on RNA

The RNA chain extension was performed in a total volume of 200 ul which consisted of 100 ul of nuclei in glycerol storage buffer and 100 ul of reaction buffer (40 mM NaCl, 100 mM KAc, 2.5 mM MgCl₂, 300 mM glycine, 10 mM Tris-Cl, pH 8). 1 mM of unlabelled ATP, CTP, GTP and 100 uCi of [α-³²P]UTP (400 Ci/m mole, Amersham) were included in each reaction. The reaction mixture was incubated at 20°C for 20 minutes. After reaction, chain-extended RNA was isolated essentially according to Groudine et al. (1981). In brief, the reaction mixture was treated with 25 μg of DNase I for 10 minutes at 20°C. The reaction volume was adjusted to 500 ul and the final SDS and EDTA concentration was brought to 1% and 5 μM respectively. The mixture was treated with 100 μg/ml of proteinase K at 65°C for 30 minutes and extracted three times with phenol/chloroform/isoamyl alcohol. The RNA was precipitated with 5% TCA-1.5% PPI on ice for 30 minutes and collected on a 0.45 um nitrocellulose filter (HAWP 025 00 filter from Millipore) by filtration under vacuum. The filters were washed with 50 ml of 3% TCA-1.5% PPI, soaked in 1.8 ml of DNase buffer (20 mM HEPES, pH 7.5, 5 mM MgCl₂, 1 mM CaCl₂) and treated with 25 μg of DNase I at 37°C for 30 minutes. To elute RNA from the filters, filters were incubated at 65°C for 10 minutes after adjusting the EDTA and SDS concentration to 15 mM and 1% respectively, followed by washing with 1 ml of 1% SDS-5 mM EDTA-10 mM Tris-Cl (pH 7.5) at 65°C for 10 minutes. The combined elutes were treated with 100 μg of proteinase K at 65°C for 30 minutes. RNA was precipitated by ethanol in the presence of 50 μg of yeast tRNA. The final RNA pellet was dissolved in hybridization buffer and counted.

C. DNA fragments used in run-on hybridization

In order to distinguish rtALB1 and rtALB2 and compare their transcription rates, two DNA fragments were amplified by PCR. These two fragments cover the 150 bp
immediately after the translation stop codon on each albumin cDNA sequence and share low identity (60%). Primers used to amplify these fragments are:

- **ALB1-5' primer**: 5' GATAGAATTGTGATGATT 3'
- **ALB1-3' primer**: 5' GGCTTATGAGTGGATGAC 3'
- **ALB2-5' primer**: 5' GATATTTATTTTCATTGTA 3'
- **ALB2-3' primer**: 5' GGCTTAGTGGATCACAGA 3'

Using rtALB1 and rtALB2 plasmid as templates, each pair of primers amplified a 150 bp fragment from PCR reaction. The PCR fragments were separated on 1% agarose gel and transferred to nylon hybridization membranes. Southern hybridization (same condition as Northern) with rtALB1 and rtALB2 cDNA probes showed that they were specific to each albumin cDNA (see results and Fig.2-14).

D. Slot-blots and hybridization

The PCR fragments were purified from 1% agarose gel by electroelution and dissolved in TE (10 mM Tris-Cl, 1 mM EDTA). 0.5 ug of each fragment was diluted in 100 ul of water and mixed with 10 ul of 3 M HCl at room temperature for 5 minutes. The DNA fragments were denatured by adding 30 ul of 3 M NaOH and incubating at 37°C for 15 minutes. Samples were chilled immediately on ice and the solution was neutralized by adding 20 ul of 3 M HCl and 150 ul of 6 M NH₄Cl. The DNA was blotted onto GeneScreen Plus filters (DuPont) using a slot-blotter under slight vacuum, and washed with 1 x SET. 0.5 ug of pBluescript was blotted as a negative control. Filters were air dried and used in hybridization. Prehybridization, hybridization and washing conditions were identical as described in Northern hybridization. 2-4 x 10⁶ cpm run-on labeled RNA were used in a 1.5 ml hybridization solution. The hybridization was carried out for 3 days.
E. Autoradiography and densitometry

Hybridized filters were exposed to Kodak X-ray film with an intensifying screen at -70°C for 1-3 days. RNA levels were estimated by densitometric scanning of autoradiographs using ScanJet 3P (Hewlett Packard) and analysis using a computer program (NIH Image 1.52).
RESULTS

1. Northern hybridization with asALB-30mer probe

In order to isolate rainbow trout albumin cDNA clones, an oligonucleotide asALB-30mer was synthesized according to the published Atlantic salmon albumin cDNA sequence (Byrnes and Gannon, 1990). The corresponding cDNA and the deduced amino acid sequences are:

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Cys</th>
<th>Cys</th>
<th>Asp</th>
<th>Met</th>
<th>Glu</th>
<th>Asp</th>
<th>His</th>
<th>Ala</th>
<th>Glu</th>
<th>Cys</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA</td>
<td>5'</td>
<td>tgc</td>
<td>tgc</td>
<td>gac</td>
<td>atg</td>
<td>gag</td>
<td>gac</td>
<td>cac</td>
<td>gca</td>
<td>gag</td>
</tr>
</tbody>
</table>

The selection of this sequence was based on the following considerations: 1. There are three cysteine residues in this sequence. Cysteine residues are important for maintaining albumin structure and are conserved in albumin from various organisms. It is expected that these three cysteines are conserved between Atlantic salmon albumin and rainbow trout albumin; 2. One methionine, two aspartic acid and two glutamic acid residues are included in the sequence. Methionine has only one codon, aspartic acid and glutamic acid have two codons each, thus minimizing the degeneracy; 3. This oligonucleotide is located in the middle of the cDNA sequence, it would hybridize to cDNA clones no shorter than 1 kb, thus facilitating the isolation of larger cDNA clones.

To test if this oligonucleotide is suitable to isolate rainbow trout albumin cDNAs, Northern blot was carried out in which the asALB-30mer was ^32P-labelled and hybridized to rainbow trout liver RNA immobilized on a filter. Three bands were shown between the 18S and 28S ribosomal RNA (Fig. 2-1). The signals remained visible even at stringent washing conditions (2 x SSC-0.1% SDS, 52°C, 20 minutes), suggesting that the asALB-30mer was suitable for screening the rainbow trout albumin cDNA from its liver cDNA library.
Fig. 2-1  Northern hybridization of rainbow trout liver total RNA with asAlb-30mer
2. Isolation of two distinct rainbow trout albumin cDNA clones

A rainbow trout liver cDNA library was constructed in lambda-ZAP II vectors. From 7 μg of liver poly(A)⁺ RNA, 2.3 x 10⁶ recombinants were obtained, representing 3.3 x 10⁵ recombinants per microgram of poly(A)⁺ RNA. Using asALB-30mer as probe, five positive clones were isolated from 8000 plaques of the unamplified cDNA library. The lengths of inserts for each positive clone were 1.3 kb, 1.6 kb, 1.7 kb, 2.0 kb and 2.2 kb, respectively. Two clones with 2.0 kb and 2.2 kb inserts were restriction mapped, subcloned, and sequenced. DNA sequences were analyzed by DNASIS, which showed that the 2.2 kb clone is a full-length cDNA clone and the 2.0 kb clone is an incomplete clone which had about 800 bp of its 5' end deleted. The deleted 5' fragment was later cloned by reverse PCR (see below). Comparison of the two albumin clones indicated that they are more than 20% divergent on the amino acid sequence level, suggesting that they represent two distinct rainbow trout albumin genes. The 2.0 kb clone shares greater identity with Atlantic salmon albumin cDNA (91%) than the 2.2 kb clone does (81%). The 2.0 kb clone is therefore the homologous gene of the Atlantic salmon albumin cDNA and is designated as rainbow trout albumin 1 (rtALB1). The 2.2 kb clone represents a new fish albumin clone and is designated as rainbow trout albumin 2 (rtALB2). Fig. 2-2 shows the restriction maps of rtALB1 and rtALB2 cDNA. The restriction sites in the first 0.8 kb of the rtALB1 cDNA was determined on the reverse PCR clone, rtALB1-p (see below). As a result of divergent DNA sequences, some restriction sites are conserved in both clones, others are specific to each of them. Partial DNA sequences showed that the other three clones were incomplete rtALB1 clones.

3. Characterization of rainbow trout albumin 2 cDNA

The rainbow trout albumin 2 cDNA (rtALB2) was isolated as a full length clone from the liver cDNA library. This clone has 2250 bp nucleotides in total (Fig.2-3). It contains 78 bp in the 5' untranslated region, 1824 bp in the reading frame to encode 608
Fig. 2-2 Restriction maps of two rainbow trout albumin clones, rt-Alb1 and rt-Alb2.
Fig. 2-3  rtALB2 cDNA and derived amino acid sequence

The rtALB2 cDNA has 2250 bp in total. It contains 78 bp in the 5' untranslated region, 1824 bp in the reading frame to encode 608 amino acids, and 348 bp in the 3' untranslated region. The bold and underlined prints indicate the polyadenylation signal.
AGT CTG TCC GAG CAC TAC
Cys Asp Lys Glu Asp His
GAG
Asp
Glu
TGT
Ala
Met Pro
GAA GCC CGG TCT CAG
Met Gln Leu Ala Val Cys Ser Leu Leu Val Val Cys Val Ser Ala Arg Ser Gin
20
AAT CATG TG5 GCT GTC TGT AGG CTG TGT GCT TGT GTA GCA GCC GCG TCT CAG
335
Met Gln Leu Val Ala Val Cys Val Ser Leu Leu Val Val Cys Val Ser Ala Arg Ser Gin
20
AAT CATG TGT ACC ATC TTCA GAG GCC AAG AAA GAT GGA TGG AAG CCT AGC TGA
1036
Asp Gin Ile Cys Thr Ile Phe Thr Glu Ala Ala Gin Ser Gin
103
GTA GCC CTG GCT GAC ACC CTG GCC GGT ATG CAC TTC GTC
Val Gly Leu Ala Gin Asn Ser Leu Leu Asp Val Val Pro Leu Val Val Ala
60
GAA GCC TTC ACC ATG GCC ATC CAA TGG TCA CAC GAA GCC CTA GGA GAC TGG TAC
318
Glu Ala Phe Thr Met Gly Ile Gin Cys Cys Asp Ser Asp Glu Pro Leu Glu Asp Ser Gin
80
GAC GCC GCC GAC GTG TTC AGT CAG CTG GTG TCT TCT GAG ACC TTA GTG GAG AAG AGC
378
Asp Val Ala Asp Leu Gin Cys Ser Ser Gin Thr Leu Val Leu Val Cys Val Ser Ala Arg Ser Gin
20
AAT TTA GAG CAC TGC TGT GAG AAT ACT GCC ATT GAG AGA ACC TGC TGC GAC CAC
438
Arg Leu Lys His Cys Thr Ala Gin Gin Thr Gin Thr Gin Gin Thr Gin Gin Ser Gin
120
AAG GCC CAG ATT CCC CGG GCC CTG CCT TGC ACA TCT GAG TTC CCA GCT GCA GCC CAG TGT
498
Lys Ala Gin Gin Thr Gin Gin Ser Gin Thr Gin Gin Thr Gin Gin Thr Gin Gin Ser Gin
140
GAA GCC TTC ACC ATG GCC ATC CAA TGG TCA CAC GAA GCC CTA GGA GAC TGG TAC
558
Glu Ala Phe Thr Met Gly Ile Gin Cys Cys Asp Ser Asp Glu Pro Leu Glu Asp Ser Gin
80
CTG GCC ACC ATG CTA CTT CCA AGT GCT CAG ATG AGT TAC GCC GGA GGT
1088
Arg Asp Thr Met Leu Pro Pro His Val Ile Leu Ala Val Thr Tyr Gin Val Gin
180
CTG GCC ACC ATG CTA CTT CCA AGT GCT CAG ATG AGT TAC GCC GGA GGT
618
Arg Asp Thr Met Leu Pro Pro His Val Ile Leu Ala Val Thr Tyr Gin Val Gin
180
CAG ATC TCC GAA CAC TCC TTC AGG AAC CTT GTG GCC TTC AGC CTG TAC AGT CAC TTC
738
Pro Thr Phe Glu His Phe Val Arg Asp Arg Val Asn Glu Leu Ala Leu Cys Ile Gin
220
CAC AAC AAA TAT GGA GCC GCC GCC AAC AAG ATG CTC TAC CAG TAC AAG AGT CAG
798
His Asn Lys Tyr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
240
ATG CCT GCG TCG TGT GCG AAT GAG AGA GGC GAG GCC CAG ACC TGG GCC
ATC ACC CAG AAA
678
Met Pro Gin Ala Ser Gin Gin Glu Gin Gin Thr Gin Thr Gin Gin Gin Gin Gin Gin Gin Gin
30
GCC CCC TGC TGC ACG GCA GCC ATC ACA TGC ATG AAG CAG AAG GTG CTG GTA GAC
918
Ala Pro Cys Cys Ser Gin Gin Thr Gin Thr Cys Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin
280
GAG GTG TGT GGT GCC CGT AGT GTG TTG TCT GCC ATG GCG GTG AGA TGT GCC
358
Glu Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
1008
GAA CAT GCC ATC AAG GCC GCC GCC ATG AGC GCC GCC GCC GCC GCC GCC GCC GCC GCC
1038
Glu His Gin Ile Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
120
AGT CTG TCC GAG CAC TAC GAG CTG CAC GCC TAC ATG GCT GTC TGC GCC
ACC TCC ACC
1098
Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
120
AAG ACC CCT GAT GTA GCC CGG AGG ATT TAC TAT GAG CTG TCA GTG CGC CAC CCT GAG
1158
Lys Thr Pro Gin Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
360
TGC TCT CAG CAG GTT ATC CTG AGG TCC GAG GCT GCC GCC GCC GCC GCC GCC GCC GCC GCC
1218
Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
360
TGT GAC AAG GAG GCC CAC GCA GAG TGT GCC GCT GCC GCC AAG AGC ATC AAG
1278
Cys Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
400
ACG Asn Gln

AGC ATC

AAG

GGA Asp

GAC qca ttt att tqa aaaaaoaaaaaaaa

Cys Cys

AGC AAG GAG AGT GTA GAG

qaa

Gly His Tyr Phe Glu Lys Ser Met Val Tyr Thr Arg Ile Met Pro Glu Ala

TGC TTC GTC CAG CTT CAC ATG GTG TCT GAG ACT GTG GCT GAT GTT TTC CAT GAC TGC Tcc

Ser Phe Val Gln Leu His Met Val Ser Glu Thr Val Ala Asp Val Phe His Asp Cys

Glu 460

GCA GAG CAG CCA GCC CAC TTA GTC TTG CCC TGC GCA GAG AAG TCG ACC AAC GTG CTC

Lys Asp Gin Pro Gln His Leu Val Leu Pro Cys Ala Glu Glu Lys Ser Thr Asn Val Leu

GAC CAC ACG TGT AAC GAC GAC TAC TCC AGC ATT AAC CCC GCC ATC GCC CGC TGC TCC

Asp His Thr Cys Asn Asp Asp Tyr Ser Ser Ile Asn Pro Arg Ile Ala Arg Cys Cys

AAC CAG TCC TAC TGC ATG AGG AGA CCT TGC ATC CAG CCT GAC GCC GAG TTC

Asn Gin Ser Tyr Ser Met Arg Arg Pro Cys Ile Leu Ala Ile Gin Pro Ala Glu Phe

GAG 520

AGC CCT CCG GAG CTG GAT GCC AAC GAC TTC CAC ATG GCC CCC GAG CTC TGC ACC AAC

Thr Pro Pro Glu Leu Asp Ala Asn Asp Phe His Met Gly Pro Glu Leu Cys Thr Lys

GAC AAG GAC CTC CTG CTA TCC GGG AAG AAG TTG CTG TAT GAT GTG GTC AGA CAT ARG ACC

Ser Lys Asp Leu Leu Ser Gly Lys Leu Leu Tyr Val Val Arg His Lys Thr

GAG 1758

AGC ATG ACC GAG CAT CTG AAG GCC CTC TCT GCT AAT AAC CAC ATC ACG AGG GAG AAC

Ser Ile Thr Glu Asp His Leu Lys Ala Leu Ser Asn His Ser Ile Thr Arg Glu Lys

TCC TGC GCC GCT GAG GAC AAA GAA GCA TGC TTC ACT GAC GAG GCC AAG CTG CTA GCT

Cys Cys Ala Ala Gln Leu Asp Lys Glu Ala Cys Phe Thr Asp Glu Ala Lys Leu Leu Ala

GAG 1878

GAG GTG GTA GAG GTG TAA gat att att ttc att gta ttc att tca caa gag

Glu Ser Val Glu Leu Val Lys Val ***

gaa ctg gtg aat taa aaa aaa aaaa tgc tca aag ctc gtt tgt gtc tac aat atc tac 1998

cgc cat gac aac acc ata tgt acc tca cct cat cct ttc aat cat ata atg tgt tgt aac 2058

cac taa gcc tgt tca tgt atg atg atg atg aag taa ctc ttc tgt ctc tgt atg aag tga aat 2118

cat taa cct cga tct tgt atc gac tca aac ctc ctg atg atg gac aag aat ctc aat 2178

tcg tgt tgt gat atc ttc ctc ttc ctc aag ctc ctc ttc ctc ttc ctc ctc ctc 2238
amino acids, and 348 bp in the 3’ untranslated region (UTR). A polyadenylation signal AATAAA is found 13 bases before the polyA tail. Primer extension experiments were carried out to show that there are 5 more bases before the cDNA sequence (see Chapter 3), so the 5’ untranslated region of rtALB2 has 83 nucleotides in total. Compared to Atlantic salmon albumin, the two albumin cDNAs have 81% identical nucleotides. The identity between two reading frames is 84.6%, while that for 5’ and 3’ untranslated regions are 71.2% and 64% respectively. This identity is low considering that they are from two closely related fish species. Salmonids are tetraploid and each protein is encoded by two copy of genes, for example two growth hormone genes and two metallothionein genes have been isolated in one salmonid species (Agellon et al., 1988; Sekine et al., 1985; Du et al., 1993; Gedamu and Zafarullah, 1993; Olsson, 1993). The rainbow trout growth hormone 1 and 2 is 92.2% and 88.7% identical to the chum salmon growth hormone 1. It appears that rainbow trout similarly has two albumin genes and the 2.2 kb albumin cDNA clone characterized here represents a second form of albumin distinct from the Atlantic salmon albumin.

