SITE-DIRECTED MUTAGENESIS OF THE BRANCHPOINT SEQUENCE OF INTRON 4 OF THE HUMAN LECITHIN: CHOLESTEROL ACYLTRANSFERASE GENE

By
Min Li

M.D., Nanjing Medical University, PRC, 1983
M.Sc., Nanjing Medical University, PRC, 1989

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY in THE FACULTY OF GRADUATE STUDIES (Department of Pathology and Laboratory Medicine)
We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
September, 2000
©Min Li, 2000
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Pathology and Laboratory Medicine

The University of British Columbia
Vancouver, Canada

Date September 29, 2000
ABSTRACT

The lecithin:cholesterol acyltransferase (LCAT) gene encodes a plasma enzyme that plays a key role in the metabolism of high-density lipoproteins (HDL). Previous mutations associated with LCAT deficiency syndromes have been identified in the coding regions of the LCAT gene. Recently, our laboratory has found an intron mutation in three patients with a form of LCAT deficiency previously described as fish-eye disease (FED). The in vitro expression of the intron mutant has been shown to result in the intron retention. Since the natural mutation occurs in a putative branchpoint consensus sequence, we hypothesized that the point mutation might disrupt the splicing of the LCAT pre-mRNA.

To test the hypothesis, two other novel mutations, i.e., LCAT IVS4-MUT-1 (T→G) and MUT-2 (T→A), were introduced into the same site of the natural mutation (IVS4:T→C22). After stable transfection of the mutated LCAT minigenes into BHK cells, neither LCAT activity nor LCAT protein could be detected in the culture medium of the IVS4-MUT-1 and MUT-2 cell lines, as was previously described for the natural mutation. To determine the effects of the introduced mutations on the splicing of pre-mRNA, total RNA from transfected BHK cells was used for RT-PCR analysis. All BHK cell lines were shown to transcribe the integrated LCAT minigenes. However, the sizes of these LCAT messages indicated that intron 4 was retained in the IVS4-MUT-1 and MUT-2 cell lines. Subsequent sequence analysis of the RT-PCR products demonstrated that the unspliced intronic sequences contained the introduced mutations, suggesting that the observed retention of intron 4 of the LCAT gene is the result of the specific loss of a thymine residue two bases upstream of the branchpoint adenosine.

In attempts to investigate the possible mechanisms responsible for the defective splicing and to study further the functional significance of the branchpoint sequence, a
series of mutations was generated in the whole region of the branchpoint sequence. After these intron mutants were transiently expressed in HEK-293 cells, the efficiency of pre-mRNA splicing was analyzed using RT-PCR as well as the measurement of LCAT activity. The results revealed that (1) the mutation of the branchpoint adenosine to any other nucleotide completely abolished the splicing; (2) the insertion of a normal branch site into the intronic sequence of the natural (IVS4-22c) or the branchpoint (IVS4-20t) mutant restored normal splicing; (3) the natural mutation could be partially suppressed by changing its consensus sequence from CCCC\text{GAC} to CCCC\text{AAC}; and (4) other single-base changes, especially around the branchpoint adenosine residue, significantly decreased the efficiency of splicing and thus enzyme activity. Surprisingly, the nucleotide transversion at the last position of the branchpoint sequence (i.e. IVS4-25a or 25g) resulted in 2.7-fold increase in splicing efficiency.

These results have demonstrated that the branchpoint sequence, although only weakly conserved in mammals, can be of essential importance for accurate and efficient splicing of human nuclear pre-mRNA and have contributed to better understanding of the mechanism of branch-site selection during pre-mRNA splicing. The findings also suggest that a DNA polymorphism involving the branchpoint sequence of an intron might affect the efficiency of RNA splicing and thus have significant clinical implications.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>xvi</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>xvii</td>
</tr>
<tr>
<td><strong>CHAPTER 1</strong> Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Lecithin:cholesterol Acyltransferase</td>
<td>3</td>
</tr>
<tr>
<td>1.1.1 The Key Role of LCAT in the Metabolism of HDL and Reverse</td>
<td>3</td>
</tr>
<tr>
<td>Cholesterol Transport</td>
<td></td>
</tr>
<tr>
<td>1.1.2 Modulation of LCAT Activity in Transgenic Animals</td>
<td>6</td>
</tr>
<tr>
<td>1.1.2.1 Overexpression of Human LCAT Gene</td>
<td>6</td>
</tr>
<tr>
<td>1.1.2.2 LCAT Knock-out Mice</td>
<td>9</td>
</tr>
<tr>
<td>1.1.3 Biochemistry of LCAT</td>
<td>9</td>
</tr>
<tr>
<td>1.1.3.1 Physical Properties of LCAT</td>
<td>9</td>
</tr>
<tr>
<td>1.1.3.2 Glycosylation of LCAT</td>
<td>10</td>
</tr>
<tr>
<td>1.1.3.3 The Role of the C-terminus of LCAT</td>
<td>12</td>
</tr>
<tr>
<td>1.1.3.4 The Phospholipid Substrate Specificity of LCAT</td>
<td>13</td>
</tr>
<tr>
<td>1.1.3.5 The Region of LCAT Involved in Interfacial Recognition to</td>
<td>13</td>
</tr>
<tr>
<td>Lipoproteins</td>
<td></td>
</tr>
<tr>
<td>1.1.3.6 The Catalytic Mechanism of LCAT</td>
<td>15</td>
</tr>
<tr>
<td>1.1.4 The Measurement of LCAT Activity</td>
<td>16</td>
</tr>
</tbody>
</table>
1.2.3.1 Spliceosomal E or Early Complex 42
1.2.3.2 Spliceosomal A Complex 43
1.2.3.3 Spliceosomal B Complex 44
1.2.3.4 Spliceosomal C Complex 47
1.2.4 The Exon/intron Definition Model 48
1.2.4.1 Internal Exon Splicing 48
1.2.4.2 First and Last Exon Splicing 52
1.2.4.3 Intron-definition Model 53
1.2.5 Enhancer Elements 54
1.2.6 Protein Factors Involved in RNA Splicing 55
1.2.6.1 mBBP/SF1 and U2AF 55
1.2.6.2 SR Proteins 56
1.3 Rationale 58
1.4 Specific Aims 59

CHAPTER 2 T → G or T → A Mutation Introduced into the Natural Mutation Site of Intron 4 of Lecithin:cholesterol Acyltransferase (LCAT) Gene: intron retention causing LCAT deficiency 61

2.1 RATIONALE 62
2.2 MATERIALS AND METHODS 63
2.2.1 Materials 63
2.2.2 Molecular Biology Methods 64
2.2.2.1 Growth and Transformation of E. coli 64
2.2.2.2 Purification of DNA 64
2.2.2.2.1 Small Scale Plasmid Preparation 64
CHAPTER 3 Characterization of the Effects of the Mutations in the Branchpoint Sequence of Intron 4 on the Splicing within the Human Lecithin:cholesterol Acyltransferase (LCAT) Gene

3.1 RATIONALE

3.2 MATERIALS AND METHODS

3.2.1 Materials

3.2.2 Molecular Biology Methods

3.2.2.1 Transformation of E.coli

3.2.2.2 Purification of DNA

3.2.2.2.1 Small Scale Plasmid Preparation

3.2.2.2.2 Large Scale Plasmid Preparation

3.2.2.3 Isolation of DNA Fragments

3.2.2.4 Construction of Mutant LCAT Minigenes

3.2.2.5 Isolation of Total Cytoplasmic RNA from the Transfected Cells

3.2.2.6 Reverse Transcription and PCR Amplification (RT-PCR)

3.2.3 Eucaryotic Cell Culture

3.2.3.1 Transient Transfection of HEK-293 Cells

3.2.4 Protein Analysis

3.2.4.1 Measurement of LCAT Activity

3.3 RESULTS

3.3.1 Generation of Mutations in the Branchpoint Sequence of Intron 4 of the LCAT Gene

3.3.2 Mutations of the Branch-site Adenine Completely Abolish Splicing of Intron 4 of the LCAT Gene in vivo

3.3.3 Intron 4 of the LCAT Gene Lacks Cryptic Branchpoint Sequence
3.3.4 The Natural Mutation of Intron 4 Can Be Partially Suppressed by Changing G to A at the Third Position of the Branchpoint Sequence

3.3.5 Single-base Changes in the Branchpoint Sequence of Intron 4 Affect Splicing of LCAT Pre-mRNA in vivo

3.4 DISCUSSION

3.4.1 The Branchpoint Adenosine Is Essential for the Efficient Splicing of LCAT Pre-mRNA

3.4.2 The Insertion of A Normal Branch Site into the Intronic Sequence of the Intron Mutants Completely Restores the Splicing

3.4.3 The Replacement of the Branchpoint Sequence CCCCCGAC to CCCCCAAG Can Partially Rescue the Natural Intron Mutation

3.4.4 Individual Nucleotides in the Branchpoint Sequence of Intron 4 Are Important in Determining the Efficiency of LCAT Pre-mRNA Splicing

3.4.5 The Invariant Branchpoint Sequence (TACTAAC) of Yeast Does Not Increase the Efficiency of LCAT Pre-mRNA Splicing

CHAPTER 4 Remarks and Future Directions

4.1 Implications of the Studies

4.2 Implications of DNA Polymorphism in the Branchpoint Consensus Sequence

4.3 Future Directions

4.3.1 In vivo Studies of Splicing

4.3.2 In vitro Studies of Splicing

4.4 Conclusions

5. REFERENCES
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC1</td>
<td>ATP cassette-binding transporter 1 gene</td>
</tr>
<tr>
<td>apo</td>
<td>apolipoprotein</td>
</tr>
<tr>
<td>ASF/SF2</td>
<td>alternative splicing factor/splicing factor 2</td>
</tr>
<tr>
<td>BHK</td>
<td>baby hamster kidney</td>
</tr>
<tr>
<td>BPS</td>
<td>branchpoint sequence</td>
</tr>
<tr>
<td>CAD</td>
<td>coronary artery disease</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CE</td>
<td>cholesteryl esters</td>
</tr>
<tr>
<td>CER</td>
<td>cholesterol esterification rate</td>
</tr>
<tr>
<td>CERP</td>
<td>cholesterol-efflux regulatory protein</td>
</tr>
<tr>
<td>CETP</td>
<td>cholesteryl ester transfer protein</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>CMV</td>
<td>human cytomegalovirus</td>
</tr>
<tr>
<td>COS</td>
<td>transformed monkey kidney cell</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine triphosphate</td>
</tr>
<tr>
<td>DHFR</td>
<td>dihydrofolate reductase</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's modified Eagle medium</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetra-acetic acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FC</td>
<td>free cholesterol</td>
</tr>
<tr>
<td>FCR</td>
<td>fractional catabolic rate</td>
</tr>
<tr>
<td>FED</td>
<td>fish eye disease</td>
</tr>
<tr>
<td>FH</td>
<td>familial hypercholesterolemia</td>
</tr>
<tr>
<td>FLD</td>
<td>familial LCAT deficiency</td>
</tr>
<tr>
<td>HDL</td>
<td>high density lipoprotein</td>
</tr>
<tr>
<td>HEK-293</td>
<td>human embryonic kidney 293 cell</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-2-hydroxyethylpiperazine-N’-2-ethanesulfonic acid</td>
</tr>
<tr>
<td>HL</td>
<td>hepatic lipase</td>
</tr>
<tr>
<td>LCAT</td>
<td>lecithin:cholesterol acyltransferase</td>
</tr>
<tr>
<td>LDL</td>
<td>low density lipoprotein</td>
</tr>
<tr>
<td>LPL</td>
<td>lipoprotein lipase</td>
</tr>
<tr>
<td>mBBP/SF1</td>
<td>mammalian orthology of yeast branchpoint sequence binding protein</td>
</tr>
<tr>
<td>M-MLV</td>
<td>Moloney murine leukemia virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Neo&lt;sup&gt;R&lt;/sup&gt;</td>
<td>neomycin resistant gene</td>
</tr>
<tr>
<td>NMD</td>
<td>nonsense-mediated mRNA decay</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotides</td>
</tr>
<tr>
<td>PC</td>
<td>phosphatidylcholine</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PL</td>
<td>phospholipid</td>
</tr>
<tr>
<td>PNPB</td>
<td>p-nitrophenylbutyrate</td>
</tr>
<tr>
<td>pre-mRNA</td>
<td>precursor messenger ribonucleic acid</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>Protein G-HRP</td>
<td>Protein G conjugated to horseradish peroxidase</td>
</tr>
<tr>
<td>Py</td>
<td>polypyrrimidine</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motif</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcription-polymerase chain reaction</td>
</tr>
<tr>
<td>SAP</td>
<td>spliceosome-associated protein</td>
</tr>
<tr>
<td>SC35</td>
<td>35-kD spliceosomal component</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecylsulfate</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>snRNP</td>
<td>small nuclear ribonucleoprotein particle</td>
</tr>
<tr>
<td>SR</td>
<td>serine and arginine rich</td>
</tr>
<tr>
<td>SR-BI</td>
<td>scavenger receptor, class B, type I</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>TC</td>
<td>total cholesterol</td>
</tr>
<tr>
<td>TG</td>
<td>total triglyceride</td>
</tr>
<tr>
<td>U2AF</td>
<td>U2 snRNP auxiliary factor</td>
</tr>
<tr>
<td>UC</td>
<td>unesterified cholesterol</td>
</tr>
<tr>
<td>VLDL</td>
<td>very low density lipoprotein</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Lipids, lipoproteins, and apolipoproteins of the FED patients with a point mutation (IVS4:T-22C) in intron 4 and/or exon 4 (T123I) of the LCAT gene</td>
<td>28</td>
</tr>
<tr>
<td>Table 2</td>
<td>LCAT activities, LCAT concentrations, and CER of the FED patients with a point mutation in intron 4 (IVS4:T-22C) in intron 4 and/or exon 4 (T123I) of the LCAT gene</td>
<td>29</td>
</tr>
<tr>
<td>Table 3</td>
<td>LCAT activity and concentration secreted by stably transfected BHK cells</td>
<td>76</td>
</tr>
<tr>
<td>Table 4</td>
<td>Oligonucleotide primers used to create the desired intron mutations</td>
<td>95</td>
</tr>
<tr>
<td>Table 5</td>
<td>The effect of mutations in the branchpoint sequences of the fourth intron of the human LCAT gene on the activity of LCAT from medium of HEK-293 cells</td>
<td>101</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1</td>
<td>Schematic representation of the LCAT reaction</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Schematic diagram of reverse cholesterol transport</td>
<td>4</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Schematic representation of the biochemical characteristics of the 4 different classes of LCAT mutations</td>
<td>24</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Pedigree of the patient's family</td>
<td>26</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Schematic representation of the splicing process of an intron</td>
<td>35</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Schematic representation of the consensus of 5' splice site, 3' splice site, and branchpoint sequence</td>
<td>36</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Schematic representation of assembly of splicing complexes</td>
<td>41</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Schematic representation of exon-definition model</td>
<td>49</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Interactions in the E complex of intron-definition model</td>
<td>54</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Schematic representation of the pNUT expression vector containing LCAT minigenes</td>
<td>74</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Fluorescence-based sequencing results of subsequent LCAT minigene constructs as produced by PCR mutagenesis</td>
<td>75</td>
</tr>
<tr>
<td>Figure 12</td>
<td>RT-PCR on total RNA isolated from BHK cells that were stably transfected with pNUT-LCAT cDNA and pNUT-LCAT minigenes containing intron 4 variants</td>
<td>78</td>
</tr>
<tr>
<td>Figure 13</td>
<td>LCAT Immuno-Western blot of culture medium from BHK cells transfected with LCAT cDNA and LCAT minigene constructs</td>
<td>80</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Proposed possible mechanism for defective splicing</td>
<td>83</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Schematic presentation of the pcDNA3.1-LCAT minigene construct</td>
<td>100</td>
</tr>
<tr>
<td>Figure 16</td>
<td>The effect of the branchpoint mutations on the LCAT pre-mRNA splicing as analyzed by RT-PCR</td>
<td>103</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Insertion of a natural branch site into the sequences of mutated intron 4 of the human LCAT genes</td>
<td>105</td>
</tr>
<tr>
<td>Figure 18</td>
<td>The effect of the introduction of a normal branch site into</td>
<td></td>
</tr>
</tbody>
</table>
the intron sequence on the splicing of the LCAT pre-mRNA containing the natural (IVS4-MUT) and the branchpoint (IVS4-20t) mutations 106

Figure 19  The effect of the change of the branchpoint sequence from CCCCGAC to CCCCAAC on the splicing of LCAT pre-mRNA 109

Figure 20  The effects of the mutations in the branchpoint sequence on the efficiency of LCAT pre-mRNA splicing 112
ACKNOWLEDGMENTS

I am deeply grateful to my supervisor, Dr. P. Haydn Pritchard, for his support, advice, enthusiasm and encouragement, and patience during the course of my work in his laboratory.

I also wish to thank my supervisory committee members, Dr. John Hill, Dr. Ross MacGillivary, and Dr. Susan Porter for their guidance and advice throughout this project.

Finally, I would like to thank Dr. Jiri Frohlich and all members of Atherosclerosis Specialty Laboratory at St. Paul's hospital: Lida Adler, Aleta Allen, Amir Ayyobi, Michelle Francis, Tom Green, Scott Lear, Sandra McGladdery, Trina Mcllhargey, Dr. Mohammed Moghadasian, Yingying Yang, and Daniella Zamfir, who have together created an enjoyable and productive environment in which to work. A special thanks also goes to Dr. Jan Albert Kuivenhoven, whose previous work made the present project possible.
DEDICATION

This work is dedicated to my father, Dingheng Li, my mother, Huizhen Li, my wife, Jiakun Chen, and my son, George Shuwei Li.
CHAPTER 1

Introduction
Atherosclerotic cardiovascular disease is the leading cause of death in developed western countries. For coronary artery disease (CAD), hypercholesterolemia and dyslipoproteinemia are major risk factors. It has been well established that the plasma level of high-density lipoprotein (HDL) cholesterol is inversely associated with the risk of CAD (Gordon et al., 1989). One of the major determinants in maintaining the normal level of HDL cholesterol is a plasma enzyme called lecithin:cholesterol acyltransferase (LCAT). As a strategy to study factors that regulate the concentration of HDL cholesterol in plasma, our laboratory has studied the molecular pathology of LCAT over the past twenty years. During studies of patients with LCAT deficiency syndromes, our laboratory has found an interesting case in which an intron mutation was identified to affect the normal splicing of the LCAT gene leading to fish-eye disease (FED). As this intron mutation is located in the putative branchpoint sequence and it is one of the first branchpoint mutations that has been described to cause a human disease, it is very important to explore the potential mechanism(s) underlying this unique genetic defect, and that the elucidation of the defective splicing mechanism(s) within the LCAT gene will add to our better understanding of the mechanism of branch-site selection during pre-mRNA splicing.

I will begin by summarizing the current knowledge of the biochemistry and molecular pathophysiology of LCAT and then describe recent advances made in the understanding of splicing of nuclear messenger RNA precursors before outlining the specific aims of this study.
1.1 Lecithin:cholesterol Acyttransferase

Lecithin cholesterol acyltransferase (LCAT) (EC 2.3.1.43) is the enzyme responsible for the synthesis of the majority of cholesteryl esters (CE) in plasma via the transfer of sn-2 fatty acid from phosphatidylcholine (PC) to the 3-β-hydroxyl group of cholesterol, resulting in lysolecithin and cholesterol esters. Therefore, catalysis of the enzyme on its substrate combines two distinct activities: (1) Phospholipase A\(_2\); and (2) Acyttransferase (Fig.1).

Figure 1. Schematic representation of the LCAT reaction

1.1.1 The Key Role of LCAT in the Metabolism of HDL and Reverse Cholesterol Transport

Many epidemiological studies have demonstrated a strong inverse relationship between plasma levels of HDL cholesterol and the risk of CAD (Gordon et al., 1989). The mechanism by which HDL particles inhibit the development of atherosclerosis is still not completely elucidated. However, it is generally accepted that the protective role of
HDL against atherosclerosis involves primarily its ability to remove excess cholesterol from peripheral tissues via plasma to the liver for catabolism, a process known as reverse cholesterol transport (Fig.2).

**Figure 2. Schematic diagram of reverse cholesterol transport.** UC: unesterified cholesterol; CE: cholesteryl esters; LCAT: lecithin:cholesterol acytransferase; CETP: cholesteryl ester transfer protein; VLDL: very low density lipoproteins; LDL: low density lipoproteins; HDL: high density lipoproteins; A-I: apolipoprotein A-I.

LCAT is a key plasma enzyme in cholesterol and HDL metabolism. HDL consists of approximately 50% lipid and 50% protein. Apolipoprotein A-I (apo A-I) and apo A-II are the major proteins found in HDL. In plasma, LCAT is bound to HDL, and nascent HDL is the preferred substrate of LCAT. The principal activator of the enzyme is the major structural component of HDL, apo A-I (Fielding et al., 1972), although a variety of other apolipoproteins, including apo A-IV (Steinmetz and Utermann, 1985), apo C-I (Streyer and Kostner, 1988), apo D (Streyer and Kostner, 1988), and apo E (Zorich et al., 1985) have also been shown to activate the enzyme. The biosynthesis of nascent HDL is not clearly understood, but nascent HDL is believed to be secreted by the liver and the intestine and also to arise from the lipolysis of large triglyceride (TG) -rich lipoproteins by lipoprotein
lipase (LPL). Most recently, it has been demonstrated that an ATP-binding-cassette transporter (ABC1, also called cholesterol-efflux regulatory protein (CERP)) which mediates efflux of cholesterol from cells plays a crucial role in generation of HDL (Brooks-Wilson et al., 1999; Bodzioch et al., 1999; Rust et al., 1999). This newly discovered ABC1/CERP is necessary for the transfer of free cholesterol (FC) and phospholipid (PL) out of cells to lipid-poor apo A-I forming nascent HDL. These HDLs (belong to the pre-β fraction), are small discoidal or spherical particles, which are protein-rich, but lipid-poor. After entering the plasma, these newly secreted/generated HDL particles acquire a core of cholesteryl esters through the catalytic action of the LCAT enzyme and are thereby transformed into mature spherical HDL (Fielding and Fielding, 1995). Besides its activity on HDL (α-activity), LCAT is also active on low-density lipoproteins (LDL, β-activity), for which no cofactor is required.

Through the above described action, LCAT maintains a cholesterol concentration gradient between peripheral cells and the plasma (Fielding, 1987). This important enzyme, therefore, plays a central role in the initial steps of reverse cholesterol transport. Once cholesterol is esterified in HDL by LCAT it can be transferred to low-density lipoproteins (LDL) and very low-density lipoproteins (VLDL) in exchange for triglyceride via the action of cholesterol ester transfer protein (CETP) and these apo B-containing lipoproteins can then be cleared by hepatic apo B/E receptors. Recently, the identification of a hepatic HDL receptor (Rigotti et al., 1997), scavenger receptor, class B, type I (SR-BI), suggests an another mechanism of delivery of cholesteryl esters to liver that differs from the receptor-mediated uptake of LDL. In addition, another key enzyme, hepatic lipase (HL) also participates in HDL metabolism together with LCAT in modulating HDL particle heterogeneity and function (Dugi et al., 1997). HL acts on TG-enriched HDL (HDL₂) which results from the process of lipid transfer, and the hydrolysis
of the HDL₂-triglycerides and phospholipids by this enzyme gives rise to the synthesis of small HDL₃ (Bensadoun and Berryman, 1996). These particles will again serve as acceptors of cholesterol whereby the HDL remodeling cycle is complete. The central role of LCAT in these physiological processes is implicated by the findings of plasma and tissue accumulation of free cholesterol and the presence of nascent HDL particles in plasma of patients with LCAT deficiency syndromes (Assmann et al., 1991).