4. Characterization of rainbow trout albumin 1 cDNA

Other cDNA clones isolated from the rainbow trout liver cDNA library were further characterized. The one with a long insert (2.0 kb) was completely sequenced. This clone (rtALB1) has 2023 base pairs. Its first nucleotide corresponds to the nucleotide 781 in rtALB2, so it is an incomplete albumin cDNA clone. This clone showed 20% difference in nucleotide sequence from rtALB2. It is more identical to Atlantic salmon albumin and shares over 90% identities in DNA sequence. In order to further compare trout albumin 1 and albumin 2, the 5’ fragment of rtALB1 was isolated by reverse PCR. As shown in Fig.2-4, a 1.2 kb fragment was amplified by PCR. This fragment was purified from agarose gel and cloned into the pT7Blue vector for DNA sequence analysis. DNA
Fig. 2-4 Reverse PCR amplification of the 5' fragment of rtALB1

P. PCR product; M. molecular marker; The arrow indicates the 1.2 kb PCR band.
sequences of the PCR clone (rtALB1-p) overlapped 400 bases with the rtALB1 clone, confirming that it was the 5' fragment of rtALB1.

Fig. 2-5 shows the DNA sequences and the derived amino acid sequences of rtALB1. The rtALB1 cDNA is 2761 bp long. Its reading frame has 1821 bp to encode 607 amino acids. There are 69 bp and 872 bp on its 5' and 3' untranslated regions respectively. The first 21 nucleotide at 5' end is the position of the 5' primer used in reverse PCR. When cloning the promoter of rtALB1 gene (see Chapter 3), this sequence was shown to differ in only one nucleotide at position 5 where a cytidine in rtALB1 replaces the thymidine in asALB. Three putative poly A signals AATAAA are located at 53, 60, and 851 nucleotides after the stop codon. The sequence shown in Fig. 2-5 is the one using the third poly A signal (the most 3' one) which adds a polyA tail 18 nucleotides after the signal. The 1.7 kb cDNA clone isolated from the cDNA library used the first polyA signal, which added a poly A tail 14 bp after the signal. Two poly A signals have been found in Atlantic salmon albumin cDNA (Byres and Gannon, 1990) and three in human albumin gene (Minghetti et al., 1986). Human albumin cDNAs using the first two poly A signals have been isolated. Comparing the DNA sequences, rtALB1 is 92.6% and 81.5% identical to Atlantic salmon albumin and rtALB2, respectively. The identity in the coding region is 95.8% between rtALB1 and asALB and 85.6% between rtALB1 and rtALB2, that in the 5' untranslated region is 98.6% and 65.2% respectively.

5. The long 3' untranslated region in rtALB1 cDNA

The rtALB1 cDNA has a relatively long 3' untranslated region (3' UTR) which contains 872 bp, while rtALB2 and Atlantic salmon albumin has only 347 bp and 337 bp in their 3' UTR. Sequence analysis revealed that this long 3' UTR is caused by a repetitive sequence (Fig. 2-5 and Fig. 2-6). Between nucleotide 2055 to 2629, a sequence of 34 bases repeats 17 times to create an 578-bases insert in the 3' UTR. Before and after this insert, there are 161 bp and 134 bp in the 3' UTR which shows high identity (94.2%) with

65
Fig. 2-5 rtALB1 cDNA and derived amino acid sequence

The rtALB1 cDNA has 2761 bp in total. It contains 69 bp in the 5' untranslated region, 1821 bp in the reading frame to encode 607 amino acids, and 872 bp in the 3' untranslated region. The underlined sequence indicates the repeat sequence in the 3' untranslated region. The bold and underlined prints indicate the polyadenylation signals.
AGA TCC CTT 9

CCT CCA TCC ATC CTT CCT CTC CAT CCA TCC CTC CGT GCC TGT CCT TCC ATC 69

ATG CAG TGG TGG TCT GTC TGC ACC CGT TGG GTG CTG AGT GTG TCA GCC CGG TCT CAG 129
Met Gin Trp Leu Ser Val Cys Ser Leu Val Leu Leu Ser Val Ser Ala Arg Ser Gin 20

GCT CAG AAC CAG ATC TGT ACC ATC TTC ACG GAG GCC AAG GAA GAT GGA TCC AAA TCT TGG 189
Ala Gin Asn Gin Ile Cys Thr Ile Phe Thr Glu Ala Lys Glu Asp Gly Phe Lys Ser Leu 40

ATC CTT GTA GGG CTG CTT CAG AAT CTT CCC GAT AGT ACG TTG GAT GCC TCT GCT TCT 249
Ile Leu Val Gly Leu Ala Gin Leu Pro Asp Ser Thr Leu Gly Asp Leu Val Pro Ser Leu 60

ATC GCC GAG GCC TTA GCC ATG GCC GTC AAG TGC TGC TCA GAC ACT CTT CCA GAG GAC TGC 309
Ile Ala Glu Ala Leu Met Gly Val Lys Cys Ser Asp Thr Pro Pro Glu Asp Cys 80

GAC AGA GAT GTG GCG GAC CTG TTC CAG AGT GCG GTG TGT TCC TTC GAG ACG CTG GAG 369
Asp Arg Asp Leu Phe Gin Ser Ala Val Ser Ser Gly Thr Leu Val Glu 100

AAG AAC CAT CTG AAC ATG TGC TGT GAG AAG ACT GCC GCT CAG ACG AAG ACA CAC TGC TTT CTG 429
Lys Asn His Leu Lys Met Cys Cys Gly Lys Thr Ala Gin Arg Thr His Cys Phe Leu 120

GAC CAC AGG GCC ATG CAA TGT GCC GTC TGC TCC CTC AAA GCT GAG CTG CCC GCT GCA GAC 489
Asp His Lys Ala Leu Ala Met Arg Pro Arg Asp Ser Leu Ser Leu Ala Gin Ala Asp 140

CAG TGT GAA GAC TTC AAG AAG GAC CAG AAT TTC TGT GGG AGG AGG TTC ATC TTC AAG TTC 549
Gin Cys Glu Asp Lys Asp His Ala Val Phe Gly Arg Ile Phe Lys Phe 160

TCC AAG AGT AAG CAG ATG CTA CAA CCA CAT GTG ACT CTG ACT GCC AAA GCT TAT GGA GAG 609
Ser Lys Asn Thr Met Leu Gin Pro His Val Ile Leu Thr Ala Lys Tyr Gly Glu 180

GTT CTG ACT AGC TGC TGT GGA GAG GCT GAG GCC CAG ACT TCC TTC GAC ACC AAG AAA GCT 669
Val Leu Thr Ser Cys Gly Glu Ala Glu Gin Thr Cys Phe Thr Lys Thr 220

ACT TTC CAA GCC GCC GTC GGG AAA GCC GTG ACT CCA TCT GTC GAG GCC GTG TTC ATC TCC CAC 729
Thr Phe Gin Ala Val Lys Gly Thr Val Thr Leu Arg Ala Leu Cys Ile Val His 220

AAG AAA TAT GGA GAC GCC GTT GTC AAG GCC AAG AAG CTG ACT CAG TAC AGT CAG AAG ATG 789
Lys Lys Tyr Gly Arg Val Val Val Lys Ala Lys Ile Gin Tyr Ser Gin Gly Met 240

CCT CAG GCC TCT TTC CAG GAG ATG GGA GCC ATG GTA GAC AAG ACT GTA GCC ACT GTT GCC 849
Pro Gin Ala Ser Phe Gin Glu Met Gly Gly Met Val Asp Lys Ile Val Ala Thr Val Ala 260

CCC TGC TGC AGC GGA GAC ATG GTC AGA TGC AAT AAG GAG AAG GCC GTG CTG GAT GAG 909
Pro Cys Ser Gly Asp Met Val Thr Cys Met Lys Glu Arg Ala Leu Val Asp Glu 280

GTT TGT GCT GAT AAG AGT GTG TTC TCT CTC GCC GCG GGT GTG TCT GTA TCC GGT TGT AAG GAG 969
Val Cys Ala Asp Lys Ser Val Leu Ser Arg Ala Ala Gly Leu Ser Cys Met Lys Ala Gly 300

GAT GCA GTA CAC AGA GGG TTC TGT GTT GAG GCC ATG AAG CCA GAC TCT AAG CCA GAC GGT 1029
Asp Ala Val His Arg Ser Cys Glu Ala Met Lys Pro Ser Lys Pro Asp Glu 320

CTG TCT GAG CAC TAC GAC GTC AAG GCG GTT CAG GTA GCC AAAG CTA GCA GAC ATG ACC 1089
Leu Ser Gin His Tyr Thr Val His Ala Asp Ile Ala Ala Val Cys Gin Thr Phe Thr Lys 340

ACC CGG GAT GTA GCC ATG AAG GAG TGG TCT GAT TAT GAG ATC TCA GTG GCT GAT GAC 1149
Thr Pro Asp Val Ala Met Gly Leu Val Thr Tyr Val Ser Val Arg Asp Glu Ser 360

TCT CAG GAG ATT CTG AAG TTC GCC AAG GAG GCT GAG CAG GCC TGG TCT CAG TGC TGT 1209
Ser Gin Glu Val Ile Leu Arg Phe Ala Lys Glu Ala Glu Ala Leu Gin Cys Ser 380

GAC AAG GAG GAC CAT GCA GAG TGT GTC AGA ACC GCT CTA GGA GCA GAG CAT ATT GAC AAG 1269
Asp Lys Glu Asp His Ala Glu Cys Val Arg Thr Ala Leu Ala Gly Ser Asp Ile Asp Lys 400
AAG ATC ACT GAT GAG ACT GAC TAC TAC AAG AAG ATG TGT GCT GTT GAG ACT GCC ATG AAC 1329
Lys Ile Thr Asp Glu Thr Asp Tyr Tyr Lys Lys Met Cys Ala Val Glu Thr Gly Met Asn 1920
AAT AAC GCC TTT GAG AAG ATG ATG GTG CAC TAC ACA AGG ATA ATG CCC CAG GCC TCC
Asn Asn Ala Phe Glu Lys Ser Met Met Val His Tyr Arg Met Pro Glu Ala Ser 440
TTC GAC CAG CTC CAC ATA GTG TCA GAG AGG GTG CAT GAT GTC CTC CAT GAC TGC TGC AAG 1449
Phe Asp Gln Leu His Ile Val Ser Glu Arg Val His Val His Cys Cys Lys 460
GAC GAG CCA GCC CAC TTC ATC CTG CCT TGT GCA GAG AGG CTG ACC GAT TCC ATC GAC 1509
Asp Glu Pro Gly His Phe Ile Leu Pro Cys Ala Glu Lys Thr Asp Ser Ile Asp 480
GCC ACA TGT GAG GAC TAC GAC CCC TCC AGC ATT AAC CCC CGC ATC GCC CAC TGC TGC AAG 1569
 Ala Thr Cys Glu Asp Tyr Asp Pro Ser Ser Ile Asn Pro Arg Ile Ala His Cys Asn 500
CAG TCC TAC TCC ATG AGG AGA CCC TGT ATC CTG GCC ATC CAG CCT GAC AGG GAG TTC ATG
Gln Ser Tyr Ser Met Arg Arg Pro Cys Leu Ala Ala Glu Pro Thr Gly Met Pro Thr 520
CCC CCA GAG CTG GAT GCC AGC AAC TTC CAC ATG GCC CCC GTC ACC AAG GAC ACC 1689
Pro Pro Glu Leu Asp Ser Asn Ser Asp Ser Pro Cys Leu Ala His Met Pro Gly Leu Lys Asp Ser 540
AAG GAG CTG CTG CTC TCT GGG AAG AAA CTA CTG TAT GGT GTG GTC AGA CAT AAG ACC ACC
Lys Glu Leu Leu Leu Ser Gly Lys Lys Leu Leu Tyr Gly Val Arg His Lys Thr Thr 560
ATC ACT GAG GAG CTG AAG TTC TCC ACC AAC TAC ACT AAA TAT CAC AGT ATG AAG GAG AAG TGC
Ile Thr Glu Glu Leu Leu Ser Ile Ser Ile Thr Tyr Arg Pro Cys Met Thr Lys Tyr His Ser Met Cys Lys 580
TGT GCT GAG GAC CAA GCA GCA TGC TTC ACT GAG GAC CCA CCC AAG CTG GGT GCT GAG
Cys Ala Ala Glu Asp Gin Ala Ala Cys Phe Thr Glu Glu Ala Pro Lys Leu Val Ala Glu 600
AGT GCA GAC CTG GTC AAG GCT TAA gat aga att gtt atg att tta att tca aaa gat gaa 1929
Ser Ala Glu Leu Val Lys Ala ***
cgg tct gtc atg att taa aaa aat aat aat taa aat ggg ttc tgt ctg tat cqa tcc atc tgc tcc 1989
ccc agc cta cat tta cac cat ggt gac ctc atc cac cta tcc aca acq tct tgt cat cca 2049
cct ata agc ctg tct atc atg gat ctc atg tga tca tgc act cat gag tct gtc tat cag tgc tct 2109
gtg att gac tca taa gta gtc tgt cta tca gtt gtc atc tgt gat cga tca tca gtt gtc atc 2169
agt gat ctg tga tgc act aag tct tgt caa gtt cag tgt ttc agg tga tgt ctc gtc atg tgt ctc 2229
tct atc act gat tga ctc taa atg agt ctc atg tca tcc ctc ctc gtt ctc atc 2289
gag gaa tct tat cag tga tct ctc gat gaa tca taa gtt cag tcc ctc ctc gtt cta tact gag 2349
tga ctc atc atg gat ctc tct tga tct act cat gag tct gtc tct cag tga 2409
tct tgt atc gac tca tga gtc tgt ctc ctc ctc ctc tca gtt atc tgt gta ctc atc gat ctg 2469
atc agt gat ctg tga tgc act aag tct gtc tat cag tga tct gtc ggt att gag tca tca tga 2529
gtc tgt ctc ctc ctg gtt gtt atc tgc atg tga ctc tct ctc atg gat ctc tct ctc tgc 2589
act cat gtc gtt ttc ttc ctc gtt gtt gtt att gtc ttt gtt tct ctc ctc ttt ttc gtt gtt gtt 2649
atc aaa cct cta tct ggg atg aat cac gcc gtt gtt ctt ctc ata gac atg aat gac 2709
tcc tgt cca tca aat aca tga ata act gaa taa gtt ctt tat cca ctc cta aat aat gaa 2769
aaa aaa aat aat aat aat aat aat
Fig. 2-6 Comparison of the 3' untranslated region of rtALB1 and Atlantic salmon albumin (A) and comparison of the 34 nucleotide sequence in the 3' UTR of three fish albumin cDNAs (B)
Atlantic salmon albumin 3' UTR (Fig. 2-6a). The sequence of the 34 bases is shown in Fig.2-6b. It corresponds to base 2059-2092 in Atlantic salmon albumin cDNA. Among the 17 repeats, only two sites in the 34 bases (site 1 and 27) showed variations: A or G for site 1, and T or C for site 27. On the sixth repeat, two deletions on the first two sites were found. Other than these changes, the 17 repeats are well maintained. The 34-bases sequence is also found in rtALB2 3' UTR at nucleotide 2063-2097 (Fig. 2-6b). Comparing the 34 bases from three salmonid albumin cDNAs, it is found that Atlantic salmon albumin has completely identical nucleotides and rtALB2 has 29 identical nucleotides, suggesting that this sequence is very conserved in salmonid albumin mRNAs. Homology search in Genebank did not find any similar repeat sequences. How the repetitive insert originate in rtALB1 and the significance of the 34-bases segment is presently unknown.

6. Comparison of fish albumin

The two rainbow trout albumin polypeptide sequences derived from their cDNA sequences together with that of Atlantic salmon albumin are compared in Fig. 2-7. rtALB1 and rtALB2 transcripts encode 607 and 608 amino acids respectively. rtALB2 has two deletions at site 19-20 comparing to the other two albumin, while the latter two have two deletions at site 189-190. rtALB1 has one less amino acid than rtALB2 and Atlantic salmon albumin, which is caused by a deletion at site 176. Overall, the three fish albumin polypeptides align well. Identities among fish albumin are 77.5% for rtALB2 and rtALB1, 77.8% for rtALB2 and asALB, and 90% for rtALB1 and Atlantic salmon albumin (Fig. 2-8). Like asALB, both rainbow trout albumin have 14 amino acids in their signal peptide and 4 amino acids in their propeptide. The pre and propeptide are processed to generate the mature proteins of 589 and 590 amino acids for rtALB1 and rtALB2, respectively. The most important feature of albumin structure is that the conserved cysteines distribute along the polypeptide to form disulfide bridges which fold the protein into three
Fig. 2-7 Comparison of three fish albumin amino acid sequences

* represents a deletion at this position. Bold prints indicate the putative glycosylation sites.
rtALB2: MQLLAVCSLLVLLCVSARS**QNICTIFTEAKKGDFKAMALVGLAQNLp 50
rtALB1: --W-S--------S------Q-A---------E------S-LI-------
asALB: --W-S--------S--LS--QA----------------E------S-LI-------

DSLLDDVPLVAEAFTMGHIQQCSDPELEDCNRVDADLFQSACVCSSETLVE 100
--T-G-L----I-Q-FA---T-P---D------------------
--T-G-L----I---LA----VK-----T-L---E----------

KSNLKHCCENTAIERTHCFVDHAKFKRPDSLITSEPAADQCEFDFKDRK 150
-NH---M---K-----A---------L----------KA-L---------H---
-ND---M---K-----A---------L----------KA-L---------H---

AFVESFIIPRFSKRTNMLPPHVLAVTKSYGGEVLTCCGKREAATCFNT 200
--GR---K-----S-----Q-----TA*-A------TS--------D--
--GR---K-----S-----P-----V------IAKG-------T--------

KKPTFEHFVNRVNELEKALCVHKNYSGDRFKAKKMQYQSMRFQSQFE 250
--A-----QRA-GK-----T-----R-----K--------VV------L----------
--A-----Q-A-MK-----A-----RS-----K--------VV------LV------

MRGIVYKYIVATTAPCCSGDMITCMKQRKLVDVCQGGRSVMRMAFLKMC 300
--G-M-D------V--------V------E------A-------A-K------WA----SA--
--G-M-D------V--------V------E------T------A------E-------A------SA--

CKEHAIDRGSCVEAMKPDTPKDSLSEHYDLHADIAAVCHTFTKTPDVALG 350
--D-VH-------S-----G---------V--------Q----------M--
--D-VH-------P--------G------I---------Q----------M--

KLIYELSVRHPESSQVILFSKERAEQAEFLQCDKEDHAEVRTALADSN 400
--V-----I------D----------A---------L---------G-D
--V-----I------Q-------L-----A-Q-------L-----M-----K-----G-D

IDKMIadeIIEYYKNNMCAAEAVLGHYNEFKSMMVYYTRIMPQASFQVQLHMV 450
--K-T-----TD------K-----V-TGNMNA-------H--------D---I-
--K-T-----TD------K-----AVSDS--------D------

SETVADVHFCKKDQPGLPCLAEKSTNLDHCTNCDDYSSINPRIAR 500
--R-H---L-A----E-------FI--------L-DSI-A--E-Y-P----------H
--H-L-A------EQ--------L-DAI-A--D-Y-P----------H--H

CCNQSYSMRRCILAQPDAETFPPEDANDFHMPGELCTKNSKDDLSSG 550
---------------------------------M----------SN------
---------------------------------H---------T-------SS--------D------

KLLLYDVVRHKTSSTEDHLKALSANHSTREKCCAAEDKEACTFDEAALK 600
------G------T-------EQ-NSI-TKYHS-------QA------E--P--
------G------T-------TI-TKYHTMK-------QA------E--P--

LAESVELVKV
V-----A------A
VS-----A------
Fig. 2-8 Identities between the three fish albumins (figures in brackets are identities on cDNA level)
Fig. 2-9 Schematic representation of the three-domain structure of rtALB2
homologous domains (Brown 1976). In the three fish albumins, all the 36 cysteines are conserved. Fig. 2-9 shows the structural organization of rtALB2 as a representative for fish albumins.

Based on amino acid compositions, molecular masses and isoelectric points were calculated to be 65.2 kD and 5.45 for rtALB1, 66.1 kD and 5.72 for rtALB2, which are similar to those of mammalian albumin (e.g. the calculated molecular mass and isoelectric point is 66.5 kD and 5.6 for human albumin). Fish albumin contain potential N-linked glycosylation sites (Asn-X-Ser/Thr) in their amino acid sequences: all three salmonid albumins have a glycosylation site at 501-503. Mammalian albumins are not glycosylated, while one of the two frog albumins is glycosylated (Haefliger et al. 1989). Glycosylation makes the molecular weight of frog albumin increased to 74 kD. Whether rainbow trout and Atlantic salmon albumin are glycosylated need to be determined by further experiments.

7. Tissue-specific expression of rainbow trout albumin gene

Tissue-specific expression of albumin gene in rainbow trout was examined by Northern analysis. Total RNA from different tissues were hybridized with rtALB2 cDNA probe. Since two albumin RNAs share 80% of identities, and in some regions the identity is as high as 95% over a stretch of 120 nucleotides, the rtALB2 cDNA probe hybridizes to both rtALB1 and rtALB2 mRNA. As shown by Northern blot (Fig. 2-10A), albumin mRNA is easily detected in adult liver, but not in other tissues like eye, gill, heart, intestine, kidney, muscle, spleen, and stomach. Even at less stringent conditions (2 x SET-0.1% SDS-0.2% PPI, at room temperature for 1 hour and at 65°C for 30 min), no signals were detected in other tissues. The rainbow trout albumin genes, as with all other vertebrate genes, are liver specific.

In the Northern hybridization with rtALB2 cDNA probe, three albumin mRNAs of about 2.2 kb, 3 kb and 3.2 kb were detected. When a rtALB1-specific probe, the rtAlb1-
Fig. 2-10  Northern blot of rainbow trout total RNA from different tissues

A. hybridization with rtALB2 cDNA probe; B. total RNA in agarose gel; C. hybridization with rtALB1-probe B. lane 1. eye, 2. gill, 3. heart, 4. intestine, 5. kidney, 6. muscle, 7. spleen, 8. stomach, 9. liver. lane 1-8, 10 ug of total RNA; lane 9, 5 ug of total RNA
probe B, which contains 800 bp of the rtALB1 cDNA at 3' end was used, both the 3.0 kb and the 3.2 kb band were confirmed to be rtALB1 mRNA (Fig. 2-10C). The 3.0 kb mRNA is likely the mature albumin 1 mRNA using the poly A signal at bp 2738-2743 in the cDNA sequence to add its poly A tail. The 3.2 kb mRNA may be due to a longer poly A tail using the same poly A signal as the 3.0 kb mRNA or may be due to another poly A signal further downstream. The human albumin gene has three poly A signals, although only the first two were found to be used in cDNA cloning (Minghetti et al., 1986). The identity of the 2.2 kb band was further analyzed below.

8. Developmental expression of rainbow trout albumin gene

To test the expression of rainbow trout albumin gene in development, total RNA were purified from rainbow trout embryos of 5 weeks and 6 weeks, hatched fry of 8 weeks, and liver of juvenile fish and subjected to Northern analysis. As shown in Fig. 2-11A, albumin mRNA signal is very weak in the 5-week embryo (lane 1 and 2). One week later, the albumin mRNA amount is increased abruptly (lane 3). The small amount of albumin mRNA in 5-week embryo may be the mRNA expressed in yolk sac when liver has not formed, while the high level of albumin mRNA in 6-week embryo may be caused by the formation of liver at this stage and the efficient expression of the albumin gene in liver. At 8 weeks, embryos were hatched. Because the hatched fry were very small, it was difficult to separate its liver from other tissues. In order to test whether tissue-specific expression of albumin gene evolved at an early developmental stage, we cut the hatched fry into three segments: head, abdomen and tail (Fig. 2-11C), and extracted total RNA separately. In Northern hybridization, only the abdomen segment contains albumin mRNA bands, the other two parts do not (lane 4-6 in Fig. 2-11A). The abdomen segment includes liver and other tissues. This experiment provided an indirect evidence that albumin gene expression is restricted to certain tissue(s) even in the early developmental stages. The last lane (lane 7) in Fig. 2-11a shows the high level of albumin mRNA in juvenile fish liver.
Fig. 2-11  Northern blot of rainbow trout total RNA from different developmental stages

A. Northern hybridization with rALB2 cDNA probe; B. total RNA in agarose gel; C. schematic representation of a hatched fry and the three segments used in Northern analysis. lane 1-2, 5-weeks embryos; lane 3, 6-weeks embryos; lane 4-6, 8-weeks hatched fry; (lane 4, head; lane 5, abdomen; lane 6, tail); lane 7, juvenile fish liver. lane 1, 50 ug of total RNA; lane 2-7, 20 ug of total RNA
Fig. 2-12 In situ hybridization of rainbow trout embryos with rtALB2 cDNA (Done by Dr. Z. Gong)

A-B. 4-weeks embryo, C-F. 8-weeks embryo

The section of rainbow trout embryos were probed with $^{32}$P-labelled rtALB2-specific probe. A, C, and E show embryo sections under a microscope. B, D, and F show the in situ hybridization results. E and F are enlargements of a portion in C and D (frame). The rainbow trout albumin mRNA is first detected in yolk sac membrane in 4-weeks embryos when liver has not formed (A and B). In 8-weeks embryos, liver has formed and the albumin mRNA is detected in liver, yolk sack membrane and intestine (C-F).
The Northern result was further confirmed by in situ hybridization. As shown in Fig. 2-12, in the 4-week embryos when liver had not formed, albumin mRNA was detected only in yolk sac membrane; while in the 8-week embryos when liver appeared, albumin gene was highly expressed in liver as well as in yolk sac membrane and some sections of the intestine.

9. *Different levels of rtALB1 and rtALB2 mRNA*

   We have shown in Fig.2-10 that the rtALB2 cDNA hybridized to three bands in the Northern blot and two of them (the 3.0 and 3.2 kb band) are rtALB1 mRNAs using the most 3' poly A signal(s). In order to identify the nature of the 2.2 kb band, two probes specific to rtALB1 (rtALB1-probe A) and rtALB2 (rtALB2-3' probe) were used in Northern blot. The rtALB1-probe A detected all three mRNAs shown in Fig. 2-10, while the rtALB2-3' probe detected a new mRNA band at 2.5 kb, indicating that rtALB1 had three mRNA forms and rtALB2 had only one mRNA form (Fig. 2-13). The rtALB2 mRNA signal was much weaker than the rtALB1 mRNA signals, which accounted for the absence of the 2.5 kb band in Northern blot with the rtALB2 cDNA probe.

10. *Different transcriptional levels of rtALB1 and rtALB2 genes*

   The different levels of rtALB1 and rtALB2 mRNA may be caused by different transcription rates or by different mRNA degradation rates. To test this, nuclear run-on experiments were carried out to examine the relative transcriptional rate of each gene. Nuclei from individual fish were purified and nascent RNA was extended in vitro by using 32P-UTP and non-radioactive ATP, CTP and GTP. The labeled RNA were hybridized to the rtALB1-PCR and rtALB2-PCR fragments immobilized on the hybridization membrane. DNA sequences of the two fragments are shown in Fig. 2-14. Southern blot was carried out to show that they were specific to their own cDNAs (Fig. 2-14). The two fragments had the same length and were located in the same region of each cDNA: hybridization
Fig. 2-13 Different levels of the rtALB1 and rtALB2 mRNA

Northern blot of rainbow trout liver total RNA with rtALB2-3' probe (lane rt-2 probe) and rtALB1-probe A (lane rt-1a probe)
signals from these two fragments would provide a quantitative measurement of the relative transcriptional rate of each albumin gene. The nuclear run-on results showed that rtALB1 had a stronger signal than rtALB2 (Fig. 2-15). NIH Image analysis showed that the rtALB1 gene was transcribed three times higher than the rtALB2 gene, suggesting that the different mRNA levels of rtALB1 and rtALB2 is caused, at least partly, by the different relative transcriptional rates of the rtALB1 and rtALB2 gene.
Fig. 2-14 The probes used in nuclear run-on experiment

A. DNA sequences of rtALB1-PCR and rtALB2-PCR fragment. * represent a deletion. Underlined are primers used in PCR. B. Southern blot to test the specificity of the two probes. The PCR products (left) were hybridized with rtALB1 cDNA (middle) or rtALB2 cDNA (right). Lane 1, rtALB1-PCR fragment; Lane 2, rtALB2-PCR fragment; M, molecular marker.
Liver nuclei from two individual fish were isolated and examined for the relative transcription rates of the rtALB1 and rtALB2 gene. The difference between the relative transcription rates of rtALB1 and rtALB2 gene in each fish was 2.35 and 4.6 fold, respectively, thus the average difference was calculated as 3.5 fold. BST, pBluescript vector, is a negative control.
DISCUSSION

Salmonid is considered to be tetraploid, implying that many if not all proteins are encoded by a pair of genes (Ohno et al., 1968; Allendorf and Thorgaard, 1984). In this work, two rainbow trout albumin cDNA were isolated from the same liver cDNA library. On the cDNA level, these two clones share 81.5% identity and on the amino acid level, they have 77.5% identity, suggesting that they are from two distinct albumin genes. Rainbow trout albumin 1 is 92.6% identical to Atlantic salmon albumin, but rainbow trout albumin 2 is only 81% identical to Atlantic salmon albumin. Therefore rtALB1 is the homologous gene of the Atlantic salmon albumin, while rtALB2 represents a novel fish albumin. Byrnes and Gannon (1990, 1992) have reported two albumin cDNA sequences for Atlantic salmon which share 99% identical nucleotides and most of the differences are in the 3' untranslated region (2.6%). The silent mutation frequency between the two Atlantic salmon albumin cDNAs is 0.98% (Table 2-1), suggesting that they are resulted from allelic variations of the same gene, but not from two different genes. This situation has been reported for coho salmon growth hormone cDNAs (Nicoll et al., 1987; Gonzales et al., 1988; Agellon et al., 1988). Interestingly, rtALB2 is more than 20% different from rtALB1, which is comparatively higher than other gene pairs. For example, the two rainbow trout growth hormones are only 5% different (Agellon et al., 1988; Rentier-Delrue et al., 1989) and the two trout metallothioneins are 3.5% different (Bonham et al., 1987). It seems that albumin gene is less conserved than other genes. According to our genomic Southern result, there should be one or two rtALB2 gene copies in the genome (Chapter 3). If there was one rtALB2 gene, there should be one rtALB1 gene in the genome and the two albumin cDNAs isolated in this work would represent the two albumin genes caused by the tetraploidization event. If there were two rtALB2 genes, we would expect two rtALB1 genes in the genome as well, thus in total of four albumin genes would exist in rainbow trout. In this case, the albumin gene might have been duplicated once before the
tetraploidization event to give rise to the ALB1 gene and the ALB2 gene, while the
tetraploidization event occurred later and caused the formation of two albumin 1 genes
(ALB1a and ALB1b) and two albumin 2 genes (ALB2a and ALB2b). The two hypothesis
for rainbow trout albumin gene evolution is summarized in Fig. 2-16. In order to
understand the evolution of the rainbow trout albumin genes, we compared the silent
mutation and replacement mutation of several pairs of salmonid genes. As shown in Table
2-1, the silent mutation between rtALB1 and rtALB2 is about 25%, which is higher than
that between the two growth hormone genes (8-12%) but is similar as that between the two
metallothionein genes (27%). Therefore, whether the separation of ALB1 and ALB2 genes
occurred before or after the tetraploidation event remains unclear.