Since it is well established that plasma levels of HDL cholesterol are inversely associated with the risk of CAD, LCAT is considered to be one of the most important enzymes in determining the risk of atherosclerosis. This antiatherogenic role has been demonstrated in transgenic animals overexpressing human LCAT. Studies in transgenic rabbits for human LCAT showed that overexpression of the enzyme increases plasma HDL cholesterol and apo A-I concentrations (Brousseau et al., 1997; 1998) and reduces diet-induced atherosclerosis (Hoeg et al., 1996a). Therefore, LCAT along with apo A-I provide new potential targets for therapy of patients with atherosclerosis associated with a low HDL cholesterol level.

1.1.2 Modulation of LCAT Activity in Transgenic Animals

The generation of either transgenic animals overexpressing the human LCAT gene or LCAT knock-out mice provides a new insight into the role of LCAT in lipoprotein metabolism, reverse cholesterol transport, and atherosclerosis.

1.1.2.1 Overexpression of Human LCAT Gene

In recent years, human LCAT gene has been expressed in both mice (Francone et al., 1995; Vaisman et al., 1995; Mehlum et al., 1995) and rabbits (Brousseau et al., 1996; Hoeg et al., 1996b). To assess the effects of an increased LCAT plasma
concentration on lipoprotein metabolism, Mehlum et al. generated a strain of transgenic mouse expressing a 40-fold increase in LCAT activity compared with that of control a mouse. Overexpression of human LCAT in the transgenic animal resulted in elevation of HDL-cholesterol by about 20% accompanied by an increase in both apo A-I and apo A-II plasma levels. By contrast, plasma concentrations of triglyceride as well as LDL and VLDL cholesterol were markedly reduced (Mehlum et al., 1995). This anti-atherogenic lipoprotein profile was also observed in another study in which three different transgenic mouse lines overexpressing human LCAT at plasma levels 11-, 14-, and 109-fold higher than non-transgenic mice were established (Vaisman et al., 1995). A high correlation between plasma LCAT activity and total as well as HDL cholesterol levels was found, which demonstrated that in mice LCAT modulates plasma HDL concentrations. Recent studies of transgenic mice expressing human LCAT have indicated that LCAT facilitates the reverse cholesterol transport pathway through modulating the rate by which cholesterol is effluxed from cell membranes onto HDL, esterified and transported to the liver (Francone et al., 1997). Although LCAT transgenic mice with the apparent anti-atherogenic lipoprotein profile are not protected from diet-induced atherosclerosis development (Berard et al., 1997; Mehlum et al., 1997), cholesterol-fed mice which are transgenic for both human LCAT and simian CETP do not develop atherosclerosis (Fruchart and Duriez, 1998; Foger et al., 1999), indicating that over production of LCAT reduces atherosclerosis when CETP is expressed in the animals.

Transgenic rabbits overexpressing the human LCAT gene have been developed as a much better animal model for assessing the effects of the transgene on lipoprotein metabolism and atherosclerosis susceptibility (Brousseau and Hoeg, 1999). Human LCAT overexpression in rabbits resulted in 1.5 and 2.5-fold increases in total cholesterol and HDL-cholesterol, respectively, with a 3.1-fold increase in LCAT activity (Hoeg et al.,
These changes were attributed to an increased concentration of large cholesteryl ester- and apo E-enriched HDL-1 particles, with a virtual absence of apo B-containing lipoproteins. Metabolic studies demonstrated that the hyperalphalipoproteinemia was due to the delayed fractional catabolic rate (FCR) of apo A-I in a gene dose-dependent manner (Brousseau et al., 1996), whereas reduced concentrations of atherogenic lipoproteins resulted from the increased FCR of LDL apo B-100 in LCAT transgenic rabbits (Brousseau et al., 1997). Compared with nontransgenic controls, these LCAT transgenic rabbits in the presence of CETP were protected from diet-induced atherosclerosis with significant reductions determined by both quantitative planimetry (-86%) and quantitative immunohistochemistry (-93%) (Hoeg et al., 1996a). To further assess the role of LCAT in the prevention of atherosclerosis, the human LCAT transgene was introduced into LDL receptor-deficient rabbits (i.e., WHHL rabbits). Overexpression of human LCAT in WHHL rabbits increased HDL-cholesterol and apo A-I concentrations by both decreasing apo A-I catabolism and increasing apo A-I synthesis, thus correcting the metabolic defects responsible for the hypoalphalipoproteinemia observed in WHHL rabbits (Brousseau et al., 1998). Since overexpression of human LCAT enhanced LDL catabolism without influencing production rate (Brousseau et al., 1997), the reduced LDL concentrations observed in both chow- and cholesterol-fed LCAT transgenic rabbits might be due to up-regulation of the LDL receptor pathway. This hypothesis has been proven to be the case. Human LCAT expression in heterozygous WHHL rabbits resulted in approximately 70% reduction in LDL apo B pool size due to significant increases in LDL apo B-100 FCR. However, LDL apo B pool size and LDL apo B-100 FCR as well as aortic atherosclerotic lesions (Hoeg et al., 1998; Brousseau et al., 2000) were nearly identical when LCAT transgenic and nontransgenic homozygous WHHL rabbits were
compared. These results established that LCAT modulates LDL metabolism via the LDL receptor pathway, ultimately influencing atherosclerosis susceptibility.

1.1.2.2 LCAT Knock-out Mice

LCAT knock-out mice have been established using targeted gene disruption techniques in two separate laboratories (Ng et al., 1997; Sakai et al., 1997). LCAT deficiency resulted in significant reductions in the plasma concentrations of total cholesterol, HDL cholesterol, and apoA-I in both homozygotes and heterozygotes (Sakai et al., 1997). The HDL particles are heterogeneous and include numerous discoidal particles, resembling those observed in LCAT deficient humans (Ng et al., 1997). Despite the similarities in most biochemical parameters with human familial LCAT deficiency (FLD), mouse models do not develop the severe clinical complications present in FLD (i.e., corneal opacity, and renal insufficiency) (Sakai et al., 1997). The other remarkable characteristics include elevated triglyceride levels and up-regulation of the adrenal SR-BI mRNA due to severe depletion of adrenal lipid stores (Ng et al., 1997; Sakai et al., 1997). These observations further suggest that LCAT may have a direct role in the catabolism of triglyceride-rich, apo B-containing lipoproteins and that the reduction in plasma HDL-cholesterol from deficiency of LCAT may have an impact on the cholesterol flux to the adrenal gland via a selective uptake pathway (Ng et al., 1997).

1.1.3 Biochemistry of LCAT

1.1.3.1 Physical Properties of LCAT

LCAT is a glycoprotein with an apparent molecular weight of about 65,000 to 69,000 daltons (Marcel et al., 1982). The human LCAT mRNA contains 1550
nucleotides that encode 416 amino acids with a hydrophobic leader sequence of 24 residues (McLean et al., 1986a; Yang et al., 1987). The unglycosylated mature protein has a calculated molecular weight of 47,090 daltons, and therefore the carbohydrate content is estimated to represent approximately 25% of total LCAT mass. The enzyme is primarily synthesized in the liver and secreted into the plasma where it is primarily associated with HDL particles. The primary structure of human LCAT has been determined by the cloning and sequencing of the LCAT cDNA (McLean et al., 1986b). The chemical analysis of LCAT protein has determined six cysteine residues, four of which form disulfide bonds; Cys-50 with Cys-74 and Cys-313 with Cys-356. The remaining two, Cys-31 and Cys-184 retain free sulphydryl groups (Yang et al., 1987). Neither of the two free Cys residues has been shown to play a role in the enzymatic activity (Francone and Fielding, 1991a). Analysis of the LCAT protein sequence also reveals four potential N-linked glycosylation sites (Asn-X-Ser/Thr) at residues 20, 84, 272 and 384 (McLean et al., 1986b).

1.1.3.2 Glycosylation of LCAT

Site-directed mutagenesis has been performed to confirm that all four N-linked glycosylation sites in the LCAT sequence are occupied. Previous studies (Francone et al., 1993) showed that elimination of the glycosylation sites at positions 20, 84 or 384 had only minor effects on either the phospholipase or acyltransferase activities of LCAT, whereas the oligosaccharide moiety at position 272 had an essential role in acyltransferase activity. Data from another experiments (Qu et al., 1993) suggested that carbohydrates at residues 20 and 384 have little or no effect on the specific activity of the mutant LCAT. Substitution of Asn to Thr at position 84 reduced the secretion and specific activity to 58% and 26% of the wild type, respectively, while no secreted LCAT
could be detected with the Asn<sub>272</sub>→Thr mutant, indicating that this residue is essential for intracellular processing. Studies from our lab (O et al., 1993a) have shown that loss of carbohydrates at position 20, 84 or 272 resulted in decreased specific activity by 18%, 82%, and 62% respectively. By contrast, the mutant without glycosylation at position 384 displayed a 2-fold increase in enzyme activity, suggesting that the presence of carbohydrate at each site exerts diverse effects on the enzyme activity. In addition, elimination of all four N-linked glycosylation sites by generating a quadruple mutant significantly decreased not only the specific activity (to 5%) but also the secretion (to 10%) of the enzyme as compared to wild-type control. The discrepancy among the results obtained from several laboratories may result from the different cell lines employed (i.e., COS-6, COS-1 or CHO), and the different amino acids mutated (i.e., Asn→Thr or Asn→Gln). Furthermore, it can not be ruled out from these experiments that the effects of the mutations might be due to the changes in the amino acid itself at these sites, which may also affect the structure and function of the protein.

Since LCAT can esterify cholesterol in HDL as well as LDL, further studies were designed to assess the influence of N-linked glycosylation of LCAT on lipoprotein substrate specificity (O et al., 1995). Elimination of the carbohydrate moiety at position 20 was associated with a lower activity on HDL, but no effect on LDL compared with wild type. Substitution of Asn-84→Gln or Asn-272→Gln displayed a decreased ability to esterify cholesterol in either HDL or LDL, whereas the loss of a carbohydrate chain at position 384 led to an increase in enzyme activity for both HDL (1.5-fold) and LDL (2.5-fold) substrates. These results suggest that individual carbohydrate chains confer specific structural and functional properties to LCAT. Recently, two unanticipated O-linked glycosylation sites were found at positions Thr407 and Ser409 (Schindler et al.,...
1995); however, the function of these carbohydrate structures has not been identified yet.

1.1.3.3 The Role of the C-terminus of LCAT

Human LCAT has an unusual hydrophilic, proline-rich region at the C-terminus (McLean et al., 1986b), containing 8 proline residues out of the final 15 amino acids. The C-terminal region of LCAT encoded by exon 6 corresponds to about 50% of the mature protein (amino acids 226-416). The function of the C-terminal region of LCAT has been highlighted by the presence of several natural mutations (Hill et al., 1993a; Klein et al., 1992; Assmann et al., 1991) as well as by the presence of the last N-glycosylation site at 384 and by studies of deletion mutagenesis in this region (Lee et al., 1997; Francone et al., 1996). It has been shown that the proline-deletion mutants were similar to wild type LCAT in terms of phospholipase and acyltransferase activities with various interfacial substrates, but resulted in approximately an 8-fold increase in the specific activity of LCAT towards the water-soluble substrate, p-nitrophenylbutyrate (PNPB) (Lee et al., 1997). This suggests that the C-terminal proline-rich region may interfere with access of water-soluble substrates to the active site of LCAT and that this region may therefore provide a protective covering of the active site while the enzyme is in solution. Further deletions in the C-terminal region beyond Pro402 impaired the secretion of LCAT from cells, implying that this region may play a critical role in either the secretion or folding of LCAT in COS-1 cells. Similarly, Francone et al. also demonstrated that the proline-rich C-terminus of LCAT is not required for effective enzyme activity (Francone et al., 1996). However, it is interesting to note that all the deletion mutants used in the experiment, spanning the C-terminus from residue 416 to 315 of LCAT, were secreted from CHO cells at similar rates. This may reflect the
different secretory mechanisms of handling the produced LCAT proteins in the two cell types.

1.1.3.4 The Phospholipid Substrate Specificity of LCAT

Human LCAT has a higher reactivity with PC species containing oleic or linoleic acid in the sn-2 position compared to arachidonic acid, whereas rat or mouse LCAT displays the opposite preference profile. A discrete region of the primary sequence of LCAT is proposed to be responsible for the fatty acyl specificity of the enzyme. Subbaiah et al. have shown that the central region (amino acids 130-306) of LCAT is essential in determining the fatty acid specificity in the acyltransferase reaction through the construction of a hybrid human-mouse LCAT gene (Subbaiah et al., 1994). By using a comparative species approach that focused on the amino acid differences between human and rat LCAT, Wang et al. recently identified a single amino acid residue in the human LCAT, the glutamic acid at position 149, that is responsible for conferring fatty acyl specificity (Wang et al., 1997). When E149 of human LCAT was mutated to an alanine, which is present in the rat sequence at position 149, the fatty acyl specificity of human enzyme was converted to that of rat LCAT, with an increase in activity toward the PC substrate containing arachidonic acid. Whether or not this substrate specificity indicates the importance of plasma CE composition in determining the atherogenic risk in terms of the susceptibility to atherosclerosis among several species (Lui et al., 1995) remains to be determined.

1.1.3.5 The Region of LCAT Involved in Interfacial Recognition to Lipoproteins
LCAT is a surface-active enzyme (Weinberg et al., 1995), which acts on lipid substrates on the surface of both HDL and LDL particles. The interfacial recognition domain of LCAT plays an important role in regulating the interaction of the enzyme with its lipoprotein substrates. Based on observations with other interfacial lipases (i.e., pancreatic lipase, hepatic lipase, and lipoprotein lipase), Adimoolam et al. have identified a surface region of LCAT that is involved in the binding of LCAT to lipoproteins (Adimoolam et al., 1997). This domain of LCAT consists of 25 amino acids linked by a disulfide bond (C50-C74). An in vitro expression study with the deletion mutant (LCAT delta 53-71) has shown that the deleted region is located on the surface of the enzyme and is not required for folding because the expression and secretion of the LCAT mutant are at levels similar to wild type LCAT. The loss of activity on both HDL and LDL suggests that the deletion abolishes the binding of LCAT to lipoprotein surfaces, whereas the retention of partial LCAT activity on a water soluble substrate, p-nitrophenyl butyrate (PNPB) indicates that the catalytic domain of the LCAT mutant is preserved. Therefore, it is concluded that this region (C50-C74) forms part of the interfacial binding domain of LCAT. Recently, the purified lid-deletion mutant of LCAT (Δ53-71) has been demonstrated to exhibit no binding to LDL by surface plasmon resonance method (Jin et al., 1999), indicating that the region of LCAT between residues 53 and 71 is essential for interfacial binding. Another segment of human LCAT (154-171), the most stable amphipathic helix in the LCAT sequence, has also been identified as a candidate domain that is involved in the association of the enzyme with its lipid substrate (Peelman et al., 1997). Interestingly, this domain is in the close neighborhood of residues 140-150, which determine the substrate specificity of LCAT (Wang et al., 1997).
1.1.3.6 The Catalytic Mechanism of LCAT

The enzymatic reaction of LCAT on lipoprotein particles consists of a series of steps. The first step is the binding of the enzyme to the lipoprotein surface. This step is followed by activation with apo A-I and binding of PC to the active site of the enzyme. Then, formation of the acyl-enzyme intermediate and lyso-PC are accompanied by binding of cholesterol to the active site and esterification of cholesterol in the active site. Finally cholesteryl esters formed are diffused into the core of the lipoproteins followed by release of the enzyme.

It is generally accepted that the catalytic mechanism of LCAT is similar to that of serine proteases and lipases involving a Ser, a His, and an acidic amino acid residue. Ser181 in LCAT has been previously identified as a catalytic residue by site-directed mutagenesis (Francone & Fielding, 1991b). Recently, a molecular model was built for the human LCAT based upon the structural homology between this enzyme and lipases (Peelman et al., 1998). In this model, it is proposed that LCAT, like lipases, belongs to the α/β hydrolase fold family, and that the central domain of LCAT consists of a mixed seven-stranded β-pleated sheet with four α-helices and loops linking the beta-strands. Based on this three-dimensional model, two other catalytic residues, Asp345 and His377 have been identified by mutational analysis (Peelman et al., 1998; Adimoolam et al., 1998a). Thus, the catalytic triad of LCAT consists of Asp345 and His377, as well as Ser181. This model explains the structural defects linked to the point mutations identified in LCAT, which cause familial LCAT deficiency syndromes (Peelman et al., 1999). Most mutations causing FLD are either clustered in the vicinity of the catalytic triad or affect conserved structural elements in LCAT, while most mutations causing FED are localized on the outer hydrophilic surface of the amphipathic helical segments. These mutations affect only minimally the overall structure of the enzyme, but are likely
to impair the interaction of the enzyme with its co-factor and/or substrate (Peelman et al., 1999). However, this proposed architecture for LCAT has to wait to be confirmed by X-ray crystallography of the LCAT protein, which is being actively carried out in our lab.

1.1.4 Measurement of LCAT Activity

Two methods are routinely used to determine the activity of LCAT in plasma, which were initially established for the differential diagnosis of HDL deficiency syndromes. These two assays are distinct from each other with respect to the utilization of different substrates, i.e., exogenous and endogenous substrates.

1.1.4.1 LCAT Activity

In this assay, an exogenous substrate (reconstituted HDL) is employed to measure the activity of LCAT in plasma or tissue culture medium. The activities obtained are generally proportional to the amount of LCAT protein present in the reaction. The substrate used in this assay is an excess of artificially prepared HDL analogue consisting of known amount of $[^{3}\text{H}]$ cholesterol, phosphatidylcholine, and apo A-I. The ratio of apo A-I to cholesterol and phospholipid is optimized for maximal activation of LCAT. The activity is determined by measuring the rate of conversion of $[^{3}\text{H}]$ cholesterol to cholesteryl esters, and is expressed as nmoles of cholesterol esterified per hour per milliliter of sample.

1.1.4.2 Cholesterol Esterification Rate (CER)

LCAT activity can also be measured using the cholesterol from plasma lipoprotein pools. The cholesterol esterification rate (CER) measures the endogenous rate of cholesteryl esters synthesized in whole plasma in the absence of any exogenous
substrate. In this assay, [3H] cholesterol is first equilibrated with the pool of endogenous unesterified cholesterol in plasma at 4°C followed by subsequent incubation of the plasma at 37°C. LCAT activity is then calculated as the rate of synthesis of [3H] cholesteryl ester. Although the units of CER are identical to those stated for LCAT activity (also nmoles of cholesterol esterified per hour per milliliter) the two assays are not equivalent. CER is influenced by not only the amount of LCAT present in plasma but also the endogenous plasma lipoprotein substrates and cofactors and other plasma constituents as well. Therefore, CER is usually applied for the differential diagnosis of FLD and FED (see Section 1.1.5.2 below).

1.1.5 Molecular Pathophysiology of the LCAT Gene

1.1.5.1 LCAT Gene Structure

The LCAT gene is localized in the q21-22 region of chromosome 16 (Teisberg et al., 1975). It consists of 6 exons interrupted by 5 introns encompassing a total of 4.2 kilobases (McLean et al., 1986a). Southern blot hybridization data suggest that there is only one LCAT gene in humans. An unusual feature of the LCAT message is that the poly(A) signal AATAAA overlaps the carboxyl-terminal glutamic acid and stop codons, and the 3' untranslated region contains only 23 bases (McLean et al., 1986b). So far, little is known about the regulation of the expression of the LCAT gene with respect to the specific DNA sequences and trans-acting factors. A minimal promoter sequence extending 71 nucleotides (nt) before the transcriptional initiation site has been found to determine the transcriptional activity of the LCAT promoter (Meroni et al., 1991). In a recent study, Hoppe et al. identified the presence of two non-consensus GC-rich regions in the proximal LCAT promoter that stimulate transcription in both HepG2 (human hepatocellular carcinoma) and HeLa (human cervix epitheloid carcinoma) cells.
(Hoppe et al., 1998). Site-directed mutagenesis revealed that these two regions designated sites A (-29 to -47) and B (-49 to -65) are equally important in promoter activity. Both Sp1 and Sp3 as the transcription factors can interact with these sites. Sp1 but not Sp3 activates the human LCAT promoter and when Sp1 is co-transfected along with Sp3, Sp3 functions as a dose-dependent repressor of Sp1-mediated activation, suggesting that the levels of Sp3 or the nuclear Sp1/Sp3 ratio may play an important role in determining the transcriptional activity of the LCAT promoter in vivo.

In humans (McLean et al., 1986a), mice (Warden et al., 1989), and monkeys (Smith et al., 1990), LCAT gene expression has been primarily detected in the liver and, to a much lower extent, in the brain and testis. The latter indicates that this enzyme is also essential for the removal of excess cholesterol from these tissues isolated from the bloodstream by the blood-brain or -testis barrier. The regulatory elements required for tissue-specific expression remain to be identified. However, the generation of human LCAT transgenic mice provides a model to study the regulation of this gene in addition to its effect on HDL composition and structure. A human LCAT genomic fragment containing 561 bp of 5' untranslated region and the 3' untranslated sequence extended 550 bp after the stop codon is used to create the transgenic mice (Francone et al., 1995). The expression of this LCAT genomic construct exclusively in the liver suggests that the sequence-specific elements required for liver expression are present in the 561 bp of the 5' untranslated region.

All exon-intron splice sites of the LCAT gene confirm to the GT-AG rule and to reasonable variants of the surrounding consensus sequences (McLean et al., 1986a). The largest intron of the gene, intron 5 contains three Alu type repetitive elements in the same 3' to 5' orientation. Interestingly, an alternative LCAT mRNA with a splice-mediated insertion of a 95 bp Alu cassette at the junction of exon 5 and exon 6 has
been identified originally in a patient with complete LCAT deficiency, later in humans and the great apes (Miller et al., 1997). In humans, the alternative transcript represents 5-20% of the complete LCAT mRNA in cultured fibroblasts and liver. The presence of this alternative splicing variant among primates suggests a potential role of Alu elements in the evolutionary diversity of proteins.

1.1.5.2 Mutations in the LCAT Gene Cause Familial LCAT Deficiency Syndromes

The physiological consequences of a decrease or total loss of cholesterol esterification by LCAT in plasma are quite severe, as illustrated by pathological changes observed in patients with LCAT deficiency syndromes (Assmann et al., 1991). Extensive genetic studies over the last 10 years have clearly demonstrated that mutations in the human LCAT gene underlie either FLD or FED, which are both inherited in an apparently autosomal recessive manner (Kuivenhoven et al., 1997). However, the clinical and biochemical phenotypes of carriers of FLD or FED are different.

1.1.5.2.1 Familial LCAT Deficiency (FLD)

The first naturally occurring mutation of LCAT associated with FLD was reported by Norum and Gjone (Norum et al., 1967) in 1967. Since then, a number of unrelated families with FLD have been diagnosed. The clinical manifestations of FLD patients include extensive corneal opacification, anemia, proteinuria, and renal insufficiency (Assmann et al., 1991). These clinical symptoms are due to highly elevated unesterified cholesterol and phospholipid levels in the affected cells and tissues. The most consistent biochemical characteristic of FLD is high plasma concentrations of unesterified cholesterol (UC) and phosphatidylcholine (PC) and residual levels of low
concentrations of cholesteryl esters (CE) and lysolecithin. In homozygous patients, both LCAT activity and protein may be absent, or, in certain patients, LCAT is immunologically detectable but functionally inactive. In addition, all classes of lipoproteins exhibit abnormal profiles, have altered morphologies, and contain high amounts of UC and PC. The concentrations of HDL cholesterol and apo A-I are markedly reduced and are accompanied by a heterogeneous distribution of HDL size and shape (Glomset et al., 1989). Heterozygous individuals do not present any clinical symptoms but the activity of LCAT in plasma is reduced to half the normal level. The concentrations of HDL cholesterol and apo A-I in these heterozygous patients are also low compared with unaffected family members. However, FLD is generally not accompanied by atherosclerosis, although a few cases have been reported (Glomset et al., 1989).

It was not until 1992 that Skretting and colleagues described the specific molecular defect underlying the first case of familial LCAT deficiency (Skretting et al., 1992). Sequence analysis of the LCAT genes in the probands of the FLD revealed that they were homozygous for a single nucleotide substitution in exon 6 of the LCAT gene that resulted in the exchange of a methionine for a lysine residue at position 252 of the mature protein. In vitro expression of this missense mutation was shown to lead to an inactive but normally secreted enzyme (Hill, 1994).

1.1.5.2.2 Fish-eye Disease (FED)

A second genetic disorder of LCAT metabolism, known as FED, was originally described in two Swedish families (Carlson et al., 1979; 1982). In FED patients, the only clinical signs are the remarkable corneal opacity, which is responsible for the unusual name of the disease, since the appearance of the eyes of affected individuals is similar
to those of a boiled fish (Carlson et al., 1985). No other clinical symptoms are evident in FED. A distinguishing feature of this disorder is the presence of an apparently normal ability of LCAT to esterify cholesterol in plasma, resulting in a normal percentage of plasma CE. Other characteristics of FED include HDL deficiency, elevated triglyceride levels, and partial reduction in LCAT concentration (Carlson et al., 1983). In vitro studies have shown that LCAT from FED patients does not esterify cholesterol in HDL (α-activity) but retains its ability to esterify cholesterol in LDL (β-activity) (O et al., 1993b). As the cholesteryl ester content of HDL was markedly reduced whereas that of VLDL and LDL was normal, it was supposed that LCAT exhibits two activities in normal plasma: α-LCAT activity that is specific for HDL (with α-mobility upon gel electrophoresis), and β-LCAT activity specific for VLDL and LDL (with pre-β- and β mobility, respectively). Therefore, FED was classified as α-LCAT deficiency whereas FLD resulted from loss of both α- and β-LCAT activities. Cloning of the LCAT gene demonstrated that α- and β- activity represent two functions of the same protein (McLean et al., 1986b). In addition to the markedly decreased plasma concentrations of HDL cholesterol, the levels of apo AI and apo AII in FED patients are also significantly decreased. In vivo kinetic studies in a FED patient have shown that the hypoalphalipoproteinemia was caused by marked hypercatabolism of apo AI and apo AII, which could be the consequence of structural abnormalities due to the selective LCAT deficiency (Elkhalil et al., 1997). Similar to familial LCAT deficiency, individuals with FED have usually been reported not to be associated with premature coronary artery disease. However, this general concept has been challenged by recent findings that several cases of male patients with FED have been reported to suffer from premature atherosclerosis (Kuivenhoven et al., 1995a; 1995b).
The genetic defect associated with FED was first described in a kindred of German origin (Funke et al., 1991) and in a Dutch family (Kastelein et al., 1992). Homozygosity for a missense mutation in exon 4 resulting in T123I was identified to be the molecular defect for the FED phenotype. Other defects in the LCAT gene associated with the biochemical and clinical phenotype of FED involve the missense mutations in codons 10 (Skretting and Prydz, 1992), 131 (Kuivenhoven et al., 1995b), 391 (Hill et al., 1993a) and a deletion of codon 300 (Klein et al., 1993). Recently, studies with the purified T123I proteins from stably transfected CHO cells have shown that the mutant has an intact core active site and catalytic triad (Adimoolam et al., 1998b). The T123I mutant is unreactive with reconstituted HDL (rHDL) but substantially reactive with LDL and the PNPB substrate as well. Since the binding affinity of the T123I mutant to rHDL as measured with isothermal titration calorimetry and the surface plasmon resonance method is only slightly decreased compared with wild type LCAT, it is concluded that the T123I mutant is most likely defective in activation by apo A-I (Adimoolam et al., 1998b).

1.1.5.3 Classification of LCAT Gene Mutations

To date, around 40 different mutations have been described that result in either FLD or FED. The mutations associated with FED are localized to exons 1, 4, and 6, whereas the mutations resulting in FLD are dispersed throughout the LCAT gene (Glomset et al., 1989). In a recent review, a new classification for the natural mutations of the LCAT gene has been proposed (Kuivenhoven et al., 1997). Based on the characteristics of the clinical and biochemical phenotypes of the patients with LCAT deficiency syndromes as well as the data of functional assessment from the in vitro expression of these mutations, this classification categorizes all known mutations of the
LCAT gene into four distinct classes. The first two classes define the clinical phenotype of FLD in which the complete or near complete loss of LCAT activity is due to null mutations (Class 1), and missense mutations (Class 2). Partial deficiency of LCAT activity associated with FED phenotype is divided into class 3 and 4 according to their lipoprotein substrate specificity. In class 3 missense mutations in the LCAT gene result in an intermediate phenotype between the classes 1, 2 and class 4, respectively. Specifically, these mutations are characterized by either a selective partial loss of activity using LDL as substrate, or a combined partial loss of activity against both HDL and LDL. Class 4 is the typical FED phenotype, i.e., missense mutations which cause the specific loss of LCAT activity against HDL analogues while retaining the activity against apo B-containing lipoprotein (Fig. 3).
Figure 3. Schematic representation of the biochemical characteristics of the 4 different classes of LCAT mutations. This figure is adapted from that in Kuivenhoven et al., 1997.
1.1.6 The Identification of an Unique Genetic Defect in the LCAT Gene

Causing FED

 Mutations in the promoter region, coding sequence, and intron-exon junctions of a gene are usually known to be the underlying cause of human genetic disease, but mutations of this type cannot be always identified in patients with a clinical diagnosis of a certain inherited disorder. This situation was encountered for a case of FED. Only after extending the search for mutations further into the noncoding sequence of the LCAT gene was our lab able to identify a point mutation in intron 4 of the gene as the underlying genetic defect responsible for the biochemical and clinical phenotype of three particular patients with FED (Kuivenhoven et al., 1996).

1.1.6.1 Genetic Studies Identified a Point Intron Mutation in the LCAT Gene

Sequencing of the LCAT gene of all three probands (Fig.4, III 2, 5 and 6) revealed compound heterozygosity for two point mutations: (a) a missense mutation in exon 4 (C\textsuperscript{2182} → T), resulting in the exchange of a threonine residue for an isoleucine at position 123 (T123I) of the mature protein, and (b) a T\textsuperscript{2327} → C nucleotide substitution in intron 4 (IVS4:T→22C), 22 bases upstream of the 3' splice site. Screening of more family members identified five T123I carriers (Fig.4, III 1 and 3, IV 1, 2 and 3) and three carriers of the intron mutation (Fig.4, III 9, 10 and 11). The results of DNA analysis of the family members revealed that the two mutations were located on different alleles. Subsequent screening of 70 normolipidemic controls and 98 unrelated subjects with HDL deficiency demonstrated that this unique intron mutation is not a variant of the LCAT gene at all (Kuivenhoven et al., 1996).
Figure 4. Pedigree of the kindred investigated. This figure shows the results of screening the DNA of family members for the presence of two point mutations, i.e., C\textsuperscript{2182} → T in exon 4 resulting in T\textsubscript{123}I, and T\textsubscript{2327} → C in intron 4, respectively. Squares and circles represent males and females, respectively. The arrow indicates the index patient. Right-sided shading (grey) indicates heterozygosity for T\textsubscript{123}I defect whereas left-sided shading (black) indicates heterozygosity for the intron mutation. Diagonal bar; deceased. Adapted from the figure in Kuivenhoven et al., 1997.
1.1.6.2 Clinical and Biochemical Characteristics of the Compound

Heterozygous Patients with FED

In addition to the dense corneal opacities, the three compound heterozygous sisters were characterized by severely reduced plasma HDL-cholesterol and apo A-I levels (Table 1), a near total loss of plasma LCAT activity, markedly decreased plasma LCAT concentration, and residual specific LCAT activities (Table 2). However, their plasma was still able to generate cholesteryl esters as shown by the low normal CERs. Heterozygotes for either the T123I defect or the intron mutation exhibited low normal HDL-cholesterol and apo A-I levels and approximately 50% reduction in LCAT activity compared with the normal controls, but showed low normal CERs. In addition, plasma LCAT concentrations were in the normal range for the T123I defect whereas carriers of the intron mutation had a half-normal LCAT protein in their plasma (Kuivenhoven et al., 1996). All of these clinical and biochemical characteristics are typical features of FED.

Homozygosity for T123I is known to underlie FED (Funke, 1991; Klein, 1992). The clinical and biochemical phenotype caused by compound heterozygotes is more complicated than that of homozygotes, especially when the two mutations on the two different alleles have distinct impacts on the phenotype. Usually, the phenotype is determined by the defect in the less severely affected allele. Heterozygous carriers of either defect were shown to have different characteristics. The observation that carriers of the intron mutation exhibited highly significantly reduced LCAT concentrations in plasma as compared to T123I carriers suggested that the intron defect might have a much more profound impact on gene expression, and thus underlie the FED phenotype for these patients.
Table 1. Lipids, Lipoproteins and Apolipoproteins of the FED Patients with a Point Mutation (IVS4:T-22C) in Intron 4 and/or exon 4 (T123I) of the LCAT Gene

<table>
<thead>
<tr>
<th>Genetic status</th>
<th>Subject</th>
<th>TC (mmol/L)</th>
<th>TG (mg/dl)</th>
<th>LDL-C (mmol/L)</th>
<th>HDL-C (mmol/L)</th>
<th>apo A-I (mg/dl)</th>
<th>apo B (mg/dl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound heterozygous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For T123I and the intron mutation</td>
<td>III-2</td>
<td>8.43</td>
<td>9.80</td>
<td>nd</td>
<td>0.08</td>
<td>&lt;25</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>III-5</td>
<td>4.96</td>
<td>4.71</td>
<td>2.31</td>
<td>0.05</td>
<td>26</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>III-6</td>
<td>5.33</td>
<td>4.35</td>
<td>3.24</td>
<td>0.11</td>
<td>30</td>
<td>174</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>6.2±1.6</td>
<td>6.3±2.5*</td>
<td>2.8±0.5</td>
<td>0.08±0.02*</td>
<td>19±13*</td>
<td>173±32*</td>
</tr>
<tr>
<td>Heterozygous for T123I</td>
<td>III-1</td>
<td>3.16</td>
<td>0.65</td>
<td>2.13</td>
<td>0.74</td>
<td>90</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>III-3</td>
<td>5.04</td>
<td>1.79</td>
<td>3.13</td>
<td>1.10</td>
<td>147</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>IV-1</td>
<td>4.61</td>
<td>1.53</td>
<td>2.94</td>
<td>0.97</td>
<td>132</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>IV-2</td>
<td>3.60</td>
<td>0.95</td>
<td>2.34</td>
<td>0.83</td>
<td>98</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>IV-3</td>
<td>2.98</td>
<td>0.94</td>
<td>1.91</td>
<td>0.64</td>
<td>81</td>
<td>58</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>3.9±0.8*</td>
<td>1.2±0.4</td>
<td>2.5±0.5*</td>
<td>0.9±0.2</td>
<td>110±25</td>
<td>75±13</td>
</tr>
<tr>
<td>Heterozygous for intron mutation</td>
<td>III-9</td>
<td>3.49</td>
<td>0.69</td>
<td>2.44</td>
<td>1.19</td>
<td>133</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>III-10</td>
<td>4.50</td>
<td>2.77</td>
<td>2.69</td>
<td>0.56</td>
<td>84</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>III-11</td>
<td>6.33</td>
<td>1.48</td>
<td>4.92</td>
<td>0.75</td>
<td>77</td>
<td>148</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>4.9±1.0</td>
<td>1.7±0.9</td>
<td>3.4±1.1</td>
<td>0.8±0.3</td>
<td>98±25</td>
<td>118±28</td>
</tr>
<tr>
<td>Normal</td>
<td>III-4</td>
<td>5.04</td>
<td>0.84</td>
<td>2.79</td>
<td>1.87</td>
<td>203</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>III-7</td>
<td>4.65</td>
<td>1.40</td>
<td>2.88</td>
<td>1.13</td>
<td>137</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>III-8</td>
<td>6.10</td>
<td>1.38</td>
<td>4.19</td>
<td>1.28</td>
<td>161</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>III-12</td>
<td>4.17</td>
<td>2.11</td>
<td>2.57</td>
<td>0.65</td>
<td>91</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>III-13</td>
<td>6.24</td>
<td>2.66</td>
<td>4.11</td>
<td>0.94</td>
<td>110</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>III-14</td>
<td>6.94</td>
<td>2.65</td>
<td>4.69</td>
<td>1.06</td>
<td>129</td>
<td>172</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>5.5±1.0</td>
<td>1.8±0.7</td>
<td>3.5±0.8</td>
<td>1.2±0.4</td>
<td>139±36</td>
<td>116±36</td>
</tr>
</tbody>
</table>

The status of family members was determined by DNA analysis. Data from each group of patients (compound heterozygotes and heterozygotes for either the T123I mutation or intron mutation) were compared with unaffected family members using the student's test: *P<0.009; ‡P<0.024; §P<0.047. Note: subject III-10 was not fasted. This table is adapted from that in Kuivenhoven et al., 1997.
Table II. LCAT Activities, LCAT Concentrations, and CER of the FED Patients with a Point Mutation in Intron 4 (IVS4:T-22C) and/or exon 4 (T123I) of the LCAT Gene

<table>
<thead>
<tr>
<th>Genetic status</th>
<th>Subject</th>
<th>LCAT activity</th>
<th>LCAT concentration</th>
<th>Specific Activity</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(nmol/h/ml)</td>
<td>(µg/ml)</td>
<td>(nmol/h/µg)</td>
<td>(nmol/h/ml)</td>
</tr>
<tr>
<td>Compound heterozygous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For T123I and the intron mutation</td>
<td>III-2</td>
<td>1.40</td>
<td>2.62</td>
<td>0.53</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>III-5</td>
<td>1.16</td>
<td>1.79</td>
<td>0.65</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>III-6</td>
<td>1.10</td>
<td>nd</td>
<td>-</td>
<td>74</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>1.2±0.1*</td>
<td>2.2±0.44</td>
<td>0.6±0.06*</td>
<td>84±7</td>
</tr>
<tr>
<td>Heterozygous for T123I</td>
<td>III-1</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td></td>
<td>III-3</td>
<td>19.7</td>
<td>5.21</td>
<td>3.78</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>IV-1</td>
<td>17.7</td>
<td>5.31</td>
<td>3.33</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>IV-2</td>
<td>19.5</td>
<td>5.00</td>
<td>3.90</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>IV-3</td>
<td>16.7</td>
<td>3.80</td>
<td>4.39</td>
<td>92</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>18.4±1.3*</td>
<td>4.8±0.6</td>
<td>3.9±0.4*</td>
<td>107±13</td>
</tr>
<tr>
<td>Heterozygous for intron mutation</td>
<td>III-9</td>
<td>23.2</td>
<td>1.92</td>
<td>12.10</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>III-10</td>
<td>19.0</td>
<td>2.52</td>
<td>7.55</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>III-11</td>
<td>16.9</td>
<td>2.90</td>
<td>5.82</td>
<td>93</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>19.7±2.6$^+$</td>
<td>2.5±0.4$^+$</td>
<td>8.5±2.6$^+$</td>
<td>97±17</td>
</tr>
<tr>
<td>Normal</td>
<td>III-4</td>
<td>29.3</td>
<td>4.87</td>
<td>6.02</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>III-7</td>
<td>28.0</td>
<td>4.40</td>
<td>6.36</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>III-8</td>
<td>28.8</td>
<td>4.81</td>
<td>5.99</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>III-12</td>
<td>34.1</td>
<td>4.55</td>
<td>7.48</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>III-13</td>
<td>23.2</td>
<td>3.51</td>
<td>6.60</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>III-14</td>
<td>29.6</td>
<td>4.24</td>
<td>6.97</td>
<td>192</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td></td>
<td>28.8±3.2</td>
<td>4.4±0.5</td>
<td>6.6±0.5</td>
<td>131±30</td>
</tr>
</tbody>
</table>

The status of family members was determined by DNA analysis. LCAT activity, LCAT concentration and CER were determined in whole plasma samples. Data from each group of patients (compound heterozygotes and heterozygotes for either the T123I mutation or intron base substitution) were compared with unaffected family members using the Student's test: *P<0.0005; $^+$P<0.001; $^+$P<0.007; $^+$P<0.0008. Adapted from Kuivenhoven et al., 1997.
1.1.6.3 In vitro Studies Demonstrated that the Intron Mutation Resulted in Defective Intron Splicing

The T→C point mutation is located in the putative branchpoint sequence (BPS) of intron 4 of the LCAT gene, a region that plays a crucial role in the splicing of nuclear mRNA precursors. In order to investigate the hypothesis that the intron mutation affects the RNA splicing or causes a change in the expression of the allele, studies of the in vitro expression of an LCAT minigene containing the intron mutation were performed (Kuivenhoven et al., 1996). The results revealed that this genetic defect causes a null allele as the result of the complete intron retention, which indicated that the intron mutation is indeed causative for the half normal LCAT concentration observed in heterozygous carriers. These observations also suggest that homozygosity for the intron mutation would be associated with classic familial LCAT deficiency whereas carriers of the T123I defect can be defined as heterozygote for FED. Most importantly, the finding that a point mutation in a lariat branchpoint consensus sequence can result in a human genetic disease indicates the functional significance of the branchpoint sequence in the splicing of human nuclear mRNA precursors.

1.2 RNA Splicing

Our identification of the mutation in the branchpoint region of intron 4 of the LCAT gene that causes a null allele as the result of the intron retention (Kuivenhoven et al., 1996) is one of the first to be characterized fully. More mutations in the branchpoint sequence of introns have been recently reported to underlie human genetic diseases although these mutations have different consequences with regard to splicing patterns, such as exon skipping, the use of alternative splice site, and intron retention. All of
these mutations in the branchpoint sequence have demonstrated the importance of this sequence in mRNA splicing.

1.2.1 Other Mutations in the Branchpoint Consensus Sequence of Introns that Cause Human Genetic Disorders

In addition to the point mutation in intron 4 of the LCAT gene in patients with FED, more mutations in the potential branchpoint sequence have been described to cause human genetic disorders. Perhaps, the first case that suggested an association between a mutation in a branch point signal and a human disease was reported by Rosenthal et al. A single intronic A→C base change 19 bp upstream of a splice acceptor site in the gene encoding the neural cell adhesion molecule L1 was identified in a family with X-linked hydrocephalus (Rosenthal et al., 1992). This point mutation that resides in a potential branch point signal gives rise to an abnormal L1 mRNA containing a 69 bp insertion and a 116 bp deletion, which could result from the utilization of an alternative branch point sequences and splice sites. On the other hand, the deletion of a branchpoint sequence in the androgen receptor gene is the probable cause of a case of X-linked Reifenstein syndrome (partial androgen insensitivity). The mutation was a large deletion (>6 kb) in intron 2 that left 18 bp of normal sequence intact at the 3' splice end of the intron, including the intron:exon junction (Ris-Stalpers et al., 1994). The deletion of the intron 2 branchpoint sequence results in 90% inhibition of wild-type splicing. Analysis of mRNA by RT-PCR showed the presence of two products, a major one (92%) in which the downstream exon 3 was skipped and a minor one that was normally spliced, apparently through the use of a cryptic branch point sequence.

Recently, Brand et al. reported a A→G mutation in intron 1 of the hepatic lipase gene that has been shown to be the underlying molecular defect in a patient with
hepatic lipase (HL) deficiency (Brand et al., 1996). This mutation that is located 13 bp upstream of the 3' splice site destroys a potential branchpoint signal and generates an additional 3' splice site, which disrupts normal splicing of intron 1 and leads to alternative splicing as well as the accumulation of unspliced mRNA. Another example of a mutation that destroys the only branchpoint consensus sequence of intron 9 of the LDL receptor gene was identified in a patient with familial hypercholesterolaemia (FH) (Webb et al., 1996). The destruction of the branchpoint consensus sequence, especially the loss of the adenosine residue, caused by a minor rearrangement 30 bp upstream of the 3' splice site, has been shown to result in retention of the intron although a small amount of alternative splicing products derived from the use of alternative branchpoint sequences and splice sites in exon 10 could also be detected.

More recently, a point mutation in intron 32 (IVS32:T→25G) of a collagen gene has been described to result in skipping of exon 33 in two British families with Ehlers-Danlos syndrome (Burrows et al., 1998). Unlike the mutations mentioned above, this mutation which lies only 2 bp upstream of a highly conserved adenosine in the consensus branch-site sequence causes a 45-bp deletion corresponding to exon 33 from the mRNA in ~60% of transcripts of the mutant gene. Because intron 33 is relatively small (135 bp), its internal branchpoint sequence could be preferred to the mutated branch site in intron 32, which consequently leads to the loss of COL5A1 exon 33. These results demonstrate that the base changes in the branchpoint sequence of introns are the underlying causes of defective splicing for certain human diseases.

In addition to mutations in the branch point sequence, base changes in the vicinity of the branchpoint sequence also affect the splicing and are associated with human genetic disorders. Maslen et al. have identified a point mutation (IVS30:G26→T) immediately upstream of the branchpoint sequence of intron 30 of the fibrillin-2 gene in
a large well-characterized kindred with congenital contractural arachnodactyly (Maslen et al., 1997). Mutational analysis of cDNA from the patients revealed the skipping of exon 31 with the mutant transcripts constituting ~25% of the total mRNA species. This mutation may influence the branchpoint selection due to the fact that this particular intron lacks the lengthy polypyrimidine tract which is characteristic of the introns of higher eukaryotes. Thus, missplicing occurs when the next downstream 3' splice site is utilized as an alternative. By contrast, a point mutation immediately located downstream from a potential branch point at -26 of intron 1 of ferrochelatase gene has been suggested to be the molecular defect in a patient with erythropoietic protoporphyria (Nakahashi et al., 1992). A C→T base change, 23 bp upstream relative to the 3' splice site, inhibits the normal splicing of the wild type of intron 1, and thus results in the alternative splicing between exon 1 and exon 3. It was supposed that the C→T mutation might affect the interaction between the putative branch site at -25 to -31 and the U2 small nuclear ribonucleoprotein (U2 snRNP) and lead to exon 2 skipping.

Taken together, these observations clearly indicate that the consensus branchpoint sequence, although only weakly conserved in humans, is essentially required for the efficient and accurate splicing of nuclear pre-mRNA, and defects in mRNA splicing due to the mutations in or near the branchpoint sequence can indeed cause human genetic disorders.