In our Northern blot, both rainbow trout albumin 1 and albumin 2 are detected only
in adult liver, not in any other tissues tested. This result is in accordance with the
expression of mammalian albumin gene which is highly liver-specific. Powell et al. (1984)
detected mouse albumin mRNA only in liver by Northern blot and nuclear run-on. In fact,
the mammalian albumin gene has been used as a model to study liver-specific gene
transcription (Tronche et al., 1990). Atlantic salmon albumin gene expression is detected in
muscle as well as in the liver (Byrnes and Gannon, 1990), the mRNA level in muscle is
3% of that in liver. Since in their Northern experiments, filters were washed at very low
stringent condition (twice for 15 min in 0.6 M NaCl at room temperature, twice for 15 min
in 0.1 M NaCl with 1% SDS at 37°C and then rinsed in 0.1 M NaCl at room temperature),
some non-specific binding may have accounted for this observation. It is necessary to
further confirm the tissue-specific expression of Atlantic salmon albumin gene using more
stringent washing conditions.

In this study, we also tested the developmental expression of rainbow trout albumin
genes. Our results showed that rainbow trout albumin mRNAs were first expressed at very
low level (before 5 weeks) and then the expression was increased abruptly (after 6 weeks).
Leod and Cooke (1989) found that rat albumin expression in yolk sac was about 100-1000
Table 2-1 Comparison of silent mutation and replacement mutation of some genes in salmon.

<table>
<thead>
<tr>
<th>gene comparison</th>
<th>silent mutation</th>
<th>replacement mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>asALB1:asALB2(608aa)</td>
<td>0.98%</td>
<td>0.64%</td>
</tr>
<tr>
<td>rtALB1:rtALB2(607-608aa)</td>
<td>25.25%</td>
<td>12.11%</td>
</tr>
<tr>
<td>rtGH1:rtGH2(210aa)</td>
<td>8.25%</td>
<td>2.27%</td>
</tr>
<tr>
<td>chGH1:chGH2(210aa)</td>
<td>12.40%</td>
<td>2.68%</td>
</tr>
<tr>
<td>rtMT1:rtMT2(60-61aa)</td>
<td>27.03%</td>
<td>2.10%</td>
</tr>
</tbody>
</table>

as, Atlantic salmon; rt, rainbow trout; ch, chum salmon; ALB, albumin; GH, growth hormone; MT, metallothionein.
Fig. 2-16. Two hypothesis of rainbow trout albumin gene evolution

A. The rtALB1 and rtALB2 genes were caused by tetraploidization event (T);
B. The rtALB1 and rtALB2 genes were caused by gene duplication event (D) before the tetraploidization event. A, common ancestral gene.
times lower than that in liver. Chicken albumin mRNA is first detectable in yolk sac on the third day of egg incubation, while liver albumin mRNA is detectable at day 6.5 of egg incubation (Dimattia and Lazier, 1993). It is possible that the albumin gene is first expressed in yolk sac in rainbow trout embryos when the liver has not formed. After the liver is formed, the albumin gene is mainly expressed in liver in high efficiency. In our in situ hybridization, embryos of 4 weeks have not developed livers and albumin mRNA is found in yolk sac membrane, while in embryos of 8 weeks, the liver has formed and albumin is expressed in much higher level in liver than in yolk sac. Combining the Northern and in situ hybridization results, we can conclude that rainbow trout liver appears between the 4th and 8th week during embryonic development and the albumin gene expression is highly increased after liver formation. The expression of albumin in fetal intestine is also detected in rat by Northern hybridization (Sell et al., 1985).

Rainbow trout albumin 1 and albumin 2 genes are transcribed differently as detected by Northern hybridization and nuclear run-on experiments. It has been noticed that, in salmonids, two genes for the same protein type may have different transcription activities. When screening a cDNA library, Sekine et al. (1985) isolated 6 growth hormone I clones and 2 growth hormone II clones from chum salmon; Rentier-Delrue et al. (1989) isolated 7 growth hormone I clones and 4 growth hormone II clones from rainbow trout; and Bonham et al. (1987) isolated 14 metallothionein A clones and 3 metallothionein B clones from rainbow trout. When using chinook salmon prolactin I cDNA as a probe, Xiong et al. (1992) isolated a gene for prolactin II and detected its expression in a Northern blot using a specific probe. The failure to detect a prolactin II cDNA may be due to the lower abundance of Chinook salmon prolactin II with respect to prolactin I. By Northern analysis, Hiraoka et al. (1993) showed that isotosin I mRNA level is higher than isotosin II in sockeye salmon, chum salmon, masu salmon and rainbow trout. However, the pattern of two vasotosin mRNA levels does not coincide among species. The mRNA level of vasotosin I is higher than that of vasotosin II in masu salmon and in rainbow trout, while the opposite
situation is found in sockeye salmon. In this study, we isolated four albumin 1 cDNA clones and one albumin 2 clone. Due to the high divergence between the two albumin cDNA sequences, we were able to design specific probes to test their mRNA levels in Northern blot and the results showed that the albumin 1 mRNA level is about ten times higher than that of albumin 2. Such difference can be caused by different transcriptional rates or by different RNA degradation rates. To further investigate these possibilities, we carried out nuclear run-on experiments and found that the two albumin genes were transcribed differently. The relative transcriptional rate of albumin 1 gene was three times higher than that of albumin 2 gene. This work provides the first evidence that the two genes for one protein in salmonid can be transcribed differently. In Atlantic salmon, only albumin 1 was isolated from cDNA library screening (Bynes and Gannon, 1990 and 1992). The failure to get albumin 2 cDNA may imply that the albumin 2 gene is also less active than albumin 1 gene in Atlantic salmon. The mammalian albumin has only one single gene copy which is actively expressed to produce large quantity of serum albumin. It is possible that the expression from one rainbow trout albumin gene is sufficient to provide enough albumin, such that the second gene is not under pressure to maintain expression and its regulatory region may have changed to a lower activity. The two rainbow trout albumin genes may represent two different transcription regulation mechanisms. Mammalian albumin genes are regulated by their proximal promoters within -170bp (Gorski et al., 1986; Heard et al., 1987) and upstream enhancer (Pinkert et al., 1987; Frain et al., 1990; Hayashi et al., 1992). How the two rainbow trout albumin genes are regulated is still unknown. It will be of interest to isolate the two albumin genes and to study their transcription regulation.
Chapter Three

Isolation and Characterization of Two Rainbow Trout Albumin Promoters
INTRODUCTION

The albumin gene is almost exclusively expressed in the adult liver and the regulation of its expression occurs at the transcriptional level. *In vivo* and *in vitro* studies of rat and mouse albumin genes showed that the DNA sequences necessary and sufficient for tissue specificity resided within the 150 bp preceding the cap site. This sequence contains a TATA box and five other *cis*-acting elements for liver-specific or ubiquitous transcription factors such as HNF1, C/EBP, DBP, NFY and NF1 (Lichtsteiner et al., 1987; Herbomel et al., 1989; Tronche et al., 1990). Comparison of the promoter sequences of albumin genes from rat, mouse, human, bovine, chicken and *Xenopus* (Frain et al., 1990; Power et al., 1994; Schorpp et al., 1988) revealed that all these elements were conserved among species, implying that similar regulatory mechanisms might be used by different organisms. The mouse albumin gene has an enhancer located far upstream (-10.5 to -8.5 kb) (Pinkert et al., 1987) and the human albumin gene has three enhancers located in the proximal region (-486 to -221) as well as the far upstream region (-1.7 kb and -6 kb) (Frain et al., 1990; Hayashi et al., 1992). In application, the albumin promoter and enhancer have been successfully used to direct the liver-specific expression of target genes in transgenic studies (Pinkert et al., 1987; Heckel et al., 1990; Kuriyama et al., 1991).

So far, mechanisms governing fish albumin gene transcription have not been explored. The previous chapter reported the isolation of two distinct albumin cDNA clones (rtALB1 and rtALB2) from the rainbow trout liver cDNA library and the differential transcription rates of the two albumin genes. We are interested in finding out the reasons why the two albumin genes are expressed differently. Here we summarize the isolation and partial DNA sequencing of the rtALB2 gene from a genomic DNA library. The transcriptional activity of the 5' flanking region up to 1.2 kb was examined for the rtALB2 gene in transient expression system and three functional regions involved in the tissue specificity were identified. The rtALB1 promoter was isolated by PCR and DNA sequence
comparison of rtALB1 and rtALB2 promoters revealed a striking difference in the core promoter regions. Functional analysis showed that the rtALB1 promoter, a TATA box-containing promoter, was more active than the rtALB2 promoter which did not have a typical TATA box. Site-directed mutations on the rtALB2 promoter suggested that it was a TATA-less promoter.
1. Preparation of DNA from rainbow trout testis

0.3-0.6 g of frozen testis of rainbow trout were cut into small pieces and washed in 20 ml of DNA extraction buffer I (10 mM Tris.Cl, pH 8.0, 100 mM NaCl, 10 mM EDTA). The tissues were dispersed in 20 ml of fresh DNA extraction buffer I, followed by the addition of 20 ml of DNA extraction buffer II (buffer I plus 1% SDS and 4% β-mercaptoethanol). The mixture was preincubated at 50°C for 30 minutes and then digested by proteinase K (200 μg/ml) at 50°C for overnight. The mixture was extracted three times with phenol:chloroform: isoamyl alcohol (25:24:1). The aqueous phase was transferred to a dialysis tube and dialyzed against 50 mM Tris.Cl and 10 mM EDTA (pH 8.0) at 4°C until the OD260 of the solution is less than 0.05. The DNA was precipitated with absolute ethanol and dissolved in TE (10 mM Tris.Cl and 1 mM EDTA, pH 8.0).

2. Restriction digestion and fractionation of high molecular weight DNA

50 μg of high molecular weight DNA were partially digested by a restriction enzyme MboI (!GATC). The first two nucleotides of MboI sites were filled with dGTP and dATP by Klenow fragment, which left 5’-GA overhangs to match the partially fill-in XhoI sites (C!TCGAG). The digested DNA was precipitated, dissolved in 70 μl of TE and loaded on two Beckman SW40 polyallomer tubes that contained 10-40% sucrose gradient solution (prepared in 10 mM Tris.Cl, 10 mM NaCl, 1 mM EDTA, pH 8.0). DNA was fractionated by centrifugation at 22,000 rpm for 22 hours at 20°C in a Beckman SW40 rotor. After centrifugation, DNA fractions were collected from the tubes and the size of each fraction was examined by electrophoresis on 1% agarose gel. DNA fractions with sizes of 9-23 kb were pooled together and dialyzed against TE buffer. Finally, DNA was precipitated and dissolved in water for genomic DNA library construction.
3. Construction and screening of genomic DNA library

The rainbow trout genomic DNA library was constructed by using the Predigested Lambda FIX II vector cloning kit (Stratagene). The Lambda FIX II vector provided in the kit has been digested by XhoI (C!TCGAG) and followed by partial fill-in with dTTP and dCTP, which creates a 5'-TC overhang. This vector can accept genomic DNA inserts that have been prepared by MboI partial digestion and filled-in with dGTP and dATP. 0.5 ug of fractionated rainbow trout DNA was ligated to 1 ug of vector arms for overnight and packaged in Gigapack II XL packaging extracts (Stratagene). Both Lambda FIX II vector and the Gigapack II XL packaging extracts would preferentially select for large inserts of about 9-23 kb. The packaged phage was incubated with freshly prepared plating SRB(P2) cells in 10 mM MgSO4 to let the phage attach to the cells. The mixture was then plated onto NZY plates [1 liter contains 5 g NaCl, 2 g MgSO4·7H2O, 5 g yeast extract, 10 g NZ amine (casein hydrolysate) and 15 g agar]. In total, about 2.5 x 10⁶ recombinant plaques were obtained, representing 5 x 10⁶ recombinants per microgram of DNA.

The SRB(P2) cells were used to enhance the stability of methylated genomic sequences and non-standard DNA structures. One colony of SRB(P2) cells on a NZY/Kanamycin plate (15 mg of Kanamycin per liter) was inoculated into LB broth supplemented with 0.2% maltose and 10 mM MgSO4 and cultured at 30°C for overnight with vigorous shaking. Cells were harvested by centrifugation on the second day and gently resuspended in 0.5 volumes of 10 mM MgSO4. The cells were diluted to OD₆₀₀ = 0.5 with 10 mM MgSO4 before use.

To screen the genomic DNA library, bacteriophage were plated on NZY plates at a density of 0.5-1 x 10⁵ plaques per 100 mm plate. Plaques were transferred to Colony/Plaque Screen Membranes (NEN, Du-Pond) or Hybond-N membranes (Amersham) in duplicates. The probe used in genomic DNA library screening was the rtALB2 cDNA insert released by EcoRI and XhoI digestion and purified by electroelution. About 0.2-0.5 ug of rtALB2 insert was labeled with α-3²P-γCTP (3000Ci/mmol, Du-
Pond or Amersham) to a specific activity of $2-4 \times 10^8$ cpm/ug by using the Random Primer Labeling Kit (BRL). $1 \times 10^6$ cpm of the labeled probe was added to every one milliliter of hybridization buffer. The hybridization and washing condition was the same as that in Northern hybridization (see Chapter 2). Secondary and tertiary screening were carried out until a single plaque was distinguished and isolated.

4. Purification of bacteriophage DNA

Preparation of bacteriophage was essentially as described by Sambrook et al. (1989). A single phage plaque was cored out and incubated in 1 ml of SM buffer (1 liter contains 5.8 g NaCl, 2.0 g MgSO4·7H2O, 50 ml 1M Tris-Cl, pH 7.5, 5 ml 2% gelatin) at 4°C for several hours. This phage stock was amplified by passing through bacterial cells on NZY plates and collected in SM buffer. To purify bacteriophage DNA, $0.2-2 \times 10^8$ pfu of phage stock were mixed with 3 ml of SRB(P2) cells (containing $4 \times 10^9$ cells, 1 OD600 = $8 \times 10^8$ cells) and cultured in 200 ml of prewarmed NZY media for overnight. On the second day, 4 ml of chloroform was added to the bacteriophage lysate followed by incubation with 1ug/ml of DNase I and RNase A at room temperature for 30 minutes. Solid NaCl was added into the lysate to a final concentration of 1 M (11.7 g per 200 ml of culture) and the mixture was kept at 0°C for 1 hour. The lysate was centrifuged at 10,000 rpm for 10 min (at 4°C). The supernatant was kept and phage particles were precipitated by adding 20 ug of PEG (MW = 8,000, final concentration of 10%) and incubated at 0°C for more than 1 hour. The phage particles were collected by centrifugation and resuspended in 4 ml of SM buffer. The PEG and cell debris was extracted with an equal volume of chloroform. The solution was treated with 200 ug of proteinase K in 20 mM EDTA and 0.5% SDS at 65°C for 1 hour, followed by extraction with phenol:chloroform:isoamyl alcohol. DNA was then precipitated with isopropanol and dissolved in 600 ul of TE buffer for RNase A digestion (5 ug of DNase-free RNase A). The bacteriophage DNA was further
purified with PEG. Finally, the phage DNA was dissolved in water and the concentration was calculated by measuring the UV absorbency at 260 nm (1 OD$_{260}$ = 50 µg/ml).

5. *Restriction mapping of genomic DNA clone*

To map the large DNA insert in the Lambda FIX II vector, phage DNA was digested by NotI to release the insert and the flanking T3 and T7 promoter sites (Fig. 3-1). The insert DNA was then subjected to partial digestion with SstI to produce a collection of DNA fragments. The digested DNA fragments were separated by electrophoresis and transferred to hybridization membrane in duplicates for Southern analysis with $^{32}$P-labeled T3 and T7 primers. By measuring the size of each hybridized fragment, the locations of SstI sites relative to the T3 or T7 promoter were determined. The maps obtained with T3 and T7 primer hybridization were complementary to each other, which confirmed the position of SstI restriction sites.