1.2.2 Splicing Process

1.2.2.1 Role of Splicing in Gene Expression

Eukaryotic genes are frequently interrupted by non-coding intervening sequences (or introns) that are removed from the primary transcripts. Nuclear pre-mRNA splicing is the process by which these intervening sequences in messenger RNAs are precisely
removed and the functional coding sequences (exons) ligated. The removal of the introns from the pre-mRNA to yield the mature mRNA is an essential step in gene expression since the intronic sequences often contain a premature stop codon that generates truncated, i.e., nonfunctional, proteins upon translation. In addition, the presence of introns within pre-mRNA also provides a means for the regulation of gene expression. In higher eukaryotes some pre-mRNAs undergo alternative (differential) splicing, which can generate different proteins in a cell-specific and temporal manner from a single gene (Lopez, 1998).

1.2.2.2 Two Steps of Trans-esterification Splicing Reaction

Splicing of the pre-mRNA occurs via a two-step transesterification reaction carried out by a large and complex ribonucleoprotein particle called the spliceosome. Splicing is initiated by the nucleophilic attack of the 2' hydroxyl group of the branch point adenosine residue of an intron on the phosphodiester bond at the 5' splice site. At the same time when cleavage at the 5' splice site occurs, the 5' end of the intron is joined to the branch site adenosine in an unusual 2'-5' phosphodiester bond, producing a free 5' exon and the intron-3' exon lariat intermediate. During the second step, the 3' hydroxyl of the 5' exon attacks the phosphodiester bond at the 3' splice site, resulting in the ligation of the two exons and the release of the intron in a lariat form (Fig.5).
Figure 5. Schematic representation of the splicing process of an intron. Boxes and lines denote exons and introns, respectively. The branch site adenosine residue within the intron is in boldface.

1.2.2.3 Sequences Required for Splicing

Introns are functionally defined by consensus sequences at the 5' splice site, the 3' splice site, and the branch site.

1.2.2.3.1 5' Splice Site Sequence

In higher eukaryotes almost all introns have the dinucleotide GU at their 5' ends. Sequence comparison of the 5' splice sites from a large number of genes have yielded the consensus: AG/GURAGY (/, cleavage site; R, purine; Y, pyrimidine) (Fig.6) in mammals (Padgett et al., 1986). Mutations of the invariant GU dinucleotide completely inactivate the 5' splice site. By contrast, mutations at other positions within the 5' splice
site may significantly reduce the amount of correctly spliced mRNA, activate a nearby cryptic 5' splice site, or both. In the yeast, the 5' splice site is more highly conserved with the sequence consensus being AG/GUAUGU.

\[
\begin{align*}
5' \text{ SS} & \quad \text{BPS} & \quad 3' \text{ SS} \\
\text{AG|GURAGU} & \quad \text{YNYURAY} & \quad (Y)_{n}\text{NYAG|G} \\
& \quad \text{18-40 nt}
\end{align*}
\]

Figure 6. Schematic representation of the consensus of 5' splice site, 3' splice site, and branchpoint sequence. SS: splice site; BPS: branchpoint sequence.

1.2.2.3.2 3' Splice Site Sequence

The mammalian 3' splice site confirms to the consensus sequence YAG/G and is preceded by a stretch of ten or more pyrimidines in the pre-mRNA (Fig. 6). The AG dinucleotide that defines the 3' end of the intron is invariant and the string of pyrimidines is referred to as the polypyrimidine tract (Py tract) (Mount, 1982). For yeast, 3' splice sites have the consensus sequence CAG/G and lack the extensive polypyrimidine tract.

In most introns, the first AG downstream of the BPS usually serves as the 3' splice site, and this AG is typically located 18-40 nucleotides downstream of the BPS (Green, 1991). The AG dinucleotide appears to be recognized by a scanning mechanism that begins at the branchpoint in a 5' to 3' direction (Smith et al., 1989; Helfmen et al., 1989). The 3' splice site YAG motif is dispensable for the first step of splicing, but is absolutely required for the second catalytic step. Mutation or deletion of the AG dinucleotide blocks exon ligation although cleavage at the 5' splice site and
formation of the lariat intermediate still occur (Reed & Maniatis, 1985; Ruskin & Green, 1985a).

The important role of the polypyrimidine tract (Py tract) of mammalian introns in splicing is underscored by the fact that the efficiency of the first step of splicing is markedly reduced when a large portion of the Py tract is deleted (Ruskin and Green, 1985a). The length of the polypyrimidine tract also provides distinct properties for splicing. Progressive reduction of the Py tract results in a decreased splicing efficiency in vivo and can eventually result in utilization of a cryptic site at another AG dinucleotide (Santen et al., 1985; Wieringa et al., 1984). This is probably because both the length and content of the Py tract affect the binding of U2 snRNP auxiliary factor (U2AF) to this region between the BPS and the 3' splices site (see Section 1.2.6.1 below). There is also evidence that mutations in the Py tract that result in defective splicing could be the underlying cause of genetic disorders. For example, a T→A transversion in the Py tract of intron 4, 11 bp from the 3' splice site, of the gene for steroidogenic acute regulatory protein causes a deletion corresponding to all of exon 5 in a patient with congenital lipoid adrenal hyperplasia (Tee et al., 1995). Another Py tract mutation was described in a patient with severe poikilocytic anemia (Fournier et al., 1997). Similarly, in this mutation a T→G base change 13 bp upstream of the 3' splice site of intron 19 of the alpha-spectrin gene creates a new acceptor splice site, AT→AG, and produces two abnormal mRNAs. One of the mRNAs contains a 12 bp intron sequence insertion upstream of exon 20, while there is skipping of exon 20 in the other. These observations further emphasize the important role of the Py tract in RNA splicing.

1.2.2.3.3 The Branchpoint Sequence (BPS)
The formation of a lariat intermediate that contains a 3'-exon and a branched intron with the 5' end of the intron linked to an internal branchpoint adenosine residue via a 2'-5' phosphodiester bond during splicing suggests that a specific sequence within the intron participates directly in the pre-mRNA splicing process. This important sequence known as the BPS is usually found at a distance of 18 to 40 nucleotides upstream of the 3' splice site. In yeast, the branch site UACUAAC is absolutely conserved. Mutation of the conserved branch site in yeast leads to a block in both steps of splicing in vivo (Langford et al., 1984; Vijayraghavan et al., 1986) and in vitro (Newman et al., 1985). A series of elegant genetic experiments has shown that base pairing occurs between the BPS and a conserved region (5'-GUAGUA-3') of U2 small nuclear RNA (snRNA) in both yeast and mammalian cell lines (Parker et al., 1987; Wu & Manley, 1989; Zhuang and Weiner, 1989). Although the U2 small nuclear ribonucleoprotein particle (snRNP) binds to this sequence element through complementary base pairing, the BPS exhibits a weak consensus YNYURAY (where A represents the site of branch formation; Y, pyrimidine; R, purine; N, any nucleotide) (Fig.6) in mammals, suggesting that the binding of U2 snRNP is inherently weak. Association of U2 snRNA with the branchpoint region occurs early during the splicing pathway and requires U2AF bound to the adjacent polypyrimidine tract (Ruskin et al., 1988; Zamore et al., 1992; see Section 1.2.6.1). It has been shown that the binding of U2AF to the polypyrimidine tract between BPS and the 3' splice site facilitates the U2 snRNP-the lariat BPS interaction. This explains, at least in part, why changes within the lariat branchpoint sequence can be tolerated to a high degree. On the other hand, it is particularly difficult to detect the effects of branch site mutations on the splicing, due to cryptic branch site activation (Padgett et al., 1985; Ruskin et al., 1985; Reed & Maniatis, 1988). In either case, mutations within the BPS might result in a slightly decreased
efficiency of splicing and therefore would not present a disease phenotype. However, growing evidence, as reviewed in Section 1.2.1 and the studies from our lab as well, have indicated that these are not always the case. Mutations within the lariat branchpoint sequence can indeed cause human genetic disease, which further underscores the importance of the BPS in RNA splicing.

1.2.3 The spliceosome assembly pathway

Nuclear pre-mRNA splicing is carried out by a multicomponent complex termed spliceosome. The assembly of a series of highly dynamic spliceosomal complexes that assemble on pre-mRNA occurs in the order of E→A→B→C, with the catalytic steps of splicing taking place in the C complex (Fig. 7). These complexes consist of five small nuclear ribonucleoprotein particles (snRNPs) (U1, U2, U4, U5 and U6) and more than 50 distinct non-snRNP splicing factors (Moore et al., 1993; Madhani and Guthrie, 1994; Reed, 1996; Will and Luhrmann, 1997).

Mammalian snRNPs, present in about $10^4$-$10^6$ copies/cell, are comprised of the small, metabolically stable RNA component, the snRNA (100-200 nt), and a set of eight common "Sm" proteins B', B, D1, D2, D3, E, F and G which form the structural core of the snRNPs except U6. These proteins play an essential role in the biogenesis of the snRNPs and are the major targets for the so-called anti-Sm auto-antibodies from patients with systemic lupus erythematosus (Tan, 1989). The association of these proteins is dictated by the presence of a conserved Sm binding sequence in the snRNA (Luhrmann et al., 1990). Each snRNP also contains a set of specific proteins. The snRNAs are RNA polymerase II transcripts and contain a unique trimethylguanosine cap nucleotide. The exception is the highly conserved U6 snRNA that does not contain a Sm binding site but is immunoprecipitable by an anti-Sm antibody by virtue of its
extensive base-pairing interaction with the U4 snRNA. Specific snRNA-pre-mRNA and snRNA-snRNA interactions play key roles in spliceosome assembly (Fig.7) and are generally believed to be involved in catalyzing the two steps of trans-esterifications of the splicing reaction (Newman, 1994; Neilsen, 1994) (Fig.5 and 7). Recently, the contribution of non-snRNP protein factors, especially the serine-arginine rich (SR)-proteins, to the splicing of nuclear pre-mRNA has been extensively investigated. These proteins play an essential role in the recognition and pairing of the 5' and 3' splice sites. In conjunction with snRNA, they may even directly contribute to the spliceosome's active sites (Will and Luhrmann, 1997).
Figure 7. Schematic representation of assembly of splicing complexes
1.2.3.1 Spliceosomal E or Early Complex

Splicing of the pre-mRNA is initiated by the ATP-independent formation of the spliceosome E complex that commits the substrate to the splicing pathway. The spliceosomal E complex in the mammalian system contains U1 snRNP as well as non-snRNP protein factors, including the U2AF and members of the SR protein family (Staknis and Reed, 1994a). The earliest specific event in the formation of the spliceosome is the binding of U1 snRNP to the 5' splice site. A similar ATP-independent complex in yeast, known as the commitment complex, was demonstrated to contain only the U1 snRNA (Ruby and Abelson, 1988; Seraphin and Rosbash, 1991). Unlike the commitment complexes in the yeast, U1 snRNP-containing complexes are difficult to visualize by native polyacrylamide gels in the mammalian system. The abundant U1 snRNP contains a 165 nucleotide long U1 snRNA. In addition to the set of Sm proteins common to other abundant snRNPs, the purified U1 snRNP also contains a set of three specific proteins, the 70K, A and C proteins (Query et al., 1989a; Scherly et al., 1989). The 70K and A proteins bind specifically to stem/loop I and stem/loop II of the U1 snRNA, respectively (Query et al., 1989b; Surowy et al., 1989). The U1 C protein does not bind to naked U1 snRNA, but is associated with U1 70K and Sm proteins (Nelissen et al., 1994).

The interaction of U1 snRNP with the 5' splice site involves base pairing between the 5' end of the U1 snRNA and the consensus 5' splice site sequence. This was first suggested from the observed base-pairing potential between these two sequences (Lerner et al., 1980; Rogers and Wall, 1980) and has been supported by a number of experiments. The most conclusive evidence comes from the in vivo suppression of 5' splice site mutants with compensatory changes in U1 snRNA (Zhuang and Weiner, 1986). Further evidence also comes from in vitro psoralen crosslinking of U1 snRNA to
the 5' splice site (Wassarman and Steitz, 1992). The role of the individual U1 specific proteins in mediating this base-pairing interaction was also investigated in biochemical depletion experiments. The base-pairing interaction was significantly reduced with a U1 snRNP lacking the C protein suggesting that the C protein plays an important role in mediating the interaction with the 5' splice site (Heinrichs et al., 1990; Jamison et al., 1995). The mechanism whereby the U1-C protein facilitates E-complex formation is not well understood, but studies have suggested that its amino-terminal domain is sufficient for its activity (Will et al., 1996) and that, when present in the U1 snRNP, U1-C protein interacts directly with the pre-mRNA (Rossi et al., 1996).

In addition to the 5' splice site, efficient E complex formation also requires the polypyrimidine tract at the 3' splice site. U2AF bond to the pyrimidine tract is essential for this early step of spliceosome assembly, which suggests a functional association between the 5' and 3' splice sites as early as E complex (Michaud and Reed, 1993). In some yeast introns (Schizosaccaromyces pombe), U1 snRNA also base pairs to the 3' splice site, but this interaction has not yet been observed in other organisms (Reich et al., 1992; Ares and Weiser, 1995; Madhani and Guthrie, 1994).

1.2.3.2 Spliceosomal A Complex

In the first ATP- and temperature-dependent step of spliceosome assembly, the U2 snRNP stably binds to the branch site to form the spliceosomal A complex, the first complex that can be detected by native gel electrophoresis in the pathway (Konarska and Sharp, 1986; 1987). The mammalian U2 snRNP consists of 187 nucleotide highly conserved RNA sequence associated with the common Sm proteins and unique proteins, A' and B". The function of U2 snRNA in splicing is mediated by the proteins of the U2 snRNP. Among the spliceosomal snRNAs, U2 has the most extensive
posttranscriptional modifications, including a 5’ trimethylguanosine cap, ten 2’-O-methylated residues and 13 pseudouridines, which are essential for spliceosome assembly as well as splicing (Yu et al., 1998). A series of genetic experiments have demonstrated that the binding of U2 snRNA to the branch site involves the base-pairing interaction (Parker et al., 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989). U2 snRNA base pairs with BPS during spliceosome assembly, resulting in a helix from which the branchpoint adenosine is bulged (Madhani and Guthrie, 1994). Direct evidence for a bulged nucleophile has recently been provided by a set of experiments using site-specific modifications in the branch-site sequence (Query et al., 1994). The branchpoint adenosine itself also plays a role in A complex assembly, and site-specific labeling studies have identified a 14 kDa protein that appears to specifically recognize this base in the complex (Query et al., 1996). In addition, a set of six highly conserved U2 snRNP proteins (Kramer et al., 1995; Gozani et al., 1996; Wells et al., 1996) UV cross-link to pre-mRNA near the BPS. Interaction of these proteins with a 20-nucleotide region upstream of the BPS, designated the anchoring site, is essential for anchoring U2 snRNP to pre-mRNA during the A complex assembly (Gozani et al., 1996). It has also been shown that the U2 snRNP protein SAP 155 UV cross-links to pre-mRNA on both sides of the BPS in the A complex. Since SAP 155’s downstream cross-linking site is immediately adjacent to the U2AF binding site, it is proposed that U2AF binds to the pyrimidine tract in the E complex and then interacts with SAP 155 to recruit U2 snRNP to the BPS (Gozani et al., 1998). The association of U2 snRNP with the branch site also requires other non-RNP factors, such as SF1 and SF3 (Kramer and Utans, 1991; Kramer, 1992) and SC35 (Fu and Maniatis, 1992a; 1992b).

1.2.3.3 Spliceosomal B Complex
Following the stable binding of U1 and U2 to the pre-mRNA, U4, U5, and U6 snRNPs join the spliceosome as a tri-snRNP particle, which leads to the formation of spliceosomal B complex (Konarska and Sharp, 1987; Bindereif and Green, 1987; Jamison et al., 1992). In the tri-snRNP, U4 and U6 snRNPs are linked through extensive intermolecular base pairing. U6 snRNA is the most highly conserved spliceosomal snRNA, and functional analysis (Moore et al., 1993; Rymond and Rosbash, 1992) has drawn particular attention to sequences in the central region of U6, including the invariant motif ACAGAG.

After the addition of the U4/U6-U5 triple snRNP to the spliceosome, a conformational rearrangement takes place where the base-pairing interaction between U4 and U6 snRNAs is disrupted resulting in the dissociation of the U4 snRNP (Lamond et al., 1988). Now the U6 snRNA can interact directly with U2 snRNA that is still bound to the branch site. Residues in the U6 snRNA adjacent to the invariant ACAGAG base pair with a sequence in U2 that is immediately upstream of the branchpoint recognition region, which forms a short helix (helix I) proposed to be very close to the catalytic center of the spliceosome (Hausner et al., 1990. Wu and Manley, 1991; Datta and Weiner, 1991; Guthrie, 1991).

The U1–5' splice site base pairing interaction is also disrupted prior to the first catalytic step and is replaced by an interaction between U6 and 5' splice site. Genetic evidence exists for an interaction between residues in the ACAGAG sequence of U6 and the conserved sequence at the 5' splice site (Lesser and Guthrie, 1993; Sun and Manley, 1995). U6 is also observed crosslinked to the 5' splice site as well as with the branch site (Wassarman and Steitz, 1992; Sawa and Abelson, 1992). Furthermore, another intermolecular helix (helix III) between U2 and U6 snRNA has been recently
identified (Sun and Manley, 1995). These interactions would help to directly juxtapose the 5' splice site and the branch site to initiate the first catalytic step in splicing.

The U5 snRNA also appears to be involved in the 5' and 3' splice site definition. U5 snRNA carries an invariant sequence of nine nucleotides (5'-GCCUUUAC-3') at the top of a stem structure (Frank et al., 1994). Biochemical and genetic data show that this U5 loop interacts with exon sequences at the 5' and 3' splice sites in the first and second steps of splicing (Newman and Norman, 1992; Wyatt et al., 1992; Cortes et al., 1993; Sontheimer and Steitz, 1993; Newman, 1994). The U5 loop contacts the 5' exon in the pre-mRNA and this interaction is maintained through both catalytic steps, while contact between the 3' exon and the U5 loop occurs only after the first catalytic step. Based on these observations, it has been proposed that the function of the loop is to hold the 5' exon after the first step of splicing and possibly to help align it with the 3' exons for the second step of splicing (Newman and Norman, 1992; Sontheimer and Steitz, 1993). As exon sequences at the splice sites are not well conserved, these weak base-pairing interactions between the conserved first loop of U5 snRNA and the exon sequences suggest that U5 snRNP proteins are required for this stabilization. U5 snRNP contains a number of specific proteins in addition to the common Sm set, among which the human U5-220 kD protein has been implicated for this role, since it has been shown to make extensive contacts with pre-mRNA sequences at the 5' and 3' splice sites, including those contacted by the U5 snRNA loop (Wyatt et al., 1992; Umen and Guthrie, 1995).

Taken together, all of these interactions between snRNPs and with the 5' splice site and the branch site are proposed to align the substrate pre-mRNA into a Holliday-like structure for the first step of the splicing reaction (Steitz, 1992).
1.2.3.4 Spliceosomal C Complex

The appearance of the free 5' exon and lariat-intermediate from the first step of the splicing reaction marks formation of the C complex. Multiple snRNA-pre-mRNA and snRNA-snRNA interactions are involved in recognition and aligning the 5' splice site and the branch site before the first catalytic step, but much less is known about events that lead to 3' splice-site selection and to the second catalytic step. Genetic and biochemical experiments have identified residues in the U2 and U6 snRNAs that are critical for this transesterification (Umen and Guthrie, 1995). It has been suggested that the newly formed U2-U6 helix in the spliceosome serves to create the catalytic site for splicing and U4 acts as a negative regulator of U6 snRNA (Guthrie, 1991). Consistent with this model, *in vitro* crosslinking of U6 snRNA to the 5' splice site is seen prior to the first step of splicing and in the vicinity of the branch site during the second catalytic step (Sawa and Shimura, 1992). Coincident with U6 interacting with the pre-mRNA a second conformational change occurs within spliceosomal C complex where the second step of the reaction takes place.

Recently, U5 snRNP has been shown to play an important role in the recognition of 3' splice site during the second step of splicing reaction in mammals. It was demonstrated that U2AF$^{65}$, which binds to the polypyrimidine tract early in spliceosome assembly, is replaced by three U5 snRNP proteins of 110, 116 and 220 kDa in the spliceosomal complex C prior to catalytic step II (Chiara *et al.*, 1997). These proteins interact with pre-mRNA in the region spanning from immediately downstream of U2 snRNP's binding site at the BPS to just beyond the 3' splice site, suggesting that U2-U5 snRNP interactions (direct or indirect) play a role in selecting the 3' splice site for the second step of splicing reaction. It is thus proposed that selection of the first AG downstream from the BPS involves a mechanism in which the U5 snRNP proteins are
positioned by interaction with the polypyrimidine tract and with U2 snRNP at the BPS. The AG closest to the 3' boundary of the U5 snRNP binding site is in turn selected as the 3' splice site (Chiara et al., 1997). The selection of the AG itself may involve an interaction between a U5 snRNP protein and an AG recognition factor, such as AG75 (Chiara et al., 1996).

After the transesterification reactions, the spliced RNA and lariat intron are released from complex C followed by disassembly of the spliceosome. The lariat intron is linearized by a debranching enzyme (Ruskin and Green, 1985b) and subsequently degraded, and the spliceosome assembly becomes available for the next cycles.

1.2.4 The Exon/intron Definition Model

1.2.4.1 Internal Exon Splicing

The investigation of the splicing process and spliceosome assembly in particular has largely been performed with pre-mRNAs containing naturally short or artificially shortened introns because of technical constraints and the results have suggested a model where the splicing machinery spans the intron. In contrast to the yeast, the splice site sequences in vertebrate introns are poorly conserved, and pre-mRNAs are often extremely long, usually containing multiple introns that can range up to 100 kb in length, whereas internal exons are, on average, 137 nucleotides in length (Berget, 1995). In addition, mRNA precursors contain many copies of sequences that match the splice site consensus sequences. Therefore, on the basis of splicing signal alone, it is almost impossible to distinguish exons from introns, especially for the pre-mRNA containing multiple exons with long introns. Berget and co-workers observed that a 5' splice site on the downstream site of an exon promotes splicing of the intron immediately upstream (Robberson et al., 1990). Based on these observations, it was proposed that initial
splice-site recognition requires an interaction between 5' and 3' splice sites across the short exons, and this process where both ends of the exon are recognized together was called the "exon-definition" (Berget, 1995) (Fig. 8).

Figure 8. Schematic representation of exon-definition model. During the formation of the E complex, the interaction of U1 snRNP with the 5' splice site facilitates the binding of 65-kDa U2 snRNP auxiliary factor (U2AF \(^{65}\)) to the polyprimidine tract (Py tract). This exon-bridging interaction is most likely mediated by an interaction between SR (serine-arginine-rich) proteins bound at the exon splicing enhancer (ESE) and the U1-70k protein of U1 snRNP and between SR proteins and the small subunit of U2AF (U2AF \(^{35}\)). The mammalian branchpoint binding protein (SF1/mBBP) binds specifically to the branchpoint sequence (BPS) and interacts with the U2AF \(^{65}\). This cooperative binding interaction (Berglund et al., 1998) would facilitate initial recognition of the BPS region. Modified from the figure in Smith and Valcárcel, 2000.