6. *Subcloning of albumin genomic clones*

SstI digestion of genomic clone 4 and 16 produced 5 fragments for each clone. The size of each fragment was 1.5 kb, 2.3 kb, 2.7 kb, 3 kb and 7 kb for clone 4: for clone 16 was 1.2 kb, 1.5 kb, 2.3 kb, 6 kb and 7 kb (Fig. 3-2). These fragments were purified by electroelution separately and subcloned to pBluescript vector on the SstI site. Subclones were designated as g4-1 to g4-5 and g16-1 to g16-5 in increasing fragment size for clone 4 and 16, respectively. Subclone g16-5 was further cut by HindIII to three fragments (Fig. 3-3) with sizes of 1 kb, 2 kb and 4 kb. The 2 kb fragment was attached to the original vector and self-ligation to produce the g16-5H2k subclone. The other two fragments were ligated to HindIII-digested pBluescript vector to create subclones g16-5H1k and g16-5H4k.
Fig. 3-1 Strategy of restriction mapping of genomic DNA clones
Fig. 3-2 Restriction maps of genomic clone 4 and clone 16
Fig. 3-3 Subclone strategy for clone g16-5

Plasmid DNA of clone g16-5 was digested by HindIII to three fragments, 1kb, 4kb and 5kb, each of them was then purified by electroelution. Self ligation of the 5kb fragment created the g16-5H2k subclone; the other two fragments were ligated to pBluescript vector to creat the g16-5H1k and g16-5H4k subclones.
7. DNA sequencing of albumin gene

In order to analyze the 5' flanking region of rainbow trout albumin genes, the subclone g16-5H2k which had been shown to contain the transcription start site (see result) was sequenced. Plasmid of g16-5H2k was digested by ApaI and SalI separately to create a 3' protruding end (ApaI site) on the Bluescript vector and a 5' protruding end (SalI site) on the insert. The insert was then subjected to exonuclease III (ExoIII) digestion (Double-stranded Nested Deletion Kit, Pharmacia). Aliquots of the digestion were stopped at time intervals to get a series of deletions of about 100-150 bp from the 5' protruding end (Fig.3-4). Exonuclease III digested the 5' protruding strand on the insert, while the other strand and the 3' protruding end of the vector was digested by S1 nuclease to blunt the two ends. The linearized DNA was then self ligated to circular DNA and used to transform to DH5α bacterial cells. DNA inserts of desired sizes were selected by PCR with T3 and T7 primers. To do this, bacterial cells from one colony were dispersed in 25 ul of 10 mM NaOH, boiled for 3 minutes, and centrifuged. 5 ul of supernatant was used to set a 50 ul of PCR reaction. The PCR products were examined on 1% agarose gels and clones with desired sizes were amplified for plasmid purification. All of the subclones obtained by nested-deletion were sequenced with T3 primer.

8. Southern hybridization

10 ug of genomic DNA was digested with restriction enzymes, separated on 0.8% agarose gels, and transferred to Hybond-N nylon membranes (Amersham) under vacuum (VacuGene Pump, Pharmacia). DNA in the gel was first transferred in 0.2M HCl for about 20 minutes to depurinate, then in 0.5M NaOH for 90 minutes. After the transfer, membranes were washed in 2 x SET for 20 minutes and dried in air. Two probes were used in genomic Southern hybridization. One was the rtALB2 cDNA insert released from the Bluescript vector by EcoRI and Xhol digestion, the other was the 1.2 kb 5' flanking region of the rtALB2 gene which was released from the palb2(-1224)CAT construct by
Fig. 3-4 Nested deletions of clone 16-5H2k

M, molecular marker; B, g16-5H2k cut by SstI and HindIII to release the 2 kb insert from the 3 kb pBluescript vector. 0-15, deletion reactions stopped at different time intervals.
HindIII and BamHI digestion. About 0.2 to 0.4 ug of DNA fragments were labeled with 10 ul of $\alpha^{-32}$P-dCTP (3000 or 6000Ci/m mole, Amersham or Du-Pond) in a 50 ul of reaction (Random primer labeling System, BRL). Probes were purified on Nick columns (Pharmacia) and about $1 \times 10^7$ cp m of probes were used per milliliter of hybridization solution. The hybridization solution and the washing condition was the same as that in Northern hybridization (see Chapter 2).

9. Primer Extension

The oligonucleotide primer was synthesized by the Biotechnology Service Center, Hospital for Sick Children, Toronto. 10-20 pmole of oligonucleotide was end labeled with $^{32}$P-$\gamma$ATP (3000Ci/m mole, Amersham) by T4 DNA kinase. About $2 \times 10^6$ cp m of primer was mixed with 40 ug of rainbow trout liver total RNA, which was then extracted with phenol:chloroform:isoamyl alcohol and precipitated with ethanol. The pellet after centrifugation was dissolved in 3 ul of DEPC-treated water, 3 ul of 10 x hybridization buffer [1 x hybridization buffer contains 40 mM PIPES (pH 6.4), 1 mM EDTA, 0.4 M NaCl, 80% formamide] and 24 ul of formamide. The primer and RNA was hybridized at 50°C for overnight. The hybridization mixture was then precipitated with ethanol and dissolved in 30 ul of reverse transcription buffer [50 mM Tris.Cl (pH 7.6 at 37°C), 60 mM KCl, 10 mM MgCl$_2$, 1 mM DTT and 1 mM dNTP, add 1 U/ul of RNAGuard (Pharmacia) before use]. The primer was extended with 1 ul of M-MuLV reverse transcriptase at 37°C for 2 hours. The reaction was stopped by adding 1 ul of 0.5 M EDTA. The RNA was digested by incubating it with 2 ug of RNase A at 37°C for 30 minutes. The reaction was extracted by phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. Finally, the pellet was dissolved in 7 ul of TE and 5 ul of stop buffer [0.3% each Bromophenol Blue and Xylene Cyanol FF; 10 mM EDTA (pH 7.5) and 97.5% deionized formamide]. 4 ul was loaded on a 6% denaturing polyacrylamide gel. As a molecular size
marker, the subclone g16-5H2k was DNA sequenced with the same primer as that used in the primer extension reaction and run beside the primer extension lane.

10. CAT constructs

To make CAT constructs with the 5' flanking region of rainbow trout albumin gene 2, the subclone g16-5H2k was used as a template to amplify two fragments by PCR. Primers used for PCR amplification were alb+40, alb-104 and T7 primers. The alb+40 primer was an antisense oligonucleotide corresponding to the sequence from nt +11 to +40 on the rtALB2 gene, while the alb-104 was a sense oligonucleotide corresponding to the region from nt -104 to -84. The sequences of these two primers are:

alb+40: 5' GGATCTAGCG ATTGAGCCCA AGAGCAGAGA 3'

alb-104: 5' GTGTACTGAG AGACCACTAT 3'

The T7 primer annealed to the pBluescript vector on the 3' of the insert. By using T7 and alb+40 primers, a fragment covering the region from nt -1224 to +40 of the rtALB2 gene was amplified; while by using the alb-104 and alb+40 primers, a fragment covering the region from nt -104 to +40 was amplified. The two PCR fragments were then cloned to pT7Blue vector (Novagen) to create clone palb2(-1224)T7Blue and palb2(-104)T7Blue. The orientations of the inserts were determined by DNA sequencing (Fig.3-5).

The inserts in clone palb2(-1224)T7Blue and palb2(-104)T7Blue were released by BamHI and XbaI digestion and purified from agarose gels. They were then ligated to a promoter-less CAT vector pCAT6 (Boshart et al. 1992) between the SalI and XbaI sites, which created sense CAT constructs. To make deletions on the 5' flanking region of albumin 2 gene, convenient restriction sites were used. On the palb2(-1224)T7Blue plasmid, digestion by HincII or NcoI in combination with XbaI released the -906/+40 and -487/+40 fragments from the albumin 5' flanking region respectively. Then the two fragments were ligated to pCAT6 to construct the palb2(-906)CAT and palb2(-487)CAT. All these CAT constructs included 26 bp from the pT7Blue vector between the albumin
PCR amplification and clone to pT7Blue vector

palb2(-1224)T7Blue

cut by XbaI and BamHI, or HindII, Sphi, NcoI, ligate to CAT6 vector

CAT6 vector

cut by XbaI and BamHI, ligate to CAT6 vector

palb2(-104)CAT

palb2(-487)CAT

palb2(-605)CAT

palb2(-906)CAT

palb2(-1224)CAT

Fig. 3-5 Strategy for rainbow trout albumin 2 promoter-CAT construction

* PstI site is included in all CAT constructs except palb2(-605)CAT.
gene sequence and the CAT coding sequence. This 26 bp DNA sequence contained an ATG codon which might interfere with the correct translation frame. In order to remove the ATG codon, all constructs were cut by NdeI and XbaI and then self-ligated after blunting ends by Klenow. Thus only 8 bp (5' -AATCCATA- 3') from the pT7Blue vector were left between the albumin gene sequence and the CAT coding sequence. From palb2(-1224) CAT, digestion with SphI and self-ligation produced the palb2(-605)CAT construct. The 5' and 3' conjunction of each CAT construct were sequenced with the CAT6-60 primer and the alb-104 primer separately to confirm their identities. The CAT6-60 primer (5' CACTGCATTC TAGTG'TGTGGT TTGTCC 3') was a sense oligonucleotide 60 bp upstream from the polycloning site.

II. Cell culture

Human hepatoma cell HepG2 cells were maintained in minimum essential medium (MEM) supplemented with 20 mM HEPES buffer (pH 7.3), 10% fetal bovine serum, 0.1 mM non-essential amino acids and antibiotic-antimycotic (100 U/ml penicillin G, 100 U/ml streptomycin sulfate, 0.25 ug/ml amphotericin B). Human epithelial cells Hela cells were maintained in MEM supplemented with 20 mM HEPES (pH 7.3), 10% calf serum and antibiotic-antimycotic. Cells were incubated at 37°C with constant CO2. For subculture, cells were washed with 1 x HBSS without magnesium and calcium and treated with 0.05% trypsin for 5 minutes for HepG2 cells and 3 minutes for Hela cells. The digestion was stopped by adding 7-10 ml of culture media and cells were lifted by pipetting, collected by centrifugation and resuspended in fresh media and plated in new plates. To store cells, 1-6 x 10^6 cells were suspended in 1 ml of 7.5% dimethyl sulfoxide (DMSO) prepared in culture media and frozen slowly in -70°C freezer, then it was transferred to liquid nitrogen for long-term storage.
12. Transfection experiments

Plasmid DNA was transfected into HepG2 and Hela cells by the calcium phosphate precipitation method (Graham et al., 1973). About $2.5 \times 10^5$ HepG2 cells or $1.5 \times 10^5$ Hela cells were plated onto one 60 mm plate and cultured for overnight or until they reached 50% confluent. Plasmids used in transfection were prepared with QIAGEN columns (QIAGEN) and 8 ug of testing CAT constructs was used for each transfection. CAT5 and CAT6 were kindly gifts from Dr. G. Sachets, Institute of Cell and Tumor Biology, German Cancer Research Center. To normalize the efficiency of transfection, 2 ug of pCMV-LacZ, a reference plasmid containing the bacterial β-galactosidase gene under the control of human cytomegalvirus (CMV) immediate early promoter, was cotransfected with the test CAT construct. In brief, plasmids DNA were mixed with 31 ul of 2 M CaCl2 in a total volume of 500 ul. The mixture was added into 500 ul of 2 x HBS (1.6 g NaCl, 74 mg KCl, 20 mg Na₂HPO₄, 0.2 g glucose, 1.0 g HEPES per 100 ml, pH 7.05) and left at room temperature for 20 minutes. The DNA precipitates were added into culture media and plates were incubated at 37°C for 5-10 hours before glycerol shock treatment (3 to 5 minutes in 15% glycerol prepared in 1 x HBS). Cells were harvested in 1 x PBS (10 x PBS = 40 g NaCl, 1 g KCl, 1 g KH₂PO₄, 5.75 g Na₂HPO₄ per 500 ml) 36-48 hours later and disrupted by freeze-thawing.

The protein concentrations of cell extracts were measured with Coomassie Brilliant Blue G-250 according to Bradford (1976). The CAT assay was performed by the method of Neumann et al. (1987), in which cell extracts were mixed with chloramphenicol and $^3$H-acetyl-CoA (3.6Ci/mmol, NEN or Amersham) and incubated at 37°C for 2-3 hours. After the reaction was finished, the product $^3$H-acetyl chloramphenicol was extracted into Econofluoro-2 and CAT activity was quantified by scintillation counting. β-galactosidase activity was determined as described by Miller (1972). The CAT activity was divided by the β-galactosidase activity to normalize for transfection efficiency. All transfection experiments were repeated at least three times, and for each construct at least two
independent plasmid preparations were tested. Data was pooled and analyzed using the StatWorks program (Cricket Software, Philadelphia, PA), and the means and standard deviations were calculated.

13. PCR cloning of albumin 1 promoter

The rtALB1 promoter was isolated by PCR using genomic DNA as template. The primers used in PCR were alb-48 and alb+65. The alb-48 primer was the 5' primer and it covered the putative HNF1 binding site which was highly conserved in the 5' flanking region of rtALB2 gene and the Atlantic salmon albumin gene. The alb+65 primer was the 3' primer. It was located in the 5' untranslated region and conserved between rtALB2 and rtALB1. The alb-48 primer corresponded to the region from nt -48 to -31 and the alb+65 was complementary to the region from nt +46 to +65 in the rtALB2 gene. To facilitate subsequent cloning, a PstI site and a XbaI site was included in the alb-48 and alb+65 primer respectively. The sequences of each primer are (underlined is the restriction site):

alb-48: 5' CTGCAGGTATAGTAATAGTATGACAACTATATTGAC 3'
alb+65: 5' GCCTTAGACGAGCGCCGGGATGACGCTGATGGA 3'

To amplify the rtALB1 promoter, 0.1 ug of genomic DNA was used in a 100 ul PCR reaction. The PCR reaction gave rise to a band at about 100 bp. The PCR products were purified by electroelution and cut by PstI and XbaI, which were then ligated to the promoter-less CAT vector (CAT6) predigested with PstI and XbaI. After transformation in DH5α cells, plasmids from different clones were purified and DNA sequenced. DNA sequences showed that these PCR clones fell into two groups: the rtALB1 promoter and the rtALB2 promoter. Both promoters were transfected into HepG2 and Hela cells.

14. Site-directed mutagenesis

To introduce mutations on the albumin promoter, two oligonucleotide primers, the mutagenic primer and the selection primer, were simultaneously annealed to denatured CAT
plasmid by using the Transformer Site-Directed Mutagenesis Kit (Clontech Laboratories, Inc.). The selection primer mutated the SmaI and KpnI restriction sites on the CAT vector. Mutagenic primers introduced the desired mutations into the promoter. Sequences of each primer is (the bold letters indicate the mutation sites):

selection primer: 5' ATCGATCCCC GGATCCCGAG CTCGAAT 3'
mA primer: 5' TTCTACAAAA GTGCAAAAGA GCACCTGTC 3'
mB primer: 5' GATTGACAAC TGGCAAATTC TACAAAAGTA 3'
mC primer: 5' GATTGACAAC TGAAGCATTC TACAAAAGTA 3'
mD primer: 5' GCTAATGATT GACAATGTA TAATTCTACA 3'

The mutagenic primer and the selection primer were annealed simultaneously to one strand of the denatured palb2(-487)CAT or palb2(-104)CAT plasmid. Then they were elongated and ligated by T4 DNA polymerase and T4 DNA ligase. The mixture of mutated and unmutated DNAs were used to transform competent BMH cells which were repair-deficient. The transformed cells were cultured in 4 ml of LB/Ampicillin media for overnight and plasmid DNA was prepared from the mixed bacterial population. The isolated plasmid DNA was digested by KpnI, which linearized the wild-type plasmids but left the mutated plasmids intact. The digested DNA was then transformed into DH5α cells. The colony population was greatly reduced after the second transformation. Colonies were picked randomly and plasmid purified from each colony was sequenced to select the mutants.
RESULTS

1. Isolation of rainbow trout albumin genomic clones

The rainbow trout genomic DNA library was constructed on Lambda FIX II vector as described in the Materials and Methods. From 0.5 ug of fractionated DNA, a total of 2.5 x 10^6 pfu were obtained, representing 5 x 10^6 pfu/ug. This library was screened using rtALB2 cDNA and seven positive clones were isolated from about 5 x 10^5 plaques of primary library. Phage DNA from each positive clone was purified, cut by restriction enzymes, and separated on agarose gel (Fig. 3-6). The insert size of each clone was estimated by the sum of the lengths of each restriction fragment. As shown in Table 3-1, the sizes of the inserts vary from 8 kb to 18 kb.