According to the exon-definition model, interactions spanning the exon are first established and the exon is then defined. Following definition of the exon, neighboring exons are subsequently brought together via the spliceosome assembly process. Studies on these interactions across exons indicate that the binding of U1 snRNP to a 5' splice site facilitates the binding of the essential splicing factor U2AF \(^{65}\) to the upstream
3' splice site (Hoffman and Grabowski, 1992; Cunningham et al., 1995). This exon-bridging interaction is mediated by splicing components including members of the SR protein family (Fu, 1995; Manley and Tacke, 1996; see Section 1.2.6.2 below) between the U1 snRNP at the 5' splice site and U2AF at the 3' splice site. Thus, the strength of a 5' splice site at the downstream site of an exon could be a crucial determinant in the efficiency of splicing of the intermediate upstream intron. In agreement with this expectation, improvement of a 5' splice site that increases base-pairing of this site to U1 snRNA activated splicing of the immediate upstream intron (Kuo et al., 1991; Grabowski et al., 1991), whereas mutation of a 5' splice site depressed the removal of the upstream intron 20-fold (Talerico and Berget, 1990). Further support for the exon-definition model is provided by observations of numerous human diseases caused by a single splice site mutation (Krawczak et al., 1992). Such mutations often result in either the skipping of the entire exon or the activation of a nearby cryptic splice site as seen by a comparison of the genomic and the cDNA sequences. Exon skipping is the most common phenotype of a splice site mutation. In the exon-definition model, mutation of the splice site at one side of an exon could inhibit the recognition of the remaining wild-type splice site across the exon. In the absence of the exon-definition model, such mutations would be supposed to give rise to intron retention instead. The other common phenotype of a splice site mutation is the activation of a nearby cryptic splice site to replace the mutant site (Krawczak et al., 1992). In the presence of the mutation, the nearby cryptic splice site can be activated by the same exon-bridging interaction that normally is directed to the wild-type site.

The model also predicts constraints on the length of the exon. Most vertebrate internal exons fall within a narrow size range of 50-250 nucleotides (Hawkins, 1988). Expanding the internal exon with strong constitutive splice signals to greater than 300
nucleotides inhibits ATP-dependent spliceosome assembly (Robberson et al., 1990), and deleting the internal exon derived from the human beta-globin gene below 50 nucleotides leads to exon skipping presumably due to steric hindrance between the factors that recognize splice sites (Dominski and Kole, 1991). Thus, it seems that exon-spanning interactions are involved in the recognition of most splice sites. However, these interactions are not always essential. It is thought that long exons are uncommon because they would not allow for exon-bridging interactions (Robberson et al., 1990), but an in vivo analysis of internal exon size showed that some exons as large as 1200 bp were still spliced (Chen and Chasin, 1994), suggesting that either these exons are defined through mechanisms other than exon bridging to ensure proper splice site recognition or perhaps they contain a secondary structure that allows them to maintain communication between the two ends. Similarly, there are many exons that are very small (<20 nucleotides) (Hawkins et al., 1988) but that function quite well. It is difficult to envision how these small exons could accommodate splicing complexes at both ends with bridging factors between them. In these situations, there must be other means of recognizing short exon splice sites. In fact, just making the splice site sequences strong matches to the consensus increases splicing efficiency of internal short exons (Dominski and Kole, 1992). The enhancement of some short-exon splice sites may come from an exon-bridging interaction spanning the entire exon-intron-exon unit followed by the recognition of the intron, if the distance is in the range of maximum size (Sterner and Berget, 1993). In another case, it has been found that an intron element containing six copies of a G-rich 7 nucleotide sequence termed intron splicing enhancer is required for the splicing of a 6 nucleotide exon from chicken cardiac troponin T (Carlo et al., 1996).
Finally, the observation that a trans-splicing reaction can occur when the 3' substrate contains a downstream 5' splice site or an exonic enhancer further supports the exon-definition model (Chiara and Reed, 1995). This result suggests that efficient trans-splicing occurs only after definition of the 3' splice site by exon spanning interactions with a downstream 5' splice site or exonic enhancer element.

1.2.4.2 First and Last Exon Splicing

Since the exon-definition model can only apply to internal exons, 5' and 3' terminal exons must have other mechanisms to recognize their single splice sites. Indeed, there have been several lines of evidence implicating that the terminal exons use their unique terminal structures to provide enhancement of splice site recognition in the absence of normal exon-bridging interactions. All pre-mRNAs transcribed by RNA polymerase II have a unique 7-methylguanosine cap. A cap-binding protein complex, which specifically recognizes the monomethyl guanosine cap structure, was shown to be essential for in vitro splicing of single substrates (Izaurrealde et al., 1994). Furthermore, experiments on the in vitro splicing of a two-intron pre-mRNA revealed that the absence of the normal 5' cap nucleotide suppressed removal of the first intron but had little effect on the second intron (Ohno et al., 1987). Recently, the cap-binding complex has been demonstrated to be required for efficient recognition of the 5' splice site by U1 snRNP during formation of E complex on a pre-mRNA containing a single intron but not for splicing of the distal intron in a pre-mRNA containing two introns (Lewis et al., 1996). These results indicate that the 5' cap structure is essential for recognition and splicing of the first intron probably via interactions between factors that recognize the cap and 5' splice site.
At the other end of the transcript, the 3' terminal exon contains the poly-A site. Similar to the first exon, mutation of the AAUAAA polyadenylation consensus sequence inhibits in vitro splicing of proximal but not distal introns (Niwa and Berget, 1991), suggesting the existence of interaction between polyadenylation site and splicing factors that specify the 3' splice site. In this respect, the U1 snRNP and its component proteins have been observed to bind in the region of poly-A sites, which indicates the enhancement of a downstream poly-A site by a equivalent mechanism to exon spanning interactions in an internal exon (Wassarman and Steitz, 1993; Lutz and Alwine, 1994).

1.2.4.3 Intron-definition Model

As described above, when internal exons in a range of 50 to 500 nucleotides are flanked by long introns, interactions can occur across the exon to stabilize the binding of U1 snRNP and U2AF. When introns are short, U1 snRNP at the 5' splice site and U2AF at the polypyrimidine tract can interact across the intron through splicing factors to stabilize binding and juxtapose the splice sites in an early spliceosome complex (Reed, 1996). This process is termed "intron-definition" (Berget, 1995) (Fig.9). Experiments on pre-mRNAs with small introns, particularly in vitro have shown that the intron, rather than the exon, is used as the unit for initial recognition of splice sites (Guthrie, 1991; Ruby and Abelson, 1988; Rosbash and Seraphin, 1991). In addition, mutations of the 5' splice site in genes with small introns have a different phenotype than the same mutations in genes with large introns. These mutations resulted in intron retention rather than exon-skipping without the splicing of neighboring introns being affected, which further suggests that intron-definition occurs with short vertebrate introns (Talerico and Berget, 1994). Thus it appears that vertebrate splice sites may be recognized initially by
exon- or intron-definitions depending on the architecture of the pre-mRNA (Sterner et al., 1996).

Figure 9. Interactions in the E complex of intron-definition model. U1 snRNP proteins and cap-binding proteins (CBC20 and CBC80) first interact with pre-mRNA in the region containing the 5' splice site. Mammalian branchpoint binding protein (SF1/mBBP), 65-kDa and 35-kDa U2 snRNP auxiliary factors (U2AF<sup>65</sup> and U2AF<sup>35</sup>) recognize the branchpoint sequence sequence (BPS), the polypyrimidine tract (Py tract), and the AG at the 3' splice site, respectively. SR proteins may also function in cross-intron bridging in the E complex. Modified from the figure in Reed, 2000.

1.2.5 Enhancer Elements

A number of experiments have shown that sequences other than downstream 5' splice sites can enhance spliceosome assembly and splicing of some introns (Tanaka et
Most of these sequences have been found in the exon downstream of the stimulated introns and contain predominantly adenosine and guanosine residues. They have thus been defined as purine-rich exonic splicing enhancers. Most splicing enhancers, which are required for accurate splice site recognition and the control of alternative splicing, are located close to the splice site and have been identified in both regulated and constitutively spliced pre-mRNAs. When located in the downstream exon, these exonic enhancers can activate splicing of the upstream intron by promoting the use of a weak 3' splice site. Biochemical evidence indicates that these elements function as binding sites for SR proteins (Fu, 1995; Manley and Tacke, 1996; see Section 1.2.6.2 below).

1.2.6 Protein Factors Involved in RNA Splicing

1.2.6.1 mBBP/SF1 and U2AF

Formation of the catalytically active site of RNA splicing requires not only U snRNPs but also several other protein components. Mammalian orthologs of the yeast branchpoint sequence binding protein (mBBP/SF1) and U2AF are required for recognition of 3' splice sites (Zamore and Green, 1989; Berglund et al., 1997). The mBBP/SF1 binds specifically to the BPS, interacts with U2AF65, and promotes cooperative binding to a branchpoint sequence-polypyrinidene tract-containing RNA, which may contribute to initial recognition of the BPS during pre-mRNA splicing (Berglund et al., 1998).

Base pairing between U2 snRNA and the branchpoint sequence (BPS) is essential for pre-mRNA splicing. Because the mammalian BPS is short and highly degenerate, this interaction alone is insufficient for specific binding of U2 snRNP. The splicing factor U2AF binds to the polypyrinidene tract at the 3' splice site in the earliest
spliceosomal complex E, and is essential for U2 snRNP binding in the spliceosomal complex A (Ruskin et al., 1988). U2AF was first identified as an activity required for the stable association of U2 snRNP with the branchpoint sequence in the formation of the A complex (Ruskin et al., 1988). Human U2AF is a heterodimer composed of a 65 kDa large subunit and a 35 kDa small subunit (Zamore et al., 1992). The large subunit, U2AF$^{65}$, is an essential splicing factor containing a C-terminal RS domain and three N-terminal RNA recognition motif domains. U2AF$^{65}$ binds specifically to polypyrimidine tracts and promotes annealing of U2 snRNA to the branch site (Valcarcel et al., 1996), while U2AF$^{35}$ is associated with the large subunit through protein-protein interactions (Zamore and Green, 1991; Zamore et al., 1992). Recently, a human 56-kD U2AF$^{65}$ associated protein (UAP56) has been identified to interact with U2AF$^{65}$, and is required for the U2 snRNP-branchpoint interaction (Fleckner et al., 1997). Analysis of the sequence of UAP56 indicates that it is a member of the DEAD box family of RNA-dependent ATPases. Thus, U2AF$^{65}$ also has the function of recruiting this essential splicing factor to the pre-mRNA, which mediates ATP hydrolysis during pre-mRNA splicing. The role of U2AF$^{35}$ in constitutive splicing had eluded detection for the past ten years until a new exciting finding has been recently made by three groups (Zorio and Blumenthal, 1999; Wu et al., 1999; Merendino et al., 1999). U2AF$^{35}$ functions in the recognition of the essential AG dinucleotide at the 3' splice site early in spliceosome assembly (See Fig. 8 and 9). The discovery that U2AF$^{35}$ recognizes the AG dinucleotide at the earliest stage of spliceosome assembly adds another important factor in initial recognition of the 3' splice site.

1.2.6.2 SR Proteins
SR proteins are members of a family of essential splicing factors, each of which contains one or more RNA recognition motifs (RRM) at the N-terminus and extensive serine/arginine dipeptides (RS domain) at the C-terminus (Ge and Manley, 1991; Krainer et al., 1991; Zahler et al., 1992). This group of proteins includes ASF/SF2 (alternative splicing factor/splicing factor 2) (Ge and Manley, 1990), SC35 (35-kD spliceosomal component) (Fu and Maniatis, 1990), and four additional polypeptides of 20, 40, 55 and 75 kDa (Zahler et al., 1992). In addition, some snRNP-associated (e.g. U1 70K) and non-snRNP associated (e.g. U2AF) splicing factors also belong to the superfamily of RS domain proteins (Fu, 1995). Many of these SR proteins are important in both constitutive and alternative RNA splicing (Fu, 1995). SR proteins have been shown to bind to pre-mRNA and promote the assembly of the E complex (Staknis and Reed, 1994a). SR protein AFS/SF2 can recognize 5' splice sites (Zuo and Manley, 1994) and facilitate the binding of U1 snRNP to 5' splice site by promoting the interactions between the RS domains of the protein and RS domain of the specific 70 kDa protein present on the U1 snRNP (Kohtz et al., 1994). SR proteins bond to the exonic enhancers can also promote the interaction between U2 snRNP and the upstream branch site (Lavigueur et al., 1993), which is mediated by facilitating U2AF binding to the adjacent polypyrimidine tract (Wang et al., 1995). SR proteins also mediate specific interactions between U1 and U2 snRNPs (Fu and Maniatis, 1992a). The observations that SR proteins can interact with both the U2AF\textsuperscript{35} and U1-70 K protein (Wu and Maniatis, 1993) suggest that SR proteins may function as a bridge between components bound to the 5' and 3' splice sites. SR proteins also are required to act late in the splicing reaction by recruiting U4/U6-U5 tri-snRNP complex to the spliceosome and participating in the transition of complex A into complex B (Roscigno and Garcia-Blanco, 1995). \textit{In vitro} complementation studies revealed that high
concentrations of SR proteins could substitute for the functions of U1 snRNP or the essential splicing factor U2AF (Crispino et al., 1994; Tarn and Steitz, 1994; MacMillan et al., 1997). These results indicate that SR proteins can also help to define the 5' splice site through an U1 snRNP-independent mechanism. Under these conditions, recognition of the 5' splice site by U6 snRNP could be facilitated by SR proteins in U1 snRNP-depleted extracts (Crispino and Sharp, 1994). Recent evidence has also shown that SR proteins have not only exon-dependent but also exon-independent functions in pre-mRNA splicing (Hertel and Maniatis, 1999).

In brief, SR proteins can interact with the 5' splice site, U1 snRNP, splicing enhancer sequences, and U2AF35. Therefore, it has been supposed that these proteins provide the bridge between the 5' and 3' splice sites, both spanning the intron to pair the 5' and 3' splice sites and across an exon to enhance splice-site recognition through a network of RS-domain-mediated protein-protein interactions.

1.3 Rationale

As reviewed above, mutations in the LCAT gene have been demonstrated to underlie either familiar LCAT deficiency (FLD) or fish-eye disease (FED). Previous mutations associated with LCAT deficiency syndromes have been identified in the coding region of the LCAT gene. However, our laboratory has recently described a point mutation (IVS4:T→C22) in intron 4 of the LCAT gene in patients with FED, which interrupts the splicing of LCAT pre-mRNA.

Splicing of pre-mRNA involves the recognition of the three short consensus sequences, the 5' splice site, the BPS, and 3' splice site. The functional significance of the splice sites has been well established by the studies of a number of naturally occurring mutations and of site-directed mutagenesis at these exon/intron junctions.
Branch site mutations have been neglected as a cause of human genetic disease, and little is known about the mechanisms of branchpoint selection. However, growing evidence indicates that the branchpoint sequence, although not strongly conserved in humans, can be of great importance for accurate and efficient splicing of human nuclear pre-mRNA, and that mutations in the branchpoint region can give rise to an abnormal RNA splicing pattern that affects gene expression and causes disease.

Since the point mutation (T→C) in intron 4 of the LCAT gene that causes FED resides in the 100% matched branchpoint consensus sequence (CCCTGAC vs. YNYTRAY), I became very interested in exploring why some mutations in the branchpoint consensus sequence can significantly affect the efficiency of splicing and cause disease while others do not. In this study, therefore, I have attempted to investigate the possible mechanisms responsible for the defective intron splicing and to establish further the functional significance of the BPS by using LCAT minigene constructs that contain branch site mutations as a model to study the effects of these mutations on the efficiency of pre-mRNA splicing. The elucidation of the molecular basis responsible for the defective splicing will help us to better understand the selection of the branchpoint adenosine during pre-mRNA splicing and mechanisms of human genetic disease caused by the splicing defect.

1.4 The Specific Aims

1. To establish stably transfected baby hamster kidney (BHK) cell lines that express LCAT minigenes containing the T→G and T→A mutations at the same site of intron 4 in which the natural mutation occurs.

2. To determine the effects of these intron mutations on the expression of the LCAT gene in order to support the idea that the specific loss of a uracil
residue, two bases upstream of the branchpoint adenosine in the branchpoint consensus sequence has a significant effect on the splicing of LCAT pre-mRNA.

(3) To create a series of mutations in the branchpoint consensus sequence, including an insertion of a wild-type branch site into the intronic sequence of the natural and the branchpoint mutants through polymerase chain reaction (PCR) based site-directed mutagenesis.

(4) To express these intron mutants by transient transfection into human embryonic kidney (HEK)-293 cells.

(5) To investigate the possible mechanism(s) responsible for the defective splicing and to establish the functional significance of each intron mutant in the branchpoint consensus sequence by the analysis of pre-mRNA splicing using RT-PCR as well as LCAT activity assays.
CHAPTER 2

T→G or T→A Mutation Introduced into Natural Mutation Site in the Branchpoint Consensus Sequence of Intron 4 of Lecithin:cholesterol Acyltransferase (LCAT) Gene: intron retention causing LCAT deficiency
2.1 RATIONALE

It is estimated that up to 15% of all point mutations causing human genetic disease result from mRNA splicing defects, most of which involve a single nucleotide substitution at the splice sites (Krawczak et al., 1992). However, more mutations in the branchpoint consensus sequence of nuclear pre-mRNA have been reported to cause human disease (see Section 1.2.1). As described above, the identification of an unique genetic defect in the LCAT gene underlying three patients with FED provides such an unequivocal evidence that a point mutation in a lariat branchpoint sequence can lead to a null allele as the result of intron retention and therefore cause the phenotype of a human disease (Kuivenhoven et al., 1996).

Although the branchpoint consensus sequence is poorly conserved in mammals, it plays a central role in nuclear pre-mRNA splicing. To study further the functional significance of intron branchpoint sequence affecting the structure and function of LCAT as well as the process of the LCAT pre-mRNA splicing, two new mutations (T→G and T→A) were introduced into the same site of the natural mutation (IVS4:T→22C) in intron 4 of the LCAT gene followed by in vitro expression of the newly constructed LCAT minigenes in baby hamster kidney (BHK) cells. Thus, the objective of this experiment is to demonstrate that the thymine residue, two bases upstream of the branchpoint adenosine residue in the putative branchpoint consensus sequence in intron 4 is of great importance for proper splicing of the LCAT pre-mRNA.
2.2 MATERIALS AND METHODS

2.2.1 Materials

A full length LCAT cDNA contained within pUC19 was kindly supplied by John McLean, Genentech, Inc., San Francisco. The pNUT vector (Palmiter et al., 1987; Funk et al., 1990) and BHK cells were kindly provided by Dr. Ross MacGillivray, Department of Biochemistry, UBC. Restriction and modification enzymes for the manipulation of DNA sequences were obtained from Pharmacia Biotechnology, New England BioLab, and Life Technologies (GIBCO BRL). Geneclean and MERmaid kits (Bio 101, La Jolla, Ca.) were used for the purification of double stranded DNA fragments. Radiolabeled products such as \[^3H\] cholesterol were from Amershram. Vent\textsubscript{R} DNA polymerase was purchased from New England BioLab, and Moloney murine leukemia virus (M-MLV) reverse transcriptase and TRI\textsubscript{ZOL} Reagent from Life Technologies, and ULTRA THERM\textsuperscript{TM} Thermophilic DNA Polymerase from Bio/Can Scientific.

Oligonucleotides were synthesized from Life Technologies. All tissue culture reagents including fetal bovine serum (FBS), Dulbecco's Modified Eagle Medium (DMEM) and serum-free medium (Opti-MEM) were supplied by Life Technologies. Methotrexate (Cyanamid Canada, Inc., Montreal, Que.) for selection medium was obtained from the University Hospital Pharmacy.

Polyclonal goat anti-human LCAT antibodies were kindly provided by Dr. Andras Lacko, Texas College of Osteopathic Medicine, University of North Texas. Fort Worth. Protein G used for the detection of LCAT protein was purchased from Pharmacia. Electrophoresis grade reagents and protein G conjugated to horseradish peroxidase (Protein G-HRP) were obtained from BioRad. All other chemicals were of reagent grade or better and were obtained from either Sigma or BDH.
2.2.2 Molecular Biology Methods

2.2.2.1 Growth and Transformation of *E. coli*

*E. coli* strain DH5-α were maintained in LB (10 g/L bacto-tryptone, 5 g/L bacto-yeast extract, and 10 g/L NaCl). Frozen bacterial stocks were prepared in 15% glycerol, frozen in liquid nitrogen and stored at -70°C.

Competent *E. coli* were prepared by inoculating a single colony into a 20 ml TYM medium (20 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 5.84 g/L NaCl, 1.20 g/L MgSO₄). After the growth of the cells had achieved midlog phase (OD₆₀₀=0.2-0.8), they were added to a 100 ml of TYM. After an OD₆₀₀ of 0.5-0.9 was reached, the culture was diluted with TYM to 500 ml. When an OD₆₀₀ of 0.6 was reached, the culture was rapidly cooled on ice, pelleted at 5,000 x g for 15 min, and resuspended on ice in 100 ml of cold 30 mM K-Acetate, 50 mM MnCl₂, 100 mM KCl, 10 mM CaCl₂ and 15% glycerol. The mixture was pelleted at 5,000 x g for 10 min and resuspended in 20 ml of cold 10 mM Na-MOPS, 75 mM CaCl₂, 10 mM KCl and 15% glycerol. Aliquots of 0.1 and 0.5 ml were frozen in liquid nitrogen and stored at -70°C.

Transformation was accomplished by gently mixing competent cells with plasmid DNA in an ice bath for 30 min. The mixture was heat shocked for 5 min at 37°C and then diluted 1:10 in LB and incubated with gentle agitation at 37°C for 90 min. The mixtures were plated onto LB-agar plates containing 100 μg/ml ampicillin and incubated inverted overnight (16 h) at 37°C.

2.2.2.2 Purification of DNA

2.2.2.2.1 Small Scale Plasmid Preparation
Aliquots of 5.0 ml of LB broth containing ampicillin (100 μg/ml) were inoculated with a single bacterial colony and were incubated at 37°C overnight with vigorous agitation. Bacteria were collected by microcentrifugation and resuspended in 100 μl of 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 50 mM sucrose. The cells were then lysed in 200 μl of 0.2 M NaOH, 1% SDS, neutralized with 150 μl of 5 M potassium acetate, pH 4.8, and mixed gently by inverting 3 times. Cellular debris was removed by microcentrifugation and the supernatant was extracted with an equal volume of 1:1 phenol:chloroform mixture. The upper aqueous layer was collected after the phases were separated by centrifugation. The plasmid DNA was precipitated by the addition of 2 volumes of 100% ethanol. Following microcentrifugation, the pellet was washed with 70% ethanol to remove salts, and the pellet was finally air-dried and dissolved in 50 μl of sterile water or TE (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA).