2. Southern analysis of positive genomic clones

The positive genomic clones were restriction digested and subjected to Southern analysis. Probes used in the Southern analysis included the rtALB2 cDNA and the 5' and 3' albumin probes (Fig. 3-6). The 5' probe was the first 270 bp of rtALB2 cDNA released from subclone rtALB2-5'NcoI(0.2kb) and the 3' probe was the last 280 bp released from subclone rtALB2-5'AccI(2.0kb). All theses genomic clones hybridized with the cDNA probe. Clone 4 and 16 hybridized with the 5' albumin probe as well as the cDNA probe (Fig. 3-6a and b), suggesting that they might contain the transcription start site and the 5' flanking region of the albumin gene. Clone 8 hybridized with the 3' probe and might contain sequences 3' of the gene (Fig. 3-6c). The other four clones, clone 2, 5, 11 and 26, hybridized only with the cDNA probe (Fig. 3-6d-g), indicating that they were located somewhere in the middle of the gene. Since none of these clones could hybridize with all three probes, none of them was a full-length gene. Human albumin gene spans 17 kb from the Cap site to the first poly(A) signal (Minghetti et al., 1986). The longest clone we have,
Table 3-1  Summary of Southern results of rainbow trout albumin genomic clones

<table>
<thead>
<tr>
<th>genomic clones</th>
<th>insert length (kb)</th>
<th>Southern results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cDNA probe</td>
<td>5' probe</td>
</tr>
<tr>
<td>g2</td>
<td>17.5</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>g4</td>
<td>16.5</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>g5</td>
<td>15</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>g8</td>
<td>10</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>g11</td>
<td>8</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>g16</td>
<td>18</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>g26</td>
<td>11</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>
Fig. 3-6 Southern analysis of positive genomic clones for trout albumin gene

The phage DNA purified from each positive genomic clones were digested by different restriction enzymes, separated on agarose gel (left), transferred to Nylon membranes and hybridized with the rtALB2 cDNA probe (middle in A-C, right in D-G), the 5' probe (right in A and B), and the 3' probe (right in C). Clone g4 and g16 (A and B) could hybridize with the cDNA probe and the 5' probe; clone g8 (C) could hybridize with the cDNA probe and the 3' probe; clone g2 (D), g5 (E), g11 (F) and g26 (G) could hybridize with only the cDNA probe.
C
clone 2, is 17.5 kb in size yet it could not hybridize with the 5' probe nor the 3' probe, it is possible that the rainbow trout albumin gene is longer than 17 kb.

3. Determination of the 5' ends of clone 4 and clone 16

In order to find out the 5' flanking region of albumin gene, the two clones, clone 4 and clone 16, which might contain the 5' flanking region were further analyzed to determine their 5' ends. First of all, the two clones were restriction mapped with SstI. SstI digestion gave rise to 5 fragments for each clone, suggesting four SstI sites in both clones. The size of each SstI fragment for clone 16 is 1.2 kb, 1.5 kb, 2.3 kb, 6 kb and 7 kb. Similarly, the size of each SstI fragment for clone 4 is 1.5 kb, 2.3 kb, 2.7 kb, 3 kb and 7 kb. The SstI fragments were designated as g16-1 to g16-5 and g4-1 to g4-5 in increasing fragment size. To find out the order of each fragment along the gene, phage DNA were partially digested with SstI and analyzed by oligonucleotide T3 and T7 hybridization as described in Materials and Methods. For clone 16 (Fig. 3-7a), 5 bands were hybridized with T7 primer at 1.2 kb, 3.5 kb, 10.5 kb, 12 kb and 18 kb. The four SstI sites relative to the T7 promoter could be determined at 1.2 kb, 3.5 kb, 10.5 kb and 12 kb. Similarly, the SstI sites relative to the T3 promoter were determined at 6 kb, 7.5 kb, 14.5 kb and 16.8 kb. By comparing the SstI sites relative to the T3 and T7 promoter, the restriction map of clone 16 was determined as shown in Fig. 3-2. Since three fragments, the 7 kb, 2.3 kb and 2.7 kb fragments, could hybridize with rtALB2 cDNA probe (Fig. 3-6b), while only the 7 kb fragment could hybridize with the 5' albumin probe, the 7 kb fragment must be located 5' of the other two fragments. Thus the 5' end of clone 16 is found to be located on the T3 promoter side and the 3' end on the T7 promoter side. The restriction map and orientation of clone 4 were determined similarly (Fig. 3-7b and Fig. 2). As shown in Fig. 2, it appeared that clone 16 and clone 4 overlapped in a large area. Clone 16 was about 3 kb
Fig. 3-7  Restriction mapping of genomic clones g4 and g16.

Phage DNA from g16 (A) and g4 (B) were cut by NotI, then partial digested by SstI, separated on agarose gel (left) and transferred to Nylon hybridization membranes for Southern hybridization with oligonucleotide probes T3 (middle) and T7 (right). M, molecular marker; 1-4, SstI partial digestion stopped at different time points.
longer at the 5' end than clone 4, while clone 4 was about 1.5 kb longer at the 3' end than clone 16.

4. DNA sequencing

Since the 7 kb SstI fragment of clone 16 could hybridize with the 5' albumin probe, it might contain the transcription start site. This fragment was cloned to a plasmid vector, pBluescript, and designated as g16-5. In order to locate the transcription start site, the g16-5 clone was digested by HindIII to three fragments: 2 kb, 1 kb and 4 kb and each fragment was subcloned to pBluescript for further analysis by PCR (Fig. 3-3 and Fig.3-8). In order to find out in which HindIII fragment the transcription start site was located, an antisense oligonucleotide alb+40 was synthesized corresponding to the nt 6 to nt 35 of the rALB2 cDNA (which was the sequence from nt +11 to +40 in the albumin 2 gene as determined by primer extension, see below). The sequence of this primer is:

   **alb+40**: 5' GGATCTAGCG ATTAGCCTA AGAGGAGA 3'

This primer was used in PCR in combination with T3 or T7 primer which located at either side of the HindIII fragment on the vector. As shown in Fig.3-8, a 1.2 kb PCR band was amplified from subclone g16-5H2k with the alb+40 and T7 primer, indicating that the transcription start site was located in the 2kb-HindIII fragment. This result also indicated that the 2 kb HindIII fragment was the most 5' one in the subclone g16-5, while the 1 kb and the 4 kb HindIII fragments were located in the middle and the most 3' end respectively. Subclone g16-5H2k was then progressively digested and sequenced.

Fig. 3-9 shows the DNA sequence of g16-5H2k. It has 2238 bp in total, including 1224 bp of the 5' upstream flanking region (the transcription start site is determined by primer extension, see below), 196 bp of the first exon, 727 bp of the first intron and 91 bp of the incomplete second exon. The exon sequences are perfectly matched with the rALB2 cDNA sequences in the same region, indicating that it is the rALB2 gene. Since the
To find out which subclone may contain the transcription start site, plasmid from g16-5H2k, g16-5H1k and g16-5H4k were PCR analyzed with T7 and alb+40 primers (1) or T3 and alb+40 primers (2). The alb+40 primer could anneal to the rtALB2 gene corresponding to nt 6 to 35 in the cDNA region; the T7 and T3 primer could anneal to the vector on each side of the insert. A 1.2 kb band was obtained from the g16-5H2k subclone using T7 and alb+40 primers, indicating that the transcription start site may locate in this subclone.
Fig. 3-9 Partial DNA sequence of the rtALB2 gene.

The g16-5H2k was nested deleted for DNA sequencing. The DNA sequence contains 2238 bp in total, including 1224 bp of the 5' upstream flanking region, 196 bp of the first exon, 727 bp of the first intron, and 91 bp of the incomplete second exon. The putative TATA box and the CAAT box are indicated in bold typeface. The transcription start site was determined by primer extension.
+218 TAAATCTATGTGAGTAAATCTACGTGAGTAACTCTGTGAGTAAATCTGTGAGTAAATCT
+278 TGTGAGTAAATCTACGTGAGTAAATCTACGTGAGTAAATCTACGTGAGTAAATCTACGTG
+338 AGTAAATCTATGTGAGTAAATCTACGTGAGTAAATCTACGTGAGTAAATCTATGTGAGTA
+398 AATCTATGTGAGTAAATCTATGTGAGTAAATCTACGTGAGTAAATCTATGTGAGTAAATC
+458 TATGTGAGTAAATCTATGTGAGTAAATCTACGTGAGTAAATCTACGTGAGTAAATCTACG
+518 TGTGAGTAAATCTACGTGAGTAAAGCTGTAATAAAAAAGCTTTTTATTTAAATCACTGCCCTTTGT
+578 CAAACTCCTGAGTGGCTGACCTCAGATTTTTAACCACCTTAATACTGTGTTTTTTTACA
+638 TGGATTTGTATTTGTCTATTTTTTCTCTGCTGATTTAAAAGGGGT
+698 AACAGCGACGGTTAGTGAATAATTTTGGTGAAGTGACGGTACGACCTACTGTCTGTGTTGT

EXON 2

+758 GTGTGTGTGTTGTGTGTTGTGTGTTGTGTGTTGTGTGTGTGTGTGTTGT

Gly Leu Ala Gly Asn Leu Pro Asp Ser Leu Leu Asp Asp Val Val
+812 GGG CTT CAG AAC CTT CCA GAT AGT TTG TTG GAT GAC GTG GTG

Pro Leu Val Ala Glu Ala Phe Thr Met Gly Ile Gln
+857 CCC CTG GTT GGA GAA GCC TTC ACC ATG GGC ATC CAA

123
genomic clone 16 contained 7.5 kb DNA sequence 5' upstream of the g16-5H2k fragment (Fig. 3-2 and Fig. 3-3), clone 16 has, therefore, about 8.7 kb of the 5' flanking region in total. Genomic clone 4 was also partially sequenced and DNA sequences confirmed that it was the same gene as clone 16.

5. Determination of transcription initiation site

The transcription initiation site was determined by primer extension experiment using an antisense primer, the alb2-extn primer, which corresponds to rtALB2 cDNA at base 89-118. As shown in Fig. 3-10, an extended band terminated at the G residue 94 bases upstream from the primer, thus the transcription start site for rtALB2 gene is the guanine 83 bases upstream from the ATG translation codon. Preceding the start site is an adenine (A) residue. This structure is similar as those seen in other albumin genes, such as rat, mouse and human albumin genes, where the transcription starts from a guanine residue and with an adenine residue at the -1 position (Heard et al. 1987). The rtALB2 cDNA described in Chapter 2 started from the cytosine at +6, so the real rtALB2 transcripts should be 5 bp longer than our cDNA clone.

6. Southern hybridization

Fig. 3-11 shows the result following Southern hybridization of rainbow trout genomic DNA using rtALB2 cDNA as probe. Multiple bands were detected in the autoradiograph as a result of complicated exon and intron distribution and possible cross hybridization between the rtALB1 and rtALB2 genes. In order to find out how many gene copies the rtALB2 gene has, the 1.2 kb 5' flanking region of the rtALB2 gene was used as a probe in Southern blot (see Materials and Methods). This region was used because both its DNA sequence and restriction map was known. In addition, since it is located upstream of the cap site, there are no introns to interrupt the sequence. From the cDNA sequences,
Fig. 3-10 Primer extension mapping of the transcription start site of the rtALB2 gene.

The primer used in this experiment was the rtALB2-extn (5' ACAGCAGCACC-AACAGGCTACAGACAGCCA 3'), which corresponded to the rtALB2 cDNA at base 89-118. A similar result was also obtained with the oligonucleotide rtALB2+54 (5' GAGGGATGGATGGAAGAGGGGATCTAGC 3'), which corresponded to the rtALB2 cDNA at base 27-54. (The arrow denotes the transcription start site).
Fig. 3-11 Southern hybridization of genomic DNA with rtALB2 cDNA probe
we have noticed that the 5' untranslated region is very divergent between rtALB2 and rtALB1 (about 65% identity). It is reasonable to speculate that the 5' flanking region of rtALB2 and rtALB1 genes may not share high identities and the 1.2kb 5' flanking region probe would hybridize with the rtALB2 gene only. The 1.2 kb 5' flanking region dose not contain restriction sites for EcoRI, HindIII, PstI and SstI. When genomic DNA was cut by these restriction enzymes and subjected to hybridization with the 5' flanking region probe (Fig. 3-12), one band was detected in HindIII-digested DNA and two bands were detected in EcoRI-, PstI-, and SstI-digested genomic DNA, suggesting that rtALB2 may have one or two gene copies. The rainbow trout used in this study was from a fish farm and high polymorphism have been observed for this closed population (unpublished data in our laboratory). The two bands detected in the EcoRI-, PstI- and SstI-digested DNA might be resulted from allelic variations of the same gene. The 7 kb band in the SstI-digested DNA corresponds to the g16-5 fragment characterized in this work.

7. Sequence of the 5'-regulatory region

A striking feature of the rtALB2 5'-flanking region comes from the proximal promoter region where no typical TATA box is found in the -30 to -25 region. Instead, a typical TATA box sequence TATAAA is located 3 bp upstream of the transcription start site, while at the -25 region, an AT-rich element AAAAAATT (from bp -25 to -19) is found. All albumin and AFP promoters isolated to date have a typical TATA box located at around bp -25 to -30, while an atypical TATA box was found in human and rat DBP promoters in which a TGTAAA element replaces the canonical TATAAA sequence. A summary of elements in the proximal promoter region of other albumin, AFP and VDBP genes is shown in Table 3-2. It seems that the rtALB2 gene promoter structure is unique as compared to other albumin promoters. It is of interest to find out which element, the AT-
Fig. 3-12 Southern hybridization of genomic DNA with the rtALB2 5' flanking probe
Table 3-2 Comparison of elements in proximal promoter regions of ALB, AFP, and DBP genes

Element sequences are taken from Ray et al. (1991) and Deryckere et al. (1995). Putative elements are underlined, others are identified regulatory elements.