2.2.2.2 Large Scale Plasmid Preparation

To obtain larger quantities of plasmid DNA suitable for eucaryotic cell transfection a 5 ml overnight culture was used to inoculate 250 ml of LB broth containing 100 μg/ml of ampicillin. After overnight incubation with vigorous shaking at 37°C, cells were pelleted at 3,000 x g and the pellets were resuspended on ice in 5 ml of 25 mM Tris-HCl pH 8.0, 50 mM glucose, and 10 mM EDTA. After 20 min, the mixture was transferred to a sterile 30 ml polypropylene screw cap tube, mixed with 10 ml of 0.2 M NaOH, 1% SDS by gentle inversion, and incubated for 10 min on ice. The lysate was then treated with 7.5 ml of 5 M potassium acetate, pH 4.8. After the mixtures were centrifuged at 25,000 x g for 30 min, the supernatant was transferred to a 50 ml Falcon polypropylene screw cap tube, digested with 50 μg of pancreatic ribonuclease A (RNase
A) for 20 min at 37°C. The plasmid DNA was precipitated with 2 volumes of ice-cold 95% ethanol at -20°C for 20 min. The DNA was collected by centrifugation at 3,000 rpm for 15 min and the pellet washed 2x with 70% ethanol before it was resuspended in 2 ml of sterile water. The mixture was transferred to a 15 ml conical Falcon tube and extracted twice with equal volumes of phenol:chloroform (1:1) and once with chloroform. 1.6 ml of sample was transferred to a chilled 15 ml silanized Corex tube containing 0.4 ml of 5 M NaCl and 2.0 ml of 13% PEG (8000). After this mixture was incubated on ice for at least one hour and centrifuged at 12,000 rpm for 15 min, the pellet was washed with 70% ethanol, air-dried and dissolved in 200-500 µl of TE. The concentration of plasmid was quantified by UV absorbance at 260 nm.

2.2.2.3 Oligonucleotide-directed Mutagenesis

The natural mutation in intron 4 of the LCAT gene is a substitution of thymidine for cytosine, 22 bases upstream of the 3' splice site. To introduce new mutations at the exact same position, the pNUT-LCAT minigene that contains a full-length LCAT cDNA and the wild type sequence of intron 4 (Kuivenhoven et al., 1996) as well was used as template in a megaprimer method of mutagenic polymerase chain reaction (Sarkar and Sommer, 1990). The T→G and the T→A mutations were created by use of the following respective oligonucleotides: 5'- CAG CTG CCC GGA CCC CTT CC -3' and 5'- CAG CTG CCC AGA CCC CTT CC -3'. These sequences were derived from the LCAT sequence as published by McLean et al. (McLean et al., 1986a). In detail, the first round of PCR was performed with 10 pg of template DNA, 200 µM of each dNTP's, 0.4 µM of the mutagenic oligonucleotide as a forward primer, and 0.4 µM of the reverse universal primer (5'- CGG ACA GAA GGA GTA ACC GG -3') which hybridizes to the pNUT vector.
downstream from the LCAT minigene sequence. The PCR was performed as follows: 30 cycles of 30 sec 95°C / 30 sec 65°C / 30 sec 72°C with 1 unit of proofreading Vent DNA polymerase. After the thirtieth cycle, the reaction was extended for 5 min at 72°C. The PCR product was gel-purified and used as a megaprimer in a subsequent second round of PCR reactions. The forward universal primer (5'- CCT GCA CAC ACT GGT GC -3') is complementary to the pNUT vector sequence upstream of the LCAT minigene. To reduce non-specific products, a touchdown PCR protocol was used (Don et al., 1991).

2.2.2.4 Isolation of DNA Fragments

PCR products and/or DNA fragments from the restriction enzyme digestion of plasmids were purified by agarose gel electrophoresis. Specific bands were excised from the gel and DNA was recovered by following the protocols for Geneclean and MERmaid (Bio101). The quantity of the isolated DNA was determined by ethidium bromide staining of the agarose gels and comparison to standards.

2.2.2.5 Construction of Expression Vector

The mammalian expression vector pNUT (Palmiter et al., 1987; Funk et al., 1990) used in the studies contains a mouse metallothionein 1 promoter, pUC 18 sequences to allow replication and selection in E.coli, and a mutant form of the dihydrofolate reductase (DHFR) cDNA that permits selection of cells containing the plasmid DNA by their survival in very high concentration of methotrexate (Funk et al., 1990). Human LCAT cDNA was first released from the pUC 19 vector and then subcloned into the mammalian expression vector pNUT (Hill et al., 1993b). Both the sequences of wild type intron 4 (IVS4-WT) of the LCAT gene and the fourth intron with
the point mutation (IVS4:T-22C) were derived from the genomic DNA of a compound heterozygote by PCR. Using unique restriction sites, these intronic sequences were inserted into pNUT-LCAT cDNA (Kuivenhoven et al., 1996). For the generation of the two novel mutations, the final PCR product from above was purified and digested with BssH II and Sca I to generate 132 bp DNA fragments that contain the desired intronic mutations. These enzymes recognized unique restriction sites at the 3' end and 5' end of exons 4 and 5, respectively. The resulting 132 bp fragment which contained the mutagenized intron 4 was inserted into the wild type LCAT cDNA in pNUT vector as a cassette by the use of the same restriction sites. All final constructs were sequenced to confirm the presence of introduced intron mutations.

2.2.2.6 Isolation of Total RNA from the Transfected Cells

Total RNA was isolated from the stably transfected BHK cells using the guanidinium isothiocyanate method (Chomczynski, 1993). The cells in a 3.5-cm diameter culture dish were lysed directly by adding 1 ml of TRIZOL Reagent. After the homogenized samples were incubated for 5 min at room temperature, 0.2 ml of chloroform was added to Eppendorf tube. The tubes were vigorously shaken by hand for 15 sec and incubated for another 3 min. The samples were centrifuged at 12,000 x g for 15 min at 4°C to separate the phases, and the colorless upper aqueous phase was transferred to a fresh tube. The RNA was precipitated from the solution by mixing with 0.5 ml of isopropyl alcohol. The samples were then incubated at room temperature for 10 min and centrifuged at 12,000 x g for 10 min at 4°C. After the supernatant was removed, the RNA pellet was washed with 75% ethanol, air-dried, and dissolved in DEPC-treated H2O.
2.2.2.7 Reverse Transcription and PCR Amplification (RT-PCR)

The first-strand cDNA was synthesized from 1-5 µg of total RNA in a 20 µl reverse transcription mixture containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 200 µM dNTP's and 200 ng Oligo (dT)₁₂₋₁₈. The tubes were heated at 75°C for 3 min and cooled immediately on ice. M-MLV reverse transcriptase (200 units) was added. The tubes were incubated at 42°C for 1 h and subsequently heated at 95°C for 3 min. After reverse transcription, 10% of the reaction product was used in a PCR amplification reaction. The 50 µl PCR reaction mixture had the following final composition: 75 mM Tris-HCl (pH 8.8), 20 mM (NH₄)₂SO₄, 15 mM MgCl₂, 0.01% Tween-20, 200 µM dNTP's, 0.4 µM of LCAT primers that were derived from sequences of exon 3 (sense strand: 5'- CCT GCA CAC ACT GGT GCA GA -3') and exon 5 (antisense strand: 5'- GGC CAA TGA GGA AGA CAG GC -3'), respectively, and 1 unit of ULTRA THERM™ Thermophilic DNA Polymerase. The samples were heated at 95°C for 3 min followed by 35 cycles of 30 sec, 95°C / 30 sec, 60°C / 1 min, 72°C in a DNA Thermal Cycler (Perkin-Elmer 2400). The PCR products were separated on a 2% agarose gel, and the bands were visualized by staining with ethidium bromide.

2.2.3 Eucaryotic Cell Culture

BHK cells were maintained in DMEM containing 10% FBS and 1% antibiotic. FBS was heat-inactivated at 56°C for 30 min prior to use as a growth supplement. Cell cultures were maintained in a humidified incubator at 37°C with a 5% CO₂ atmosphere.

2.2.3.1 Stable Transfection of BHK Cells
The pNUT LCAT minigene constructs, i.e., pNUT-LCAT IVS4-WT (wild type), pNUT-LCAT IVS4-MUT natural (T→C), pNUT-LCAT IVS4-MUT-1 (T→G), and pNUT-LCAT IVS4-MUT-2 (T→A) as well as pNUT-LCAT cDNA were used to establish stable cell lines of BHK cells. Stable transfection of BHK cells was performed by calcium phosphate co-precipitation as previously described (Rosenthal, 1987). BHK cells were maintained in DMEM supplemented with 10% heat-inactivated FBS. 10 μg of the plasmid DNA in 450 μl of 0.25 M CaCl₂ was mixed with 2 X HEPES-buffered saline containing 40 mM HEPES, pH 6.96, 280 mM NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄ and 10 mM glucose. The DNA-calcium phosphate mixture was incubated at room temperature for 30 min before it was added dropwise to a 100-mm culture dish containing approximately 50% confluent BHK monolayers. After an overnight incubation at 37°C under 5% CO₂, the transfection medium was replaced with DMEM/10% FBS for 24 hours. The transfected cells were then selected over a period of 10-14 days in DMEM/10% FBS containing 500 μM methotrexate. Surviving colonies were transferred to 20-mm culture wells and grown to confluence under selective conditions. Colonies that express maximal quantities of LCAT and/or contain the integrated LCAT minigenes were identified by the LCAT activity assay as well as RT-PCR.

2.2.4 Protein Analysis

2.2.4.1 Western Blot

Western blotting was performed as previously described (Hill et al., 1993b). Samples from the culture medium were mixed with 2 X SDS sample buffer containing 10% β-mercaptoethanol and 0.1% bromphenol blue at a ratio of 1:1. The mixture was boiled for 5 min and run on 7.5% SDS-PAGE gel at a constant current of 15 mA/gel for
1-2 h. Gels were electroblotted onto nitrocellulose paper (pore size, 0.45 μM) as described by Towbin et al. (Towbin et al., 1979). Transfer was achieved at 100 V for 1 h in a cooled chamber containing transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol). The membranes were blocked with 5% non-fat powdered milk in 1 x PBS (30 min) and incubated with polyclonal goat anti-human LCAT antibodies and Protein G-HRP to detect the LCAT protein.

2.2.4.2 Measurement of LCAT Activity

To determine the presence of LCAT in the medium, appropriate BHK cell lines were incubated in DMEM containing 10% FBS for 12 h, washed three times in serum-free medium and incubated in Opti-MEM for 48 h. Culture medium was then collected for the measurement of LCAT activity.

LCAT activities in the collected media were determined using an HDL-like proteoliposome substrate containing [3H] cholesterol, phosphatidylcholine and apo A-I (O et al., 1993a; Batzri and Korn, 1973). Egg yolk phosphatidylcholine:cholesterol liposomes were prepared by ethanol injection according to Batzri and Korn (Batzri and Korn, 1973). The substrate mixture containing 4.66 nmol [3H] cholesterol (0.03 μCi/nmol), 18.46 nmol phosphatidylcholine, and 7.5 μg purified human apo A-I in 10 mM Tris-HCI (pH 7.4)-150 mM NaCl-5 mM EDTA was preincubated at 37°C for 30 min. Subsequently, 5 mM β-mercaptoethanol and 1.5% bovine serum albumin (essentially fatty acid free) were added to the substrate mixture. After the addition of 50 μl of BHK cell culture medium, the mixture was incubated at 37°C for 1 h and was terminated by the addition of 1 ml ethanol. Unesterified cholesterol and cholesteryl esters were separated by thin-layer chromatography on silica gel layers incubated in petroleum
ether:diethyl ether:acetic acid (70:12:1). The radioactivity associated with cholesterol and cholesteryl ester was determined by liquid scintillation counting.

2.2.4.3 Quantification of LCAT Protein

LCAT protein was quantified by a solid-phase LCAT immunoassay using a polyclonal goat anti-human LCAT antibody and Protein G-HRP. The immunoassay of LCAT was carried out by using nitrocellulose membranes (0.45 μM pore size) as a solid-phase support. Samples containing either culture medium or purified human recombinant LCAT were bound to the membrane in a BioRad slot blot apparatus. The membrane was then blocked with 5% non-fat dry milk in PBS for 30 min at 37°C. Polyclonal goat anti-human LCAT antibodies were incubated with the membrane overnight at room temperature in PBS containing 0.5% non-fat dry milk. The membrane was then washed 3 X for 5 min with PBS containing 0.02% Tween before it was reacted with Protein G-HRP for 1 hr. The nitrocellulose was again washed 3 times and then developed in 50 ml of PBS containing 25 mg diaminobenzidine, 15 mg CoCl₂ and 0.010 ml of 30% H₂O₂ to visualize the protein. The blot was scanned using LKB Laser Densitometer. The results were integrated using Gelscan XL software (Pharmacia Biotech).
2.3. RESULTS

2.3.1. DNA Sequence Analysis of the LCAT Minigene Constructs

The natural mutation of intron 4 of the LCAT gene in a family with FED occurs in a putative branchpoint consensus sequence (CCCTGAC) (Kuivenhoven et al., 1996) and is a T→C nucleotide substitution, 22 bases upstream of the 3' splice site. To further study the functional significance of this specific mutation in interfering with LCAT pre-mRNA processing, the LCAT IVS4-WT, LCAT IVS4-MUT natural and two minigene constructs with other base changes at the same position, i.e., LCAT IVS4-MUT-1 (T→G) and LCAT IVS4-MUT-2 (T→A) were generated. The LCAT minigenes were cloned into the mammalian expression vector pNUT (Fig.10) and subsequently sequenced to check for the correct site-directed mutagenesis (Fig.11). Except for the desired mutations, no other mutation was found in the intron 4 sequence.
Figure 10. Schematic representation of the pNUT expression vector containing LCAT minigenes. The minigenes encompass of full length wild type LCAT cDNA and mutated forms of intron 4, i.e., T→C (pNUT-LCAT IVS4-MUT natural); T→G (pNUT-LCAT IVS4-MUT-1) and T→A (pNUT-LCAT IVS4-MUT-2) intron mutations as indicated by boxes. MT-1 pro: mouse metallothionein 1 promoter; hGH 3': 3' portion of the human growth hormone; SV40: simian virus 40 promoter and origin; DHFR: dihydrofolate reductase; HBV 3': hepatitis B virus 3' end; and AmpR: ampicillin resistance gene.
Figure 11. Fluorescence-based sequencing results of subsequent LCAT minigene constructs as produced by PCR mutagenesis. Panel (a) shows the putative branchpoint consensus sequence of the wild type intron 4 of the LCAT gene; (b) indicates the natural mutation T→C that was identified in a family with FED; (c) represents the introduced T→G base change (LCAT IVS4-MUT-1); and (d) demonstrates a T→A nucleotide substitution (LCAT IVS4-MUT-2).

2.3.2. Expression of LCAT Minigenes in BHK Cells

After the pNUT-LCAT minigenes were stably transfected into BHK cells, LCAT activity was determined in the culture medium using a HDL-like proteoliposome substrate (Hill et al., 1993b; Batzri and Korn, 1973). As shown in Table 3, the uninterrupted LCAT cDNA exhibited the highest activity. As previously demonstrated (Kuivenhoven et al., 1996), the insertion of the wild type intron 4 in the LCAT cDNA resulted in an about 90% decrease of LCAT activity in the culture medium as compared
to the full length of LCAT cDNA. However, the introduction of LCAT IVS4-MUT-1 and LCAT IVS4-MUT-2 into the putative branchpoint sequence in intron 4 completely abolished the secretion of LCAT in the culture medium as previously observed for the natural mutation.

To quantify the LCAT protein in the culture medium, the mass of the secreted LCAT was assayed using a polyclonal antibody specific for human LCAT. No LCAT protein could be detected in both the medium of pNUT-LCAT IVS4-MUT-1 and pNUT-LCAT IVS4-MUT-2 cell lines as well as the cells that contained the minigene with the natural mutation. Specific LCAT activities as determined in the medium of cells transfected with either wild type LCAT cDNA or LCAT intron 4 wild type did not differ.

Table 3. LCAT activity and concentration secreted by stably transfected BHK cells

<table>
<thead>
<tr>
<th></th>
<th>(n=3)</th>
<th>pNUT empty vector</th>
<th>pNUT-LCAT cDNA</th>
<th>PNUT-LCAT IVS4-WT</th>
<th>PNUT-LCAT IVS4-MUT natural</th>
<th>PNUT-LCAT IVS4-MUT-1</th>
<th>PNUT-LCAT IVS4-MUT-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCAT activity</td>
<td>0.07 ± 0.02''</td>
<td>20.69 ± 6.24''</td>
<td>2.36 ± 0.59</td>
<td>0.07 ± 0.05''</td>
<td>0.07 ± 0.03''</td>
<td>0.06 ± 0.03''</td>
<td></td>
</tr>
<tr>
<td>Concentration</td>
<td>n.d.</td>
<td>9.28 ± 3.36''</td>
<td>1.05 ± 0.28</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td></td>
</tr>
<tr>
<td>Specific activity</td>
<td>2.34 ± 0.68</td>
<td>2.23 ±1.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Confluent BHK cells were incubated in serum-free Opti-MEM for 48 hours. The culture medium was assayed for LCAT activity and protein mass as described in the Materials and Methods. The data represent the result of three separate analyses, each carried out in duplicate. n.d. = not detected. ***p<0.0001 compared to the intron wild type.
2.3.3. Reverse Transcription and PCR Amplification

To evaluate whether the mutations in the putative branchpoint consensus sequence affect normal splicing of LCAT pre-mRNA, BHK cell lines that expressed the various constructs (i.e., pNUT-LCAT cDNA; pNUT-LCAT IVS4-WT; pNUT-LCAT IVS4-MUT natural; pNUT-LCAT IVS4-MUT-1; pNUT-LCAT-IVS4-MUT-2) were used to collect total RNA.

To eliminate the possible DNA contamination, the RNA samples were first treated with DNase I, and the isolated RNA was then subjected to RT-PCR and the products were run on a gel (see Fig.12). LCAT transcription of the LCAT minigenes was evident in all cell lines except for the one that was transfected with an empty vector (Fig.12). However, when comparing the cells that expressed the uninterrupted LCAT cDNA construct with those that expressed the minigenes, a difference in size of the RT-PCR products was observed. Specifically, the 83 bp difference indicated the retention of intron 4 in all minigenes. Subsequent sequence analysis of the transcripts demonstrated that the intron sequences contained the introduced mutations (data not shown). The cells expressing the LCAT IVS4-WT also contained correctly spliced LCAT transcripts (Fig.12, lane 3). The minor band of the correctly spliced mRNA in the LCAT IVS4 wild type as compared to the unspliced product may explain for the lower LCAT concentration and activity in the culture medium of these cells compared with the cells that expressed wild type LCAT cDNA.
Figure 12. RT-PCR on total RNA isolated from BHK cells that were stably transfected with pNUT-LCAT cDNA and pNUT LCAT minigenes containing intron 4 variants. Total RNA isolated from the cell lines were analyzed by RT-PCR using primers complementary to sequences of exon 3 and exon 5, respectively. Correct splicing results in the excision of intron 4 and generates the template for the amplification of a 270 bp fragment. The unspliced product has a size of 353 bp and contains the complete 83 bp sequence of intron 4 with the introduced mutations. Lane 1: pNUT vector without insert; lane 2: pNUT LCAT cDNA; lane 3: pNUT-LCAT IVS4-WT; lane 4: pNUT-LCAT IVS4-MUT-natural (T→C); lane 5: pNUT-LCAT IVS4-MUT-1 (T→G); lane 6: pNUT-LCAT IVS4- MUT-2 (T→A); and M: molecular weight marker.
2.3.4. Western Blotting

To test for the presence of LCAT protein in the culture medium, Western blotting was performed using polyclonal goat anti-human LCAT antibodies. As shown in the Fig. 13, BHK cells stably transfected with pNUT-LCAT cDNA and pNUT-LCAT IVS4-WT (lane 2 and 3) secreted a major protein with an apparent molecular weight of about 67 kD which is consistent with the expected size of the mature LCAT. No LCAT protein was detected in the culture medium of cells transfected with mutated minigenes. Since the sequence of intron 4 contains a premature stop codon, retention of this intron may therefore lead to the expression of a truncated LCAT of about 28 kD (when assuming full glycosylation). However, we did not detect any truncated forms of LCAT in the intron 4 mutants, suggesting that the truncated proteins were probably rapidly degraded within the cells.
Figure 13. LCAT Immuno-Western blot of culture medium from BHK cells transfected with LCAT cDNA and LCAT minigene constructs. The medium was collected and separated by 10% SDS-PAGE. LCAT protein was blotted onto nitrocellulose membranes and detected with a polyclonal goat antibody specific for human LCAT. lane 1: pNUT empty vector; lane 2: pNUT-LCAT cDNA; lane 3: pNUT-LCAT IVS4-WT; lane 4: pNUT-LCAT IVS4-MUT-natural (T→C); lane 5: pNUT-LCAT IVS4-MUT-1 (T→G); lane 6: pNUT-LCAT IVS4-MUT-2 (T→A); and M: molecular weight standard.
2.4 DISCUSSION

2.4.1 The Thymine Residue in the Putative Branchpoint Sequence Is Absolutely Required for Efficient Splicing of Intron 4 of the LCAT Gene

It has been well documented that mutations in the LCAT gene underlie either FLD or FED which are both inherited in an apparently autosomal recessive manner. As described above, our laboratory has previously identified an intron mutation of the LCAT gene (IVS4:T→22C) in patients with FED. *In vitro*, this mutation has been shown to result in the intron retention (Kuivenhoven *et al.*, 1996).

In the present study, we demonstrate that the thymine residue in the putative branchpoint consensus sequence is an absolute prerequisite for correct splicing of the LCAT pre-mRNA. LCAT minigenes containing two novel point mutations at the same position as the natural mutation in intron 4, i.e., T→G and T→A, were stably expressed in BHK cells. Subsequent analysis of the RNA of these stably transfected cells in a reverse transcriptase PCR experiment revealed that the loss of the T residue interferes with the splicing of pre-mRNA and results in the intron retention.

2.4.2 The Possible Mechanism for the Defective Splicing

The removal of introns from pre-mRNAs by the RNA splicing is an elementary step in eukaryotic gene expression. It has been demonstrated that nuclear pre-mRNA splicing involves two-transesterification steps (Green, 1986, 1991). The first step involves cleavage at the 5’ splice site and the formation of an intron lariat structure in which an adenosine residue, usually 18-40 nucleotides upstream of the 3’ splice site, plays an essential role (Green, 1986). The second step involves cleavage at the 3’ splice site and ligation of the two exon sequences. During the splicing reactions, the binding of the U2 snRNA to the BPS of pre-mRNA is suggested to enable the
branchpoint adenosine to be bulged out of the U2-intron duplex, which permits the
nucleophilic attack on the 5' splice site, leading to the formation of lariat structure
(Madhani and Guthrie, 1994, Query et al., 1996).

The natural intron 4 mutation of the LCAT gene is located in the putative
branchpoint consensus sequence (CCCTGAC), just two nucleotides upstream of the
putative branchpoint adenosine residue (Kuivenhoven et al., 1996) and substitutes a C
residue for a highly conserved T. I have mutated this very thymine residue with two
purine residues, i.e., both a guanosine and an adenosine. The results obtained from
these transfection experiments are consistent with the loss of the thymine residue
interfering with the binding of U2 snRNA to the branchpoint sequence. Subsequent
inappropriate bulging of the branchpoint A results in defective splicing, and thus intron
retention (Fig. 14).
Figure 14. Proposed possible mechanism for defective splicing. Putative interactions of U2 snRNA with the branchpoint sequence of wild type (a) and mutant (b, c, and d) intron 4 of LCAT pre-mRNAs. The bulging of the A, which is the result of the base-pairing of the branchpoint consensus sequence with U2 snRNA, is required for its usage in the branchpoint region. The loss of the T, just two bases upstream of the branchpoint A, therefore interferes with the binding of U2 snRNA to the branchpoint region. The inability of the bulging of the A disrupts the pre-mRNA splicing, thus results in the intron retention.
2.4.3 Retention of LCAT Intron 4 Probably Leading to a nonsense transcript that Is Most Likely Rapidly Degraded within the Cells

Although the efficiency of LCAT pre-mRNA splicing in BHK cells transfected with LCAT intron 4 wild type was not efficient when compared with the uninterrupted LCAT cDNA, the correctly spliced transcript could be clearly identified (Fig. 12, lane 3). By contrast, no band corresponding to the mature mRNA transcript was observed in the cell lines that expressed the LCAT intron 4 mutant minigenes (Fig. 12, lane 4-6).

Analysis of LCAT mRNA isolated from leukocytes of a compound heterozygote by RT-PCR also revealed a larger DNA fragment containing intron 4 with the point mutation as compared to the spliced RNA (Kuivenhoven et al., 1996). It seemed that the intron mutation that was harbored in the LCAT pre-mRNA also abolished the splicing of the intron because the smaller product (mature RNA transcript) exclusively contained the T123I defect. These observations suggested that the substitution of the thymine residue in the BPS resulted in the intron retention both in vitro and in vivo. Since the sequence of intron 4 of LCAT gene contains a premature stop codon signal, defective splicing of intron 4 may therefore result in the synthesis of a truncated protein with an expected size of 28 kDa (assuming full glycosylation). In this study, Western blotting was used to demonstrate the existence of LCAT proteins in the culture medium and to determine the molecular size of both the normal and the truncated LCAT proteins. However, I failed to detect any truncated LCAT in the medium from BHK cells which expressed mutated LCAT minigenes. The absence of LCAT activity and protein in the medium probably suggest that these cells are either unable secrete truncated LCAT protein or these translation products are rapidly degraded within the cells. However, the most likely situation is that the LCAT pre-mRNA with the premature termination codon in the intronic sequence is recognized and efficiently degraded in a pathway known as
nonsense-mediated mRNA decay (NMD) (Frischmeyer and Dietz, 1999; Mitchell and Tollervey, 2000) by mRNA surveillance mechanism (Hilleren and Paeker, 1999).

2.4.4 The Implications of the Findings

To date, most reports on mutations in putative branchpoint regions of human genes mention the use of alternative branchpoint sequences and consequently differential splicing (Krawczak et al., 1992; Rosenthal et al., 1992; Nakahashi et al., 1992; Brand et al., 1996). In this study, however, I have not observed alternative splicing of LCAT minigene messages containing the intron 4 mutations. This is probably due to the absence of alternative branchpoint sequences and splice sites in this short intron and downstream exonic sequences. These results indicate that the BPS, although not strongly conserved in mammals, can be of essential importance for correct splicing of human pre-mRNA. In this respect, one has to consider the unusually high conservation of the total LCAT gene sequence (Shaw, 1996) and the short length of intron 4 of the LCAT gene.

In summary, I provide further evidence that supports the notion that a naturally occurring mutation in intron 4 of the LCAT gene, which apparently seems far from the intron-exon boundary to affect RNA splicing, is the probable cause of the absence of a gene product from the LCAT allele that harbors it. The transfection experiments with the two artificially created mutants, \( T \rightarrow G \) and \( T \rightarrow A \), demonstrate that the loss specifically of T two bases upstream of the branchpoint adenosine in the branchpoint consensus sequence has a significant effect on RNA splicing, and thus results in intron retention. This study is significant since the results confirm that a single nucleotide substitution in the branchpoint consensus sequence of an intron can cause human disease. Such a mechanism may be a more widespread cause of phenotypes of genetic disease in
humans than had previously been appreciated. Therefore, we suggested that the search for mutations which cause human genetic disease should include consideration of regions other than the promoter region, coding sequences and intron-exon boundaries of the gene only.
CHAPTER 3

Characterization of the Effects of Mutations in the Branchpoint Sequence of Intron 4 on the Splicing within the Human Lecithin:cholesterol Acyltransferase (LCAT) Gene
3.1 RATIONALE

In the previous chapter, I demonstrate that the specific loss of the thymine residue two bases upstream of the branchpoint adenosine in intron 4 results in defective splicing without the synthesis of any detectable LCAT protein, suggesting that this highly conserved thymidine residue in the BPS is essential for the LCAT pre-mRNA splicing. To date, only a few mutations in the branchpoint consensus sequence of introns have been reported to cause human genetic diseases (See Section 1.2.1). However, the growing evidence as well as the results presented in my first experiment clearly indicate that the BPS can be of essential importance for accurate and efficient splicing of human nuclear pre-mRNA.

As discussed above, a series of elegant genetic experiments in both yeast and mammalian cell lines has shown that base pairs exist between the branchpoint sequence and a conserved region (5'-AUGAUG-3') in U2 snRNA (Parker et al., 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989). The BPS is absolutely conserved in yeast (UACUAAC, the underlined A indicates the branchpoint nucleotide). Mutations of the yeast branchpoint adenosine significantly reduce or abolish splicing (Langford et al., 1984). By contrast, the BPS exhibits only a weak consensus in mammals (YNYURAY, Y=pyrimidine, R=purine, and N=any nucleotide), but sequences most closely resembling UACUAAC are preferentially used (Zhuang et al., 1989). Mutations of the branchpoint adenine in mammalian introns are known to only result in moderately reduced splicing efficiency due to the utilization of cryptic BPS (Padgett et al., 1985; Reed and Maniatis, 1985; Ruskin et al., 1985). To compensate for the poorly conserved branchpoint sequence, most mammalian introns contain another cis-acting element, the Py tract, between the BPS and the 3' splice site. The specific binding of U2AF65 to the polypyrimidine tract has been shown to promote the base pairing of U2 snRNA with the
branchpoint sequence early during spliceosome assembly (Valcarcel et al., 1996). However, much less is known about the mechanism of the selection of the branchpoint adenosine, which is usually located 18-40 nucleotides upstream of the 3' splice site, in the mammalian intron with respect to the base pairs between the U2 snRNA and the pre-mRNA.

The LCAT gene has an unusually high conserved sequence (Shaw, 1997; Hixson et al., 1993). The intron 4 of LCAT gene contains only 83 base pairs and the possible absence of cryptic BPS in this short intron may render it highly sensitive to any alteration within the branchpoint region. Therefore, the branchpoint sequence in intron 4 of the LCAT gene may act as a model to identify the functional significance of conserved nucleotides contained within the BPS of the human genes.

In this chapter, I extended the studies into the whole region of the branchpoint consensus sequence. I generated a series of mutations in the BPS and expressed these intron mutants in transiently transfected HEK-293 cells followed by the analysis of pre-mRNA splicing using RT-PCR as well as the measurement of LCAT activity. Furthermore, I also performed two rescue experiments by inserting a normal branch site into the intronic sequence of the natural or branchpoint mutant and by changing one nucleotide (G→A) in the naturally mutated BPS (CCCCGAC). The objective of this study is therefore to establish the functional significance of each nucleotide within the conserved branchpoint sequence of intron 4 for the efficient splicing of human LCAT mRNA precursors and to investigate the possible mechanisms responsible for the defective splicing.
3.2 MATERIALS AND METHODS

3.2.1 Materials

Top10 One Shot™ chemically competent cells and the pcDNA3.1(-) mammalian expression vector were obtained from Invitrogen. QIAprep Spin Plasmid Kit, QIAfilter Maxi Plasmid Kit, QIAquick Gel Extraction Kit, and RNeasy Mini Kit were purchased from QIAGEN. Pfu DNA polymerase was obtained from Stratagene, Taq DNA polymerase from Boehringer Mannheim, and DNase I from Promega. The human embryonic kidney (HEK) 293 cell line was obtained from the American Type Culture Collection. Geneticin (G418) for selection medium was obtained from Life Technologies. All other restriction and modification enzymes, primers, and reagents were obtained from the same source as those described in Chapter 2, section 1.

3.2.2 Molecular Biology Methods

3.2.2.1 Transformation of E.coli

Transformation was carried out by gently mixing One Shot™ TOP10 chemically competent cells with plasmid DNA in an ice bath. After 30 min, the reaction mixture was heat shocked at 42°C for exactly 30 sec and then diluted 1:5 in SOC (20 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 0.5 g/L NaCl, and 20 mM glucose) medium and incubated at 37°C for 1 h at 225 rpm in a shaking incubator. Mixtures were plated onto LB (10 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 10 g/L NaCl) - agar plates containing 100 μg/ml ampicillin and incubated inverted overnight (16 h).

3.2.2.2 Purification of DNA

3.2.2.2.1 Small Scale Plasmid Preparation
Aliquots of 5 ml of LB broth containing ampicillin (100 μl/ml) were inoculated with a single bacterial colony and were incubated at 37°C for 16 h with vigorous agitation. The small scale plasmid preparation was carried out by using the QIAprep Spin Plasmid Kit according to the standard protocol provided by the manufacturer. Following centrifugation at 3,000 rpm for 10 min, the bacterial pellets were resuspended in 250 μl of buffer P1 (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 μg/ml RNase A) and then lysed in 250 μl of buffer P2 (200 mM NaOH, 1% SDS). After the addition of 350 μl of buffer P3 (3 M potassium acetate, pH 5.5), the neutralized mixtures were centrifuged for 10 min at 13,000 rpm in the tabletop microcentrifuge (Eppendorf, centrifuge 5417C). The supernatants were applied to QIAprep columns and centrifuged for 60 sec. The QIAprep spin columns were then washed, and DNA was eluted by the addition of 50 μl of H2O to the center of the column followed by centrifugation for 1 min at maximum speed.

3.2.2.2 Large Scale Plasmid Preparation

To obtain large quantities of plasmid DNA suitable for eucaryotic cell transfection, a single bacterial colony was inoculated into 5 ml LB medium containing 100 μg/ml of ampicillin for 8 h at 37°C with vigorous shaking (300 rpm). The 5 ml was then used to inoculate 250 ml of LB broth containing ampicillin (100 μg/ml). After overnight (16 h) incubation with vigorous agitation (300 rpm) at 37°C, cells were collected by centrifugation at 3000 rpm for 20 min at 4°C. The large scale plasmid preparation was carried out by using QIAfilter Maxi Plasmid Kit according to the standard protocol provided by the manufacturer. In brief, the pellets from 200 ml of LB medium were re-suspended in 10 ml of buffer P1 (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 μg/ml
RNase A) and lysed in 10 ml of buffer P2 (200 mM NaOH, 1% SDS). The lysates were then added to 10 ml of buffer P3 (3.0 M potassium acetate, pH 5.5), poured into the barrel of the QIAfilter Cartridge and incubated at room temperature for 10 min. After the cell lysate was filtered through the QIAfilter Cartridge into the previously equilibrated QIAGEN-tip, the DNA was eluted by addition of 15 ml buffer QF (1.25 M NaCl, 50 mM Tris-HCl, pH 8.5, 15% isopropanol) and precipitated by addition of 10.5 ml isopropanol to the elutant. The mixture was centrifuged at 15,000 X g for 30 min at 4°C. The DNA pellet was washed with 5 ml of 70% ethanol and centrifuged at the same speed for 10 min. The pellet was air-dried, redisssolved in 200-500 µl of sterile water, and stored at -20°C. The concentration of plasmid was quantitated by UV absorbance at 260 nm.

3.2.2.3 Isolation of DNA Fragments

PCR products and/or DNA fragments from the restriction enzyme digestion of plasmids were purified by agarose gel electrophoresis. Specific bands were excised from the gel and DNA was recovered by using the QIAquick Gel Extraction Kit according to the standard protocol provided by the manufacturer. The quantity of the isolated DNA was determined by ethidium bromide staining of the agarose gels.

3.2.2.4 Construction of Mutant LCAT Minigenes

The LCAT minigene containing the full length cDNA (McLean et al., 1986b) and the wild type sequence of intron 4 (Kuivenhoven et al., 1996) was first released from the pUC19-LCAT-intron 4 plasmid by the digestion of EcoR I and BamH I. The pUC19-LCAT intron 4 constructs that contained the natural mutation (IVS4-22c) and its two corresponding mutations (IVS4-22g and IVS4-22a) were also digested by the same restriction enzymes. The EcoR I-BamH I DNA fragments were then subcloned into
pcDNA 3.1(-), a mammalian expression vector using standard procedures (Maniatis et al., 1989). The pcDNA3.1 vector contains a strong human cytomegalovirus (CMV) immediate-early promoter that can be used to obtain high-level transient expression in a wide range of mammalian cells including HEK-293 cells. The LCAT intron 4 mutants were constructed by the PCR overlap technique (Ho et al., 1989) and the pcDNA3.1-LCAT intron 4 wild type plasmid used as template. The upstream (5'-' GGG AGA CCC AAG CTG GCT A -3') and downstream (5'-' CGT CGA GGC TGA TCA GCG G -3') primers used in the mutagenesis are complementary to sequences that flank the 5' and 3' sites of the polylinker in the pcDNA3.1 vector, respectively. The sequences of site-directed oligonucleotide primers used to create the desired intron mutations are listed in Table 4. In detail, the first PCR was performed with 100 ng of template DNA, 200 µM of each dNTPs, 0.4 µM of each upstream primer and reverse mutagenic oligonucleotide, or 0.4 µM of each downstream primer and forward mutagenic oligonucleotide. The PCR was performed as follows: 30 cycles of 30 sec. 95°C / 30 sec. 56°C / 60 sec. 72°C with 2.5 unit of Pfu DNA polymerase. To introduce a normal branch site into the natural (IVS4-22c) and branchpoint (IVS4-20t) mutants as well as the intron wild type (IVS4-WT), 0.4 µM of each upstream primer and antisense strand oligonucleotide (5'-' GTC AGG GGG GGC TGG GGC ACC TGC CC -3'), or 0.4 µM of each downstream primer and sense strand oligonucleotide [5'-' CCC TGA CCA GCT GCC CCG ACC CCT TC -3' (for IVS4-22c) or 5'-' CCC TGA CCA GCT GCC CTG ACC CCT TC -3' (for IVS4-20t) or 5'-' CCC TGA CCA GCT GCC CTG ACC CCT TC -3' (for IVS4-WT)] were used to amplify PCR product 1 or 2 in the presence of the respective template DNA, 200 µM of each dNTP, and 2.5 units of Pfu DNA polymerase. The resultant two PCR fragments were gel-purified and used in a subsequent fusion reaction in the presence of 0.4 µM of
each upstream and downstream primers under the same PCR condition except for the extension time being increased to 2 min. The final products purified from the agarose gel were digested with EcoR I and BamH I and then ligated into pcDNA3.1 vector digested with the same restriction enzymes. The presence of the introduced intron mutations was confirmed by dideoxy sequencing.
Table 4. Oligonucleotide primers used to create the desired intron mutations

<table>
<thead>
<tr>
<th>Name of oligonucleotide and position of the mutation relative to 3' splice site</th>
<th>Sequence of primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>IVS4-19a</td>
<td>5'- CTGCCCTGAA[<strong>CCCTCCAC</strong>] -3'</td>
</tr>
<tr>
<td>IVS4-19g</td>
<td>5'- CTGCCCTGACCCTCCAC -3'</td>
</tr>
<tr>
<td>IVS4-20c</td>
<td>5'- GCTGCCCTGGCCCTTCCA -3'</td>
</tr>
<tr>
<td>IVS4-20g</td>
<td>5'- GCT GCCCTGGCCCTTCCA -3'</td>
</tr>
<tr>
<td>IVS4-20t</td>
<td>5'- GCTGCCCTGTCCCTTCCA -3'</td>
</tr>
<tr>
<td>IVS4-21c</td>
<td>5'- AGCTGCCCTACCCCTTCC -3'</td>
</tr>
<tr>
<td>IVS4-21t</td>
<td>5'- AGCTGCCCTACCCCTTCC -3'</td>
</tr>
<tr>
<td>IVS4-23a</td>
<td>5'- CCAGCTGCCATGACCCCTT -3'</td>
</tr>
<tr>
<td>IVS4-23g</td>
<td>5'- CCAGCTGCCGTCGACCCCTT -3'</td>
</tr>
<tr>
<td>IVS4-24a</td>
<td>5'- CCCAGCTGCACTGACCCCT -3'</td>
</tr>
<tr>
<td>IVS4-24t</td>
<td>5'- CCC AGCTGCTCTGACCCCT -3'</td>
</tr>
<tr>
<td>IVS4-25a</td>
<td>5'- CCCCAGCTGACCTGACCCC -3'</td>
</tr>
<tr>
<td>IVS4-25g</td>
<td>5'- CCCCAGCTGGCTGACCCC -3'</td>
</tr>
<tr>
<td>IVS4-21a,-22c</td>
<td>5'- CAGCTGCC[<strong>CAACCCTTCC</strong>] -3'</td>
</tr>
<tr>
<td>IVS4-yeast</td>
<td>5'- CCAGCTG[<strong>TACTAAC</strong>]CCCTTCC -3'</td>
</tr>
</tbody>
</table>

The introduced mutations at the corresponding positions relative to the 3' splice site of LCAT intron 4 are in bold face and the 7 bp sequences replacing the wild type consensus sequence (5'-CCCTGAC-3') at the branch site are underlined.
3.2.2.4 Isolation of Total Cytoplasmic RNA from the Transfected Cells

Total cytoplasmic RNA was isolated from the transiently transfected HEK 293 cells using the RNeasy Mini Kit (Qiagen) according the standard protocol provided by the manufacturer. After the removal of culture medium, the cells were released by treatment with trypsin and cell pellet was collected by centrifuged at 300 X g for 5 min (Eppendorf, centrifuge 5417C). The cell pellet was resuspended in 175 μl cold bufer RLN (50 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1.5 mM MgCl₂, 0.5% NP-40), and incubated for 5 min on ice. Following centrifugation of the lysate at 4°C for 2 min at 300 X g, the supernatant was transferred to a new tube. 600 μl of lysis buffer RLT containing guanidium hydrochloride was added to the supernatant and vortexed vigorously. After adding 430 μl 100% ethanol to the mixture, the sample was mixed well by pipetting. The sample was then applied to the RNeasy mini spin column and total RNA was bound to the membrane by centrifugation for 15 sec at 13,000 rpm. After the membrane was washed by adding 700 μl buffer RW1 and 500 μl buffer RPE, respectively to the column followed by centrifugation, the RNA was eluted in 50-100 μl water by centrifuging for 1 min at 13,000 rpm. The concentration of the RNA was quantitated by UV absorbance at 260 nm.

3.2.2.5 Reverse Transcription and PCR Amplification (RT-PCR)

Before reverse transcription, the samples were treated with DNase I at 37°C for 30 min followed by inactivation of the enzyme at 56°C for 30 min. First-strand cDNA was then synthesized from 1.0 μg of cytoplamic RNA in a 20 μl reverse transcription mixture containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 1 mM dNTPs and 200 ng Oligo (dT)₁₂-₁₈. The tubes were heated at 75°C for 3 min and cooled immediately on ice.
M-MLV reverse transcriptase (200 units) was added. The tubes were incubated at 42°C for 1 h and subsequently heated at 95°C for 3 min. After reverse transcription, 10% of the reaction product was used for PCR amplification in a 50-μl final reaction volume (10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.4 μM of primers that were derived from sequences of exon 2 (5'- GCT ACC GCA AGA CAG AGG AC -3') and exon 5 (5' GGC CAA TGA GGA AGA CAG GC -3') of the LCAT gene, respectively, and 2 units of Taq DNA polymerase. To determine the efficiency of transient transfection, another primer pair complementary to the sequences (sense primer, 5'- GGG GTT CGA AAT GAC CGA CC -3'; antisense primer, 5'- CAG CTG GCA CGA CAG GTT TC -3') within the neomycin resistance gene of pcDNA3.1 vector was utilized to serve as an internal control. Twenty-five to thirty cycles of amplification were performed using a Perkin-Elmer DNA thermal cycler (System 2400). Each cycle consisted of a 30 sec denaturation at 95°C, a 30 sec annealing at 56°C, and 30 sec extension at 72°C. The PCR products were separated on a 3% agarose gel, and the bands were visualized by ethidium bromide staining.

3.2.3 Eucaryotic Cell Culture

HEK-293 cells were maintained as described in Section 2.2.3.

3.2.3.1 Transient Transfection of HEK-293 Cells

In chapter 2, BHK cells stably transfected with LCAT minigenes containing intron mutations were used to study the effect of the two novel mutations (IVS4-22g and -22a) at the same position of the natural mutation (IVS4-22c) on the splicing of LCAT pre-mRNA. In this chapter, HEK-293 cells were selected as an in vivo expressing system to investigate the possible mechanism(s) responsible for the defective splicing and to
establish further the functional significance of the BPS of intron 4 in the LCAT pre-mRNA splicing. It has previously been reported (Klein et al., 1995) that high-level expression of LCAT enzymes could be obtained by transient transfection of a LCAT cDNA plasmid containing the CMV immediate early promoter into HEK-293 cells. HEK-293 cells have also been employed to demonstrate that the mammalian branch site selection by U2 snRNP involves the base pairing during pre-mRNA splicing (Wu and Manley, 1989). Therefore, I decided to use the pcDNA3.1-LCAT minigenes that contain the intron mutations to transiently transf ect HEK-293 cells instead of performing stable transfection with these constructs.

Transient transfection of the LCAT minigene constructs into HEK-293 cells was performed using the calcium phosphate co-precipitation method (Rosenthal, 1987). The mammalian expression vectors, pcDNA3.1 and pcDNA3.1-LCAT cDNA as well as pcDNA3.1-LCAT IVS4 wild type were used as negative and positive controls, respectively. For transfection, 28 μg of each of the plasmids was added to 100-mm Petri dishes containing approximately 10^6 cells. After an overnight incubation at 37°C under 5% CO₂, the transfection medium was replaced with serum-free medium for 48 h. Culture medium was collected for the LCAT activity assays, and the cells were collected to prepare cytoplasmic RNA.

3.2.4 Protein Analysis

3.2.4.1 Measurement of LCAT Activity

As described in Section 2.2.4.2.
3.3 RESULTS

3.3.1 Generation of Mutations in the Branchpoint Sequence of Intron 4 of the LCAT Gene

To study the effects of mutations in the branchpoint consensus sequence on the efficiency of LCAT pre-mRNA splicing in vivo, the human LCAT minigene was first cloned into the mammalian expression vector, pcDNA3.1 (Fig.15a). Intron 4 of the human LCAT gene is a small intron consisting of only 83 base pairs (Fig.15b). The putative BPS (CCCTGAC, the underlined A denotes the branchpoint nucleotide) is located -19 to -25 bases upstream of the 3' splice site. For convenience, the nucleotide at -19 relative to the 3' splice site was designated as the first position of the branchpoint sequence. Based on the wild type pcDNA3.1-LCAT minigene, I made an additional series of point mutations in the branch-site (Table 5) using the synthetic oligonucleotides as described under "Materials and Methods".