<table>
<thead>
<tr>
<th>Genes</th>
<th>HNF-1</th>
<th>CAAT</th>
<th>TATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>human</td>
<td>TAGTTAATAATCTACAA</td>
<td>CCAAT</td>
<td>TATATTA</td>
</tr>
<tr>
<td></td>
<td>-65</td>
<td>-49</td>
<td>-88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-26</td>
</tr>
<tr>
<td>mouse</td>
<td>TGGTTAATGATCTACAG</td>
<td>CCAAT</td>
<td>TATATTA</td>
</tr>
<tr>
<td></td>
<td>-66</td>
<td>-50</td>
<td>-88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-27</td>
</tr>
<tr>
<td>rat</td>
<td>TGGTTAATGATCTACAG</td>
<td>CCAAT</td>
<td>TATATTA</td>
</tr>
<tr>
<td></td>
<td>-67</td>
<td>-51</td>
<td>-89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-28</td>
</tr>
<tr>
<td>Xenopus</td>
<td>AGGTTAATAATTTTCCA</td>
<td>NF</td>
<td>TATAAA</td>
</tr>
<tr>
<td></td>
<td>-67</td>
<td>-51</td>
<td>-32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-26</td>
</tr>
<tr>
<td>Atlantic</td>
<td>CAGCTAATGATTGACAG</td>
<td>-</td>
<td>TATAAAAG</td>
</tr>
<tr>
<td>salmon</td>
<td>-60</td>
<td>-45</td>
<td>-31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-27</td>
</tr>
<tr>
<td>rainbow</td>
<td>TAGCTAATGATTGACAA</td>
<td>CAAAT</td>
<td>TATAAAG?</td>
</tr>
<tr>
<td>trout 2</td>
<td>-45</td>
<td>-29</td>
<td>-69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-2</td>
</tr>
<tr>
<td>AFP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>human</td>
<td>AGGTTACTAGTTAACAG</td>
<td>CCAAC</td>
<td>TATAAAA</td>
</tr>
<tr>
<td></td>
<td>-61</td>
<td>-45</td>
<td>-69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-21</td>
</tr>
<tr>
<td>mouse</td>
<td>AGGTTACTAGTTAACAG</td>
<td>NF</td>
<td>TATAAAA</td>
</tr>
<tr>
<td></td>
<td>-64</td>
<td>-48</td>
<td>-30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-24</td>
</tr>
<tr>
<td>rat</td>
<td>AGGTTACTAGTTAACAG</td>
<td>CCACT</td>
<td>TATAAAA</td>
</tr>
<tr>
<td></td>
<td>-63</td>
<td>-47</td>
<td>-70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-23</td>
</tr>
<tr>
<td>DBP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>human</td>
<td>AGATTAATAATGATTA</td>
<td>NF</td>
<td>TGTA AAA</td>
</tr>
<tr>
<td></td>
<td>-65</td>
<td>-49</td>
<td>-29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-24</td>
</tr>
<tr>
<td>rat</td>
<td>AGATTAATAATGATGA</td>
<td>CAAAT</td>
<td>TGTAAG</td>
</tr>
<tr>
<td></td>
<td>-61</td>
<td>-45</td>
<td>-91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-19</td>
</tr>
</tbody>
</table>
1. ANF site: 
rtALB2
   -814/-805: ATTTATCTGA (8/10)
   -465/-456: ATTTATCTTT (7/10)

2. C/EBP consensus: TKNNGYAAK 
rtALB2
   -45/-36: TTAGCTAAT (8/9)
   -67/-59: ATGTGAAAG (8/9)
   -133/-125: CTAAGAAAA (7/9)
   -347/-339: GGTGGAAT (7/9)
   -1204/-1196: TGTGGGAAG (8/9)

3. DBP consensus: TGATTTTGT 
rtALB2
   -476/-468: TTATTTTGT (8/9)
   -671/-663: TTATTTTAT (7/9)
   -896/-890: TTATTTGTGT (7/9)
   -1037/-1029: TCAGTTTGT (7/9)

4. HNF1 consensus: RGTTAATNATTAACM 
rtALB2
   -44/-33: AGCTAATGATTGACA (13/15)

5. HNF3 consensus: TATTGAYTTWG 
rtALB2
   -619/-609: TATTGACAAAG (9/11)
   -40/-30: TAATGATTGAC (8/11)
   -754/-764: CAAAATGAAGT (8/11) (antisense)

6. NFY consensus: YNNNNNRRCCAATCANYK 
rtALB2
   -377/-359: TATACAGGTACAATCAGTT (17/19)
   -192/-174: CCAAGAATAGCAATCATTT (17/19)
   -55/-37: GAGAGAGGTTAGCTAAT (15/19) (antisense)

7. SP1 consensus: KRGGCKRRK 
rtALB2
   -157/-149: GAGGTTGGA (7/9)

FIG. 3-13 Potential DNA binding sites for some transcription factors in the 5' flanking region of rtALB2 gene

rich element at -25 region or the TATA element close to the cap site, serves as a functional TATA box in the rtALB2 promoter.

Mammalian albumin promoters are well conserved up to 250 bp from the transcription start site (Heard et al., 1987; Power et al., 1994). When comparing the rtALB2 5' upstream sequences with other known albumin promoter sequences, little identities were found. We then searched the 5' regulatory region for consensus sequences that are recognized by known DNA-binding proteins. The sequence revealed many putative binding sites for known transcription factors including liver-specific and ubiquitous factors. Fig. 3-13 shows potential binding sites for various transcription factors. Among them, C/EBP, HNF1, HNF3, and DBP are liver-specific or liver-enriched transcription factors, while Sp1, NFY, ANF are ubiquitous ones. All the transcription factors shown in Fig. 3-13 are positive regulators except for ANF which is a negative regulator of mouse albumin gene (Herbst et al. 1990). In mammalian albumin gene promoters, C/EBP, HNF1, DBP, and NFY have been shown to bind to the proximal promoter and play important roles in the promoter activity (Maire et al. 1989; Lichtsteiner et al. 1987; Herbst et al. 1989, 1990). Potential binding sites for these transcription factors are also found in the 5'-flanking region of the rtALB2 gene. However, the presence of the potential binding sites does not prove their involvement in the regulation of gene expression. To study the relevance of these sequences, the 5' regulatory region was examined by transient expression system (see below).

8. Characterization of the 5' regulatory region

The 5' upstream region was analyzed by measuring its ability to direct a reporter gene, the bacterial chloramphenicol acetyl transferase (CAT) gene, in transient transfection assay in human hepatoma cell line, i.e. HepG2 cells. As a negative control, Hela cells were transfected with the same CAT constructs. pCMV-LacZ construct was included in each transfection as an internal control. All the CAT activities were compared with that of CAT5.
CAT constructs    relative CAT activity

<table>
<thead>
<tr>
<th></th>
<th>HepG2</th>
<th>Hela</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT5</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>palb2(-104)CAT</td>
<td>14.4</td>
<td>0.8</td>
</tr>
<tr>
<td>palb2(-487)CAT</td>
<td>210.1</td>
<td>0.8</td>
</tr>
<tr>
<td>palb2(-605)CAT</td>
<td>194.3</td>
<td>0.7</td>
</tr>
<tr>
<td>palb2(-906)CAT</td>
<td>181.1</td>
<td>0.6</td>
</tr>
<tr>
<td>palb2(-1224)CAT</td>
<td>103.1</td>
<td>1.1</td>
</tr>
<tr>
<td>CAT6</td>
<td>0.17</td>
<td>1.47</td>
</tr>
</tbody>
</table>

Fig. 3-14  CAT activities of albumin-CAT constructs.
a herpes simplex virus (HSV) thymidine kinase (tk) promoter-driven CAT construct, and the promoter activities were expressed as the percentage of the tk promoter activity. As shown in Fig.3-14, the highest CAT activity was achieved with the palb2(-487)CAT construct, which was 210% of the tk promoter activity. Inclusion of further upstream sequences from -487 to -906 did not change the promoter activity significantly, as shown by palb2(-605)CAT and palb2(-906)CAT. However when additive upstream sequences from -906 to -1224 were included, there was a 2-fold decrease in the promoter activity, from 210% to 103%, indicating the presence of a negative element in this region. In our albumin CAT constructs, the shortest one was palb2(-104)CAT. This construct had significant expression in HepG2 cells (14.37% of the tk promoter) but no expression in Hela cells, indicating that the rtALB2 promoter resides in the proximal 100 bp sequences.

All the rtALB2 promoter-CAT constructs were expressed only in HepG2 cells, not in Hela cells, suggesting that the rtALB2 promoter is liver-specific. The activity of the palb2(-487)CAT is about 15-fold higher than that of the palb2(-104)CAT, indicating a positive regulatory element presents between bp -487 and -104. To summarize the CAT assay results, we have defined three functional distinct regions in the 5'-flanking region of rtALB2 gene: (1) a negative element located between bp -1224 and -906, (2) a positive element located between bp -487 and -104, and (3) a promoter spanning the first 100 bp preceding the cap site. This pattern is similar to that seen in human albumin 5'-flanking region where a negative element and an enhancer are located in front of the promoter in the region from bp -673 to -486 and from bp -486 to -221 respectively (Frain et al., 1990).

9. PCR cloning of rainbow trout albumin 1 promoter

In Chapter 2, it was shown that the two rainbow trout albumin genes are expressed differently. Rainbow trout albumin 2 gene is transcribed three times lower than albumin 1 gene. In order to find out the mechanism of governing this differential expression, we set out to clone the rtALB1 promoter. By comparing the promoter region of the rtALB2 gene
Fig. 3-15 PCR cloning of the rtALB1 promoter.

Using rainbow trout genomic DNA as template, alb-48 and alb+65 as primers (see Material and Methods), a 120 bp band was amplified by PCR (right lane). After cloning the PCR products to CAT vector and DNA sequencing, it was found that the PCR products contained both rtALB1 and rtALB2 promoters.
and the Atlantic salmon albumin gene, a potential HNF1 site is highly conserved between the two promoters (Table 3-2). We used the potential HNF1 site as the 5' primer and a sequence identical to the rtALB2 and rtALB1 cDNA as the 3' primer to isolate both rtALB1 and rtALB2 promoters directly from genomic DNA by PCR (see Materials and Methods). The PCR reaction amplified a 120 bp band (Fig.3-15) which was then cloned and DNA sequenced.

The PCR products included both rtALB1 and rtALB2 promoters. Fig.3-16 shows the two rainbow trout albumin promoter sequences in comparison with other albumin promoter sequences. The rtALB1 PCR product has 125 bp. Its transcription start site is determined at the G residue 12 bases upstream from the first nucleotide in the rtALB1 cDNA clone. Thus this PCR fragment contains 62 bp of the 5' flanking region and 63 bp of the transcription region. The rtALB1 promoter has a typical TATA box sequence TATAAA from nt -32 to -27. 14 bp upstream of the TATA box is the potential HNF1 binding site. This structure is representative of mammalian albumin promoters, for example, the human albumin promoter has a TATA box from nt -32 to -25 and a HNF1 binding site 18 bp upstream of the TATA box (see Table 3-2). The rtALB2 PCR product has 113 bp which contains 48 bp of the 5' flanking region and 65 bp of the transcription region. The DNA sequence from the PCR product confirmed the DNA sequence we got from the g16-5H2 clone in the same area.

Comparing the two rainbow trout albumin promoters, it is found that they are basically conserved except for two regions. One region is between the TATA box and the transcription start site where a deletion of 22 bp is found in the rtALB2 promoter, which put the TATA box very close to the transcription start site (only 3 bp apart). The other region is on the -25 region of the rtALB2 promoter in which there are 8 extra nucleotides composed mainly of adenine residues, thus an AT-rich element AAAAAATT was formed around the -25 region. Other than these two regions, the rtALB1 and rtALB2 promoters are
A.

rt2: 5' GTTAGCTAATGATTGACAAGACTGAAAAATCTACAARAAAGTATAAGAA--------
rtl: 5' GTTAGCTAATGATTGACA--------G------ATTCCAARATAGTATAAGGAAGGCCTG

rt2: --------------------------GCCCCCTGCTCTCTGCTTGGGCCTCAATCCCTAGATCCCT
rtl: CTTTTCTCCTCCATCTCCCGTCACATCCCTCCTCCGGTCTCTCC-CC-ATCCCATCCCTCCCT

rt2: CTTCCCATCCATCCCTCCCTCCC 3'
rtl: TCCTCCATCCATCCCTCCCGTCCG 3'

B.

rt2: GTTAGCTAATGATTGACAAGACTGAAAAATCTACAARAAAGTATAAGAA------------------------AG
rtl: GTTAGCTAATGATTGACAAGACTGAAAAATCTACAARAAAGTATAAGGAAGGCCTCCTCCATTCCCTGG
sal: GTCAGCTAATGATTGACAAGACTGAAAAATCTACAARAAAGTATAAGGAAGGCCTCCTCCATTCCCTGG
rat: TGTGGTTAATGATCTACGGTTATTGGTT---AGAGAAGTGAGCGAGTTTCTCTGCA-----CACAG
mou: TAT-GTTATTGG----TT--AAAGAAGmGAGCGAGTCTTTCTGCA-***-CACAG
hum: TCTAGTTAAATTATTGGTT--TT--AAAGAAGTAGTGCTAATTTCCCTCCGTTTGTCcTAG

Fig. 3-16 Comparison of albumin promoters.

A. DNA sequences of the PCR fragments. B. Comparison of albumin promoters

DNA sequences for rainbow trout albumin 2 and albumin 1 gene promoters (rt2 and rtl) are obtained from this work. Those for Atlantic salmon (sal), rat, mouse (mou) and human (hum) albumin gene promoters are taken from Deryckere et al. (1995). The HNF1 binding sites and TATA boxes are underlined. Stars indicate the conserved residues in the HNF1 binding site among different albumin gene promoters. Dashes represent deletions. Arrows indicate the transcription start sites.
aligned well to each other. Considering the difference in the two rainbow trout albumin promoter structures, it is of interest to find out whether they are functionally different.

10. Functional analysis of the two albumin promoters

The rtALB1 and rtALB2 promoter activities were analyzed by linking the two PCR fragments to a CAT reporter gene and testing the CAT activities in transient expression system. Both promoters could drive CAT gene expression in HepG2 cells but not in Hela cells, suggesting that they are tissue-specific promoters (Fig.3-17). The rtALB1 promoter is 3.7 times more active than rtALB2 promoter, 20% versus 5.4%, which coincides with the difference in the transcription rates of the two genes as detected by nuclear run-on experiments (see Chapter 2). It is likely that the different transcription rates observed for the two rainbow trout albumin genes are a result of their different promoter structures, i.e. rtALB1 promoter which has a typical TATA box is more active as compared to the rtALB2 promoter which has an abnormal TATA box.

11. Mutational analysis of the rtALB2 promoter

The rtALB2 promoter is a weaker promoter compare to the rtALB1 promoter. The lower activity of the rtALB2 promoter may be due to its abnormal TATA box. There are two possible TATA boxes in the promoter region: the consensus TATAAA element 3 bp preceding the cap site and the AT-rich element AAAAAATT around -25 region. In order to find out which one is the functional TATA box, we designed site-directed mutations on the rtALB2 promoter. As shown in Fig. 3-18, a pair of GC residues were introduced into the consensus TATA element (mA) and the AT-rich element (mB and mC) separately on the palb2(-487)CAT construct. Transient expression assay showed that the activities of the two mutant promoters were similar to that of the wild type (Fig. 3-19). Mutation A was also introduced into palb(-104)CAT (mA’ in Fig. 3-19) and the promoter activity remained the same as the wild type. It seems that neither the consensus TATAAA element from bp -9 to
-4 nor the AT-rich element around -25 serves as a functional TATA box in the rtALB2 promoter. Thus distinct from other known albumin promoters, the rtALB2 promoter might be a TATA-less promoter.

We attempted to improve the rtALB2 promoter's activity by introducing a typical TATA box into the -25 region on the palb2(-104)CAT construct (mD). mD changed the AT-rich element to TATAATT, which is a canonical TATA box sequence. When tested in the transient expression system, mD could not alter the promoter's activity. The reason for this failure is discussed below.
Fig. 3-17 Functional comparison of the two rainbow trout albumin gene promoters.
Fig. 3-18 Site-directed mutations on the rtALB2 promoter
Constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Relative CAT Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT5</td>
<td>100</td>
</tr>
<tr>
<td>palb2(-487)CAT</td>
<td>207.6</td>
</tr>
<tr>
<td>mA</td>
<td>198.6</td>
</tr>
<tr>
<td>mB</td>
<td>187.3</td>
</tr>
<tr>
<td>mC</td>
<td>312.6</td>
</tr>
<tr>
<td>palb2(-104)CAT</td>
<td>15.2</td>
</tr>
<tr>
<td>mA'</td>
<td>16.8</td>
</tr>
<tr>
<td>mD</td>
<td>11.2</td>
</tr>
</tbody>
</table>

Fig. 3-19 CAT Activities of Mutants
DISCUSSION

Transcriptional regulation of the albumin genes has been studied in various organisms from human to frog (Dimattia and Lazier, 1992; Frain et al., 1990; Maire et al., 1989; Pinkert et al., 1987; Power et al., 1994; Schorpp et al., 1988; Tronche et al., 1990). This is the first study on the regulation of fish albumin gene expression. Rainbow trout is a tetraploid and has two copies of genes to encode a particular protein. The isolation of two rainbow trout albumin cDNAs was described in chapter 2. In this chapter, we describe the isolation of the albumin 2 gene from a genomic DNA library. The length of the albumin 2 gene is estimated to be over 17 kb. About 1.2 kb of the 5' upstream regulatory region has been DNA sequenced and characterized in transient expression system. The 5' flanking region of rainbow trout albumin 2 gene is able to direct liver-specific gene expression since all of the albumin-CAT constructs are only expressed in human hepatoma cells (HepG2) but not in Hela cells. Three functional elements related to the tissue specific activity have been found in this region: (i) a negative element located between bp -1224 and -906, (ii) a positive element essential for efficient albumin transcription located between bp -486 and -104, and (iii) a promoter spanning the first 104 bp upstream of the cap site. This structure is similar to the upstream region of the human albumin gene. although they do not share any DNA sequence identities. The human albumin gene has a negative element (from bp -673 to -486) and an enhancer (from bp -486 to -221) located upstream of its promoter as detected in human hepatoma cells Hep3B (Frain et al., 1990). The rALB2 positive element increased the transcription from the promoter by 15 fold in a cell culture system, whereas the human albumin enhancer increased promoter activity by only 5 fold (Frain et al., 1990). The mouse albumin gene does not have a proximal enhancer or positive regulatory element, but instead has an enhancer located 10 kb upstream of the promoter. In transgenic mice, this enhancer could stimulate transcription by 50 fold (Pinkert et al. 1987), but in a cell culture system it only increased promoter activity by 5 fold (Herbst et al. 1989). This is not
surprising since in a cell culture system, some factors important for transcription may have been lost. One of the purposes for initiating this study was to provide a strong fish liver-specific promoter for fish transgenic study. The positive element in the rtALB2 gene seems to have stronger activation potential than other reported albumin enhancers in the cell culture system. Whether this positive element can drive the expression of heterologous gene promoters in a cell culture systems and in transgenic fish remain to be tested.

The rtALB2 gene 5' flanking region does not show much identity with other albumin genes. Computer assisted searching found many potential binding sites for liver-specific (such as HNF1, C/EBP, DBP and HNF3) and ubiquitous transcription factors (such as NF-Y). In mammalian albumin genes, binding sites for HNF1, C/EBP, DBP, NF-Y and NF-1 are clustered in the proximal promoter region of about 130 bp (Herbomel et al. 1989; Lichtsteiner et al. 1987; Maire et al. 1989). A similar feature is also seen in the rtALB2 promoter where potential binding sites for different transcription factors (HNF1, HNF3, and C/EBP) are concentrated within 70 bp (from bp -100 to -30). The rtALB2 promoter including only the core promoter and a HNF1 site, the palb2(-48), was able to direct gene expression in HepG2 cells but not in Hela cells, suggesting an important role played by HNF1 in the tissue-specificity and activity of the rtALB2 promoter. A fish HNF1 factor has been isolated from Atlantic salmon by Deryckere et al. (1995). The salmon HNF1 shares 92.8% and 86.1% identical amino acids in the homeodomain and the POU domain with the rat HNF1. The dimerization and activation domains are less conserved between the salmon HNF1 and the rat HNF1. The same authors also isolated the Atlantic salmon albumin proximal promoter and found a conserved HNF1 binding site close to the TATA box. Gel mobility shift assay showed that the salmon HNF1 could bind to the salmon and rat albumin HNF1 site. The rainbow trout albumin 2 promoter has a conserved HNF1 binding element as seen in the Atlantic salmon albumin promoter. Our experiment provide direct evidences that the HNF1 binding site is important for the liver-specific promoter activity of rtALB2 gene. Unlike the rat and mouse albumin promoter in
which a HNF1 binding site and a TATA box give almost full transcriptional activity in transient transfection system, the rtALB2 promoter needs other transcription factors binding to the region from nt -48 to -104 for efficient tissue-specific promoter activity since this region could increase the promoter activity 3 fold [compare palb2(-104)CAT with palb2(-48)CAT].

Distinct from all known albumin promoters, the rtALB2 promoter does not have a typical TATA box at the -30 to -25 region. We isolated the rtALB1 promoter which was found to contain a typical TATA box in the area of -31 to -26 and a potential HNF1 binding site, which is similar to the structure of mammalian albumin promoters. The Atlantic salmon albumin gene promoter (Deryckere et al. 1995) has exactly the same DNA sequence as the rtALB1 promoter in this region, suggesting a similar regulation mechanism between the rtALB1 promoter and the Atlantic salmon albumin promoter. The Atlantic salmon and the two rainbow trout albumin promoters have the same TATA box sequence (TATAAA) and a highly conserved potential HNF1 binding site. A striking difference between the rtALB2 promoter and the other two fish albumin promoters is the deletion of 22 bp between the TATA box and the cap site in the rtALB2 promoter, which altered the TATA box position relative to the cap site (only three bases apart). In the -25 area, the rtALB2 promoter has 8 extra nucleotides compared to the rtALB1 promoter. Since we have sequenced two genomic clones isolated from genomic DNA library (clone 16 and clone 4) and several clones from PCR with the same results, it is unlikely that the DNA sequence of the rtALB2 promoter is due to sequencing errors. The difference between the two promoter structures resulted in different activities of the two promoters. The rtALB1 promoter was 3.7 fold stronger than the rtALB2 promoter in HepG2 cells, which coincided with the difference between the transcription rates of the two albumin genes (Chapter 2). The different transcription rates of the two rainbow trout albumin genes are therefore due to their different promoter structures.
In order to understand the mechanism governing the rtALB2 promoter activity, we made two site-directed mutations in the promoter: one was designed to destroy the TATA box located at nt -9 to -4 (mA), the other was designed to destroy the AT-rich element near the -25 region (mB and mC). In our transient expression system, neither of the mutants could abolish the promoter’s activity. This result indicates that neither of these two elements are essential for the promoter’s activity and suggests that the rtALB2 promoter is a TATA-less promoter. The TATA box is usually found in mRNA-encoding gene promoters located 25 bp upstream from the transcription start site. On the other hand, many protein-encoding gene promoters are reported to be TATA-less. According to Arkhipova (1995) about one-half of known Drosophila promoters can be classified as TATA-less. TATA-less promoters often contain an initiator element (Inr) which encompasses the transcription start site to direct sufficient and accurate RNA polymerase II transcription (Weis and Reinberg, 1992; Burley and Roeder, 1996). Many TATA-less initiator-containing promoters transcribe mRNA from multiple sites, such as housekeeping gene promoters which have high GC content and no apparent TATA box. Others initiate transcription from a single start site like the terminal deoxynucleotidyl transferase (TdT) (Dynan, 1986; Weis and Reinberg, 1992). Also in some TATA-less promoters, downstream elements are reported to be able to bind proteins and are critical for the promoter activities (Ince and Scotto, 1995; Burke and Kadonaga, 1996). The initiator is usually a pyrimidine-rich element and the cap site is generally situated within the initiator. The rtALB2 gene is transcribed from one single cap site and it contains a pyrimidine-rich element from +2 to +21 (5' CCCCCGTGTCTCTCTGCTC'TT 3'), whether this element serves as an initiator remained to be investigated.

In TATA-containing promoters, the formation of the transcription preinitiation complex is nucleated from the binding of the TATA-binding protein (TBP) on the TATA box followed by an ordered addition of other general transcription factors (TFIIA, B, E, F, and H) and RNA polymerase II. Studies of many initiator-mediated transcriptions suggest
that TFIID is also required for this type of promoter (For review see Weis and Reinberg, 1992). Four mechanisms have been proposed to explain how TBP is recruited to TATA-less mRNA-encoding promoters: (i) TBP binds directly to low-affinity non-consensus binding sites in the -30 promoter region (Zenzie-Gregory et al. 1993); (ii) TBP is tethered by activators binding on the proximal promoter, such as Sp1 (Pugh and Tjian, 1991); (iii) TBP interacts with sequence-specific initiator-binding proteins (such as YY1 and TFIID) (Weis and Reinberg, 1992) or RNA polymerase II which binds to the initiator (Roy et al. 1991); and (iv) TBP-associated factor(s) (TAF) recruit TBP by interacting with the initiator region (Martinez et al. 1994). Different from TATA-containing promoters where TBP is the rate-limiting factor in their transcription, the initiator-mediated transcription is not limited by TBP, but by a factor which is able to interact with TFIID (Colgan and Manley, 1992). It is possible that the limiting factors for rtALB1 and rtALB2 gene transcription are different, the one for rtALB1 gene might be TBP, while the one for rtALB2 gene might be another protein(s). The concentration of TBP and the limiting factor for rtALB2 promoter or their affinity to interact with DNA or other components in the preinitiation complex might be different, which results in the different transcription rates of the two albumin genes.

Many studies have examined the effect of inserting a TATA box in the -25 region in TATA-less initiator-containing promoters, and different results have been obtained. The mouse terminal deoxynucleotidyl transferase (TdT) promoter is a TATA-less initiator-containing promoter. When a consensus TATA sequence (TATAAAA) replaced the wild-type sequence (GGTCTGC) on the region from nt -31 to -25, a strong stimulatory effect on the basal TdT transcription was observed (Martinez et al. 1994). A different result was obtained from the TATA-less SM50 gene promoter. The SM50 gene encodes a minor matrix protein of the sea urchin embryo spicule. When the -30 to -25 area was changed to a canonical TATA box form, no effect was observed for its function (Makabe et al. 1995). In our experiments, when changing the -25 area of the rtALB2 promoter to a TATA box, the promoter's activity was not affected. An explanation for this failure could be that the HNF1
site is too close to the TATA box in mD mutant. The HNF1 site in other albumin genes are usually located 18 bp upstream from the TATA box (Table 2). In mD mutant, a TATA box was introduced 3 bp downstream from the potential HNF1 binding site. The binding of HNF1 protein may interfere with the binding of TBP on the TATA box, or vise versa, thus the introduced TATA box could not function properly.

We can not exclude other possibilities governing the different transcription rates of the two albumin genes at this stage. For example, the consensus TATA sequence at -9 to -4 and the AT-rich element in the -25 region might serve as a redundant TATA box. As long as one of them remains intact the transcription rate will not be affected. In this case, the TATA box in the rtALB2 promoter may have lower affinity for TBP binding than the rtALB1 promoter, thus its activity is lower. Another possibility comes from the comparison of the rtALB2 promoter with the mouse ribosomal protein rpS16 promoter. The mouse ribosomal protein rpS16 promoter has an AT-rich element in the -30 region as well as an initiator around the cap site. The initiator is required for precise transcription, while the AT-rich element TGAAAAATC from nt -31 to -23 is required for efficient transcription. The AT-rich element bound to a protein which was not TBP as detected by competition gel mobility-shift assay. Replacement of this AT-rich element by a canonical TATA box sequence TATATAAAC did not change the promoter's activity, while changing it to a GC-rich element TGCTCCTC reduced the promoter's activity to 10% of the wild-type promoter (Hariharan and Perry, 1990). The AT-rich element in the rtALB2 promoter is exactly the same as that in the rpS16 promoter. Whether this sequence binds a different factor and function as a TATA-box counterpart in rtALB2 promoter remains unclear. The unique structure of rtALB2 promoter may represent a distinct transcription regulation mechanism.

Our genomic southern provided the possibility that two rtALB2 genes exist in the genome. If this is true, we are not sure at this moment whether both genes are expressed and whether the gene isolated in this work corresponds to the rtALB2 cDNA characterized
in Chapter 2. A complete DNA sequence of the ALB2 gene is required to answer this question.
Chapter Four:

Summary and Future Directions
The objective of this study was to characterize the regulation of rainbow trout albumin gene expression. As a necessary step, we set out to isolate the rainbow trout albumin cDNA from a liver cDNA library. Two distinct rainbow trout albumin cDNAs, rtALB1 and rtALB2, were isolated and sequenced. The rtALB2 cDNA was a novel fish albumin cDNA which had not been reported for other salmonid species and the rtALB1 cDNA was the counterpart of the Atlantic salmon albumin cDNA (Byrnes and Gannon, 1990). The two albumin genes were expressed only in liver, not in other tissues as detected by Northern hybridization. We also measured the transcriptional rates of the two albumin genes and found that the rtALB2 gene was transcribed three times slower than the rtALB1 gene, thus providing the first direct evidence that the two genes for the same protein in tetraploid salmonids can be regulated differently at the transcriptional level (Chapter Two).

We isolated and partially sequenced the first fish albumin gene, the rtALB2 genomic clone. A striking feature of the rtALB2 gene was that there was no typical TATA box in its proximal promoter region. As far as we know, all the reported albumin promoters ranging from mammals to *Xenopus* are TATA-containing promoters (Ray et al., 1991; Deryckere et al., 1995; Table 2 in Chapter Three). The rtALB1 promoter isolated by PCR also contained a TATA box in the -30 to -25 region and a consensus HNF1 binding site upstream of the TATA box, a structure representative of other albumin promoters. Transient expression analysis showed that the rtALB2 promoter was 3-4 times less active than the rtALB1 promoter in hepatoma cells (HepG2), which was consistent with the nuclear run-on results. These data lead to the conclusion that the different transcriptional rates of the two rainbow trout albumin genes were caused by their different promoter structures. We have analyzed the rtALB2 proximal promoter by site-directed mutagenesis and our preliminary results suggested that the rtALB2 promoter was a TATA-less promoter. Some TATA-less promoters have pyrimidine-rich initiators located around the transcription start site to direct the transcription initiation (Weis and Reinberg, 1992). We have noticed that a pyrimidine-rich element from +2 to +21 in the rtALB2 gene. Whether it
serves as an initiator remains to be clarified. Mutations by linker-scanning on the proximal promoter region and the downstream region will be necessary to define the cis element(s) essential for the transcription initiation of the rtALB2 gene. It has been reported that the TATA-binding protein (TBP) is a rate-limiting factor for TATA-containing promoters, but not for TATA-less promoters (Colgan and Manley, 1991). Cotransfection of the rtALB2 promoter-CAT constructs with a TBP expression vector will provide further information to understand the mechanisms regulating the rtALB2 gene transcription.

This work was the first study on the regulation of albumin gene in fish. Up to 1.2kb of the 5' flanking region of the rtALB2 gene was characterized in a transient expression system. Three functional regulatory regions have been identified in the 5' flanking region: a negative region between bp -1224 to -906, a positive region between bp -487 to -104, and a proximal promoter located within the first 100 bp proceeding the transcription start site. Potential binding sites for some transcription factors have been analyzed by computer (Fig. 14 in Chapter Three). DNA-protein interaction analysis by gel mobility shift and footprinting will provide information about cis elements and trans factors involved in the regulation of rtALB2 gene and will eventually lead to the isolation of some important liver-specific or ubiquitous transcription factors from fish, which will benefit the study of fish liver-specific gene transcription.

The mammalian albumin gene promoters and enhancers have been used as a tool to confer tissue-specific expression in transgenic studies. In transgenic mice, the mouse albumin promoter and enhancer directs the human growth hormone gene to express exclusively in liver, but not in spleen, intestine, pancreas, brain, heart, muscle, lung, and testis. Many of the transgenic mice grow larger than controls (Pinkert et al., 1987). Urokinase-type plasminogen activator (uPA) is a plasminogen activator. To study its function in physiologic fibrinolysis, the uPA gene has been linked to the albumin enhancer/promoter and transferred to mice. High expression of the transgene in liver and the elevated plasma uPA levels results in spontaneous intestinal and intra-abdominal
bleeding in newborn transgenic mice (Heckel et al., 1990). The possibility of using an albumin enhancer and promoter in gene therapy of hepatoma has been explored by Kuriyama et al. (1991). In their experiments, the albumin enhancer and promoter were introduced into a retroviral vector and the recombinant retrovirus was injected directly into liver or spleen. It was found that quiescent hepatocytes were not infected by the retrovirus, while the regenerating hepatocytes, which mimic hepatic cells in vivo, were infected. Our laboratory is interested in searching a strong fish liver-specific promoter for fish transgenic studies. In antifreeze gene transfer studies, the expression of antifreeze genes in transgenic fish is very low and functional concentrations of antifreeze protein can not be reached by using its own promoter. It is known that antifreeze proteins have multiple gene copies (about 50-100 copies), therefore, individual antifreeze genes might be insufficient to produce milligram per milliliter quantities of protein necessary for freeze protection (Hew, 1989). A solution for this problem might be to use a strong liver-specific promoter to increase the expression of antifreeze genes in transgenic fish. We have examined the function of mouse albumin enhancer on the ocean pout and winter flounder antifreeze gene promoter and found that the mouse albumin enhancer could not activate these heterologous antifreeze promoters in a cell culture system (data not shown). Thus, identifying a fish liver-specific strong promoter becomes desirable, and the rtALB2 promoter characterized in this study is a potential candidate for this purpose. The rtALB2 5' flanking region contained a positive regulatory region (from -487 to -104) which could increase the activity of the proximal promoter by 15 fold in HepG2 cells, but not in Hela cells. Since the proximal promoter of the rtALB2 gene is a TATA-less promoter and its activity is lower than the TATA-containing rtALB1 promoter, it might be necessary to introduce a TATA box in the proximal region to improve the activity of the rtALB2 promoter before using it in a transgenic study. An alternative approach is to test whether the positive region in the rainbow trout albumin promoter could activate the antifreeze promoter which includes all
cis elements for seasonal or hormonal regulation. We believe that this work opens many interesting fields in both basic and applied research.
REFERENCES


Martinez, E., Chiang, C.M., Ge, H, and Roeder, R.G. (1994). TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. EMBO J. 13, 3115-3126.