Except for the natural mutation at position -22, two additional mutations at the same position, and the three branchpoint mutations, most of the rest site-directed mutagenesis involved nucleotide transversion according to the consensus sequence YNYTRAY. I also made one plasmid construct in which the wild type branchpoint sequence CCCTGAC was replaced by the absolutely conserved yeast branch-site consensus TACTAAC to test whether the invariant branchpoint sequence is the most preferred sequence for mammalian pre-mRNA splicing (Zhuang et al., 1989).
Figure 15. Schematic presentation of the pcDNA3.1-LCAT minigene construct. (A) The pcDNA3.1 contains the human CMV immediate early promoter, a multiple cloning site region (MCS, shaded box), and the 3' portion of the bovine growth hormone gene (BGH) including the polyadenylation site. The human LCAT minigene insert encompasses full-length LCAT cDNA as well as the wild type sequence of intron 4. The sizes of the exons 1-4 and 5-6 as well as the intron in base pairs are indicated above and below the elements, respectively. (B) The sequence of intron 4 of the human LCAT gene. The underlined nucleotides represent the putative branchpoint sequence in which T is mutated to C in the patients with FED. The asterisked adenosine represents the putative branchpoint nucleotide in the intron, and the base-pairing interaction of a conserved sequence (5'-GUAGUA-3') in human U2 snRNA with the branchpoint sequence is also shown.
Table 5. The effect of mutations in the branchpoint sequences of the fourth intron of the human LCAT gene on the activity of LCAT from medium of HEK-293 cells

<table>
<thead>
<tr>
<th>Branch site</th>
<th>Sequence</th>
<th>Activity (nmol/ml/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3.1</td>
<td></td>
<td>0.04 ± 0.01***</td>
</tr>
<tr>
<td>LCAT cDNA</td>
<td></td>
<td>23.49 ± 1.59***</td>
</tr>
<tr>
<td>IVS4-WT</td>
<td>CCCTGAC</td>
<td>5.15 ± 1.65</td>
</tr>
<tr>
<td>IVS4-yeast</td>
<td>TACTAAC</td>
<td>3.60 ± 0.51*</td>
</tr>
<tr>
<td>IVS4-19a</td>
<td>CCCTGAA</td>
<td>0.54 ± 0.28***</td>
</tr>
<tr>
<td>IVS4-19g</td>
<td>CCCTGAG</td>
<td>0.15 ± 0.11***</td>
</tr>
<tr>
<td>IVS4-20c</td>
<td>CCCTGCC</td>
<td>0.04 ± 0.01***</td>
</tr>
<tr>
<td>IVS4-20g</td>
<td>CCCTGGC</td>
<td>0.04 ± 0.01***</td>
</tr>
<tr>
<td>IVS4-20t</td>
<td>CCCTGTC</td>
<td>0.04 ± 0.02**</td>
</tr>
<tr>
<td>IVS4-21c</td>
<td>CCCTGAC</td>
<td>0.20 ± 0.14***</td>
</tr>
<tr>
<td>IVS4-21t</td>
<td>CCCTTAC</td>
<td>2.54 ± 0.32**</td>
</tr>
<tr>
<td>IVS4-22c (MUT)</td>
<td>CCCGCAC</td>
<td>0.03 ± 0.01***</td>
</tr>
<tr>
<td>IVS4-21a,-22c</td>
<td>CCCCAAC</td>
<td>0.13 ± 0.05§§</td>
</tr>
<tr>
<td>IVS4-22g (MUT-1)</td>
<td>CCCGGAC</td>
<td>0.02 ± 0.01***</td>
</tr>
<tr>
<td>IVS4-22a (MUT-2)</td>
<td>CCCAGAC</td>
<td>0.02 ± 0.01***</td>
</tr>
<tr>
<td>IVS4-23a</td>
<td>CCATGAC</td>
<td>3.61 ± 0.57*</td>
</tr>
<tr>
<td>IVS4-23g</td>
<td>CCGTGAC</td>
<td>1.56 ± 0.19***</td>
</tr>
<tr>
<td>IVS4-24a</td>
<td>CACTGAC</td>
<td>4.43 ± 1.29</td>
</tr>
<tr>
<td>IVS4-24t</td>
<td>CTCTGAC</td>
<td>3.86 ± 0.98</td>
</tr>
<tr>
<td>IVS4-25a</td>
<td>ACCTGAC</td>
<td>14.37 ± 1.90***</td>
</tr>
<tr>
<td>IVS4-25g</td>
<td>GCCTGAC</td>
<td>14.05 ± 1.43***</td>
</tr>
<tr>
<td>Mammalian consensus</td>
<td>YNYTRAY</td>
<td></td>
</tr>
<tr>
<td>Mammalian U2 snRNA</td>
<td>AUGAU G</td>
<td></td>
</tr>
</tbody>
</table>

All mutants were generated by overlap RCR method as described in material and methods. Positions that are mutated or altered are underlined. In the mammalian consensus sequence: Y=pyrimidine; R=purine; and N=any nucleotide. Confluent HEK-293 cells were incubated in serum-free Opti-MEM for 48 h, and the medium was assayed for LCAT activity using the HDL-like analogue as the substrate. Data shown represent means ± S.D. from duplicate measurements of triplicate transfections. **p<0.0001, ***p<0.001, and *p<0.05 compared to the wild type intron (IVS4-WT). §§§p<0.01 compared to the intron natural mutation (IVS4-22c).
3.3.2 Mutations of the Branch-site Adenine Completely Abolish Splicing of Intron 4 of the LCAT Gene in vivo

As described in previous chapters, the natural mutation of intron 4 and its two corresponding mutants, IVS4-MUT-1 and MUT-2, gave rise to neither spliced RNAs nor detectable LCAT proteins as the result of the defective intron splicing. The retention of intron 4 of human LCAT gene caused by the substitution of the thymine by other nucleotides in the putative branchpoint sequence suggests that the mutations of the branchpoint adenosine residue might also result in defective splicing for this particular intron. To examine the effect of the branchpoint mutations on the splicing of LCAT pre-mRNA, the pcDNA3.1-LCAT intron 4 constructs were transiently transfected into HEK-293 cells, and the cell media were collected to determine the enzymatic activities. As shown in table 5, the LCAT cDNA exhibited the highest LCAT activity. When the wild type intron 4 was inserted into cDNA the enzymatic activity decreased to approximately 20-25% percent of the original, which was most likely due to the influence of the splicing process on the expression of the gene. However, like the natural mutation, IVS4-MUT, and its two related mutations, IVS4-MUT-1 and MUT-2, substitution of the branchpoint adenine with a cytosine (IVS4-20c) resulted in the absence of LCAT activity, similar to control transfections with the empty pcDNA3 vector. A similar result was also obtained when the branchpoint adenine was changed to G (IVS4-20g) or T (IVS4-20t).

To determine further the effect of the branchpoint mutations on the LCAT pre-mRNA splicing in vivo, RT-PCR was performed. The wild type intron was efficiently spliced out of the pre-mRNA (Fig.16, lane 3) giving a spliced RNA band corresponding to the LCAT cDNA (Fig.16, lane 2) whereas no spliced RNAs could be detected for either the IVS4-MUT natural, MUT-1, and MUT-2 (Fig.16, lane 4-6) or the branchpoint mutants, IVS-20c, -20g, and -20t (Fig.16, lane 7-9). These observations are consistent
with previous reports (Zhuang et al., 1989; Reed and Maniatis, 1988) that mutations of the highly conserved thymine at the fourth position of the branchpoint consensus sequence have a similar severe effect on splicing as do mutations at the branch-site adenosine.

Figure 16. The effect of the branchpoint mutations on the LCAT pre-mRNA splicing as analyzed by RT-PCR. Total cellular RNA isolated from HEK-293 cells that were transiently transfected with the branchpoint mutants were analyzed by RT-PCR using primers complementary to sequences of exon 2 and exon 5, respectively. Correct splicing results in the excision of intron 4 giving a fragment of 389 bp corresponding to the same size of the LCAT cDNA. The unspliced product has a size of 472 bp containing the 83 bp sequence of intron 4. Lane 1: pcDNA 3.1 vector without insert; lane 2: LCAT cDNA; lane 3: IVS4-WT; lane 4: IVS4-MUT natural; lane 5: IVS4-MUT-1; lane 6: IVS4-MUT-2; lane 7: IVS4-20c; lane 8: IVS4-20g; lane 9: IVS4-20t; and M: molecular weight marker.
3.3.3 Intron 4 of the LCAT Gene Lacks Cryptic Branchpoint Sequence

Previous studies have shown that a mutation contained within the branch site sequence does not necessarily abolish but leads to a decrease of in vitro splicing (Padgett et al., 1985; Reed and Maniatis, 1985). The reduced rate of the splicing process has been explained by the activation of a nearby cryptic branch-site (Ruskin et al., 1985). In the previous study, I have not observed any alternative splicing of LCAT pre-mRNA containing the natural or the branchpoint mutations. Examination of the intron sequence upstream from the 3' splice site (Fig.15B) does not reveal a potential branch site that matches the mammalian consensus YNYURAY sequence, and this may be the reason why a complete abolition of splicing is observed. If this hypothesis is correct, the introduction of a BPS would allow the intron mutants to become efficiently spliced. To test whether this was the case, a normal BPS CCCTGAC was inserted into the intronic sequence just six bases upstream of the mutated branch sites (Fig.17). The results of the effect of the insertion of the BPS on the splicing of LCAT pre-mRNA are shown in Fig.18. The introduction of a normal branch site totally restored the enzyme activities from the natural and the branchpoint mutants, respectively. Interestingly, the insertion of the branch site into the branchpoint mutant IVS4-20t resulted in even higher LCAT activity compared with the wild type intron. However, the insertion had no obvious effect on the intron 4 wild type (Fig.18A). The efficiently spliced RNAs from the natural and the branchpoint mutants were revealed by RT-PCR (Fig.18B, lane 6 and 8, respectively), which further confirmed the effect of the insertion of a normal branch site into the intron sequence on the splicing of these mutated LCAT mRNA precursors. These results suggested that the inserted branch site was utilized when the normal BPS contains a mutation.
Figure 17. Insertion of a natural branch site into the sequences of mutated intron 4 of the human LCAT genes. Both the inserted and mutated branch sites are underlined. *, Normal branchpoint; (*), inserted branchpoint. (A) The insertion of a natural branch site into the wild type sequence of intron 4 of the human LCAT gene; (B) the insertion of a natural branch site into the naturally mutated intron 4 of the LCAT gene. The highly conserved T was mutated to C in the patients with FED; (C) the insertion of a natural branch site into the branchpoint mutation of intron 4 of the LCAT gene. The normal branchpoint A was substituted to T in the branch site.
Figure 18. The effect of the introduction of a normal branch site into the intron sequence on the splicing of the LCAT mRNA precursors containing the natural (IVS4-MUT) and the branchpoint (IVS4-20t) mutations. (A) Enzymatic activity secreted by transiently transfected HEK-293 cells. Bar 1: IVS4-WT; bar 2: IVS4-BPS-WT; bar 3: IVS4-MUT natural; bar 4: IVS4-BPS-MUT; bar 5: IVS4-20t; bar 6: IVS4-BPS-20t. ***p<0.0001 compared to the intron wild type; (B) RT-PCR analysis of total cytoplasmic RNA from HEK-293 cells transfected with the LCAT intron 4 mutants.
containing the branchpoint sequence insertion. *Lane 1*: pcDNA 3.1 vector without insert; *lane 2*: LCAT cDNA; *lane 3*: IVS4-WT; *lane 4*: IVS4-BPS-WT; *lane 5*: IVS4-MUT natural; *lane 6*: IVS4-BPS-MUT natural; *lane 7*: IVS4-20t; *lane 8*: IVS4-BPS-20t; and M: molecular weight marker. BPS represents the inserted normal branchpoint sequence (see Fig.17).

3.3.4 The Natural Mutation of Intron 4 Can Be Partially Suppressed by Changing G→A at the Third Position of the Branchpoint Sequence

It has been shown that the recognition of the poorly conserved mammalian pre-mRNA branch site sequence by U2 snRNA involves base pairing (Zhuang and Weiner, 1989; Wu and Manley, 1989). It has been previously proposed that the substitution of the thymine residue just two nucleotides upstream from the branchpoint adenine of intron 4 of the LCAT gene might interfere with its binding to AUGAUG sequence in human U2 snRNA, and thus lead to the defective splicing (Fig.19A, scheme b). There is a Wobble G-U base pair between the third position of the branch-site and the fifth position of AUGAUG (the underlined U is the supposed nucleotide). The natural point mutation T→C at the fourth position of the branch-site might not only destroy the base-paring itself with the fourth position of AUGAUG (the underlined A is the supposed nucleotide) but also disrupt the adjacent downstream weak G-U base-pairing interaction between the branchpoint sequence and U2 snRNA (Fig.19A, scheme b). This disruption would lead to inappropriate bulging of the adenosine residue in the branchpoint region. Therefore, if the G at the third position of the branchpoint sequence is altered to A, the Watson-Crick A-U base pair between the branch site and the AUGAUG sequence in U2 snRNA might compensate, at least in part, for the natural mutation in the branchpoint sequence (Fig.19A, scheme c). As shown in Fig.19B, the alteration from the sequence
CCCCGAC to CCCCAAC (IVS4-21a,-22c) did indeed increase the LCAT activity although it was still very low as compared to the wild type intron (Table 5). Also, a correspondingly small amount of correctly spliced transcript could be clearly identified (Fig.19C, lane 5). By contrast, no band corresponding to the mature mRNA transcript was observed in the cells that expressed the natural mutant (Fig.19C, lane 4). To exclude the possibility that the apparent effect of the double mutant (IVS4-21a,-22c) on the splicing of LCAT pre-mRNA resulted from the differences in the efficiency of transient transfection rather than specific base pairing interaction, I assayed the relative abundance of each expressed construct by RT-PCR with a set of primers that amplify 502 bp fragment of the neomycin resistance gene (NeoR) which is under the control of the SV40 early promoter in the pcDNA3.1 vector. Preliminary results (data not shown) have indicated that there is a linear relationship between the starting amount of mRNA (or cDNA) and the amount of amplified product during 15 to 25 cycles of PCR amplification, suggesting that the PCR reaction is in the exponential phase under these conditions. However, after 30 cycles of amplification, the PCR reaches the plateau phase and the yield is no longer proportional to the template. Therefore, in this experiment and the RT-PCR in the following experiment described in Section 3.3.5, 25 cycles of PCR were performed to make the amplification occurred in the range of exponential phase. As shown in Fig.19C, the amount of amplified NeoR gene product was similar in the cells transfected with each plasmid construct (lane 6-10), indicating that there were no differences in the transient transfection efficiency of the HEK-293 cells by the different plasmids. These observations suggest that the affinity of the base pairing between the U2 snRNA and the branchpoint sequence plays a role in the efficiency of pre-mRNA splicing.
Figure 19. The effect of the change of the branchpoint sequence from CCCCGAC to CCCCAAC on the splicing of LCAT pre-mRNA. (A) Proposed base-pairing interaction between the branchpoint sequence and U2 snRNA, (a) U2 snRNA-wild type intron 4 of LCAT pre-mRNA, (b) U2 snRNA-the natural mutation of intron 4 of LCAT pre-mRNA, (c) U2 snRNA-the doubly mutated intron 4 of LCAT pre-mRNA; (B) LCAT activity (nmol/ml/h) for CCCCGAC and CCCCAAC; (C) Gel electrophoresis showing the size of different fragments produced.
activity from the media of HEK-293 cells transfected with the natural (IVS4-22c) and the
double (IVS4-21a,-22c) mutants. "p<0.001 compared to the natural mutant; (C) Total
cytoplasmic RNA from HEK-293 cells transfected with the natural and the IVS4-21a,-
22c mutants for either LCAT (lane 1-5) or NeoR (lanes 6-10). RNA was amplified by RT-
PCR with the primer pairs as described under “Materials and Methods”. Lane 1 and 6:
pcDNA 3.1 vector without insert; lane 2 and 7: LCAT cDNA; lane 3 and 8: IVS4-WT;
lane 4 and 9: IVS4-MUT natural; lane 5 and 10: IVS4-21a,-22c; and M: molecular
weight marker.

3.3.5 Single-base Changes in the Branchpoint Sequence of Intron 4 Affect
Splicing of LCAT Pre-mRNA in vivo

To demonstrate further the functional role of the branchpoint sequence, I
identified the effects of mutations of other nucleotides in the branchpoint consensus
sequence on the efficiency of LCAT pre-mRNA splicing. The results from the transient
transfection of HEK-293 cells with the LCAT minigene constructs containing the intron
mutations show that single-base changes in the branchpoint sequence affect splicing
significantly as analyzed by LCAT activity assay and RT-PCR (Table 5 and Fig.20).
After the intron mutants were transiently transfected into HEK-293 cells, the LCAT
activities from the media were determined (Table 5 and Fig.20A). The mutants IVS4-
19g and IVS4-21c (Fig.20A, bar 4 and 5) were associated with the lowest LCAT
activities followed by IVS4-19a (Fig.20A, bar 3). A dramatic reduction in LCAT enzyme
activity was also observed with IVS4-23g (Fig.20A, bar 8). The mutants IVS4-21t and
IVS4-23a (Fig.20A, bar 6 and 7) also had a markedly decreased LCAT activities
compared with that of the wild type. However, the C→A transversion or C→T transition
at the sixth position (IVS4-24a or IVS4-24t) of the consensus sequence (Fig.20A, bar 9
and 10) had only a mildly decreased enzyme activity as compared to wild type control. Interestingly, in contrast to these downstream mutations, the change at the last position of the branchpoint sequence, i.e., IVS4-25a or -25g, resulted in an approximately 2.7-fold increase in LCAT activity compared with the intron 4 wild type (Fig.20A, bar 11 and 12).

The effects of these intron mutations on the efficiency of RNA splicing were further determined by RT-PCR. As expected, normally spliced RNA could be detected in all of these mutants, but the efficiency of splicing of different mutants varied significantly. The mutant IVS4-19g and IVS4-21c exhibited lowest yields of spliced RNAs (Fig.20B, lane 6 and 7) followed by IVS4-19a (Fig.20B, lane 5). The amount of the spliced RNAs for the mutants IVS-23a and 23g were less than that of wild type (Fig.20B, lane 9 and 10), and the levels of spliced RNA in IVS4-24a and -24t were similar to those of wild type control (Fig.20B, lanes 11 and 12). It was interesting to note that the RNA splicing of mutants IVS-25a and -25g was much more efficient than that of the wild type, as approximately 50% of the transcripts were spliced. All of these observations are generally consistent with the corresponding LCAT activities for these mutants. Therefore, it appears that the efficiency of splicing is well correlated with the activity of LCAT secreted from the transfected HEK-293 cells into the medium. Another surprising observation in this study was that the replacement of the wild type branchpoint sequence of intron 4 with the invariant TACTAAC box of yeast did not increase the efficiency of splicing but had a mild decrease in both spliced RNA (Fig.20B, lane 4) and LCAT activity (Fig.20A, bar 2). This result was unexpected based on the previous reports (Zhuang et al., 1989) suggesting that the yeast branchpoint sequence was the optimal sequence for the splicing of mammalian introns.
Figure 20. The effects of the mutations in the branchpoint sequence on the efficiency of LACT pre-mRNA splicing. (A) LCAT activities of HEK-293 cell media for intron 4 wild type (WT) and the intron mutants. The enzymatic activities (in nanomoles of cholesterol ester formed per ml and per h) are expressed as a percentage of wild type LCAT intron 4 activity. *Bar 1: IVS4-WT; bar 2: IVS4-yeast; bar 3: IVS4-19a; bar 4: IVS4-19g; bar 5: IVS4-21c; bar 6: IVS4-21t; bar 7: IVS4-23a; bar 8: IVS4-23g; bar 9: IVS4-24a; bar 10: IVS4-24t; bar 11: IVS4-25a; bar 12: IVS4-25g. **p<0.0001, ***p<0.001,
*p<0.05 compared to the wild type intron. (B) RT-PCR on total cytoplasmic RNA from HEK-293 cells that were transiently transfected with LCAT minigene constructs carrying a series of mutations in the branchpoint sequence. Lane 1: pcDNA 3.1 vector without insert; lane 2: LCAT cDNA; lane 3: IVS4-WT; lane 4: IVS4-yeast; lane 5: IVS4-19a; lane 6: IVS4-19g; lane 7: IVS4-21c; lane 8: IVS4-21t; lane 9: IVS4-23a; lane 10: IVS4-23g; lane 11: IVS4-24a; lane 12: IVS4-24t; lane 13: IVS4-25a; lane 14: IVS4-25g; and M: molecular weight marker.
3.4 DISCUSSION

Our lab has previously shown that a point mutation in a lariat branchpoint sequence of intron 4 of the LCAT gene (IVS4:T-22C) resulted in the intron retention causing a human inherited disorder, FED (Kuivenhoven et al., 1996). In the previous chapter, I have demonstrated that the uracil residue at the fourth position of the branchpoint sequence in which the natural mutation occurs is crucial for the pre-mRNA splicing. In this chapter, I extend the studies into the whole region of the branchpoint sequence and further demonstrate that the branchpoint sequence of intron 4, especially the branchpoint adenosine residue and the nucleotides surrounding it, is essential for the efficient splicing of LCAT messenger RNA precursors.

3.4.1 The Branchpoint Adenosine Is Essential for the Efficient Splicing of LCAT pre-mRNA

The formation of a lariat structure during the first step of nuclear pre-mRNA splicing indicates that the branchpoint sequence is involved in the process. The branchpoint nucleotide is typically the adenosine residue in the sequence but splicing also can proceed using other nucleotides (Horning et al., 1986; Hartmuth and Barta, 1988). Previous studies have shown that deletion or mutation of the branchpoint sequence does not abolish splicing both in vitro and in vivo (Padgett et al., 1985; Reed and Maniatis, 1985; Ruskin et al., 1985). This paradox has been explained by the utilization of relatively inefficient nearby cryptic branchpoint sequence (Ruskin et al., 1985). However, the results presented here clearly demonstrate that this is not always the case. The single-base substitutions of the adenosine reside in the branchpoint region of intron 4 of LCAT pre-mRNA completely abolish splicing as assayed by the enzyme activity and RT-PCR. Newman et al. has previously shown that mutations of the
adenosine residue in the yeast branchpoint sequence abolish RNA splicing by preventing the cleavage at the 5' splice site (Newman et al., 1985). In this respect, the mutations of the branchpoint A as well as the natural mutation in intron 4 of the LCAT gene might have the same effect on the cleavage at the 5' splice site. Since the intronic sequence contains a stop codon signal, the failure to detect any LCAT activity and the truncated protein (data not shown) in the culture medium indicate that the translation products are probably rapidly degraded within the cells. As discussed in the previous chapter, the real case is most likely that the LCAT primary transcript containing the premature termination signal undergoes nonsense-mediated mRNA decay (NMD) pathway (Frischmeyer and Dietz, 1999; Hilleren and Parker, 1999; Mitchell and Tollervey, 2000) rather than the synthesis of nonfunctional truncated protein. Recently, another example of the branchpoint mutation that disrupts mRNA splicing in intron 9 of the low density lipoprotein (LDL) receptor gene has been described in a patient with familial hypercholesterolaemia (FH). This mutation disrupts the branchpoint consensus sequence and also causes intron retention (Webb et al., 1996). These observations, therefore, suggest that the branchpoint sequence plays an important role in the splicing of nuclear pre-mRNA although it is poorly conserved in mammals.

3.4.2 The Insertion of a Normal Branch Site into the Intron Sequence of the Intron Mutants Completely Restores the Splicing

In order to test the hypothesis that the sequence of LCAT intron 4 lacks cryptic branch sites that can be used when the normal BPS contains a mutation, I inserted a normal BPS CCCUGAC into the natural and the branchpoint mutants (IVS4-22c and IVS4-20t, respectively) just six nucleotides upstream of the mutated branch sites (Fig.17). The purpose of the insertion is to see whether this sequence could rescue the
natural or the branchpoint mutation. As expected, I observed not only the complete restoration of the LCAT activities of both the mutants but also an increase in splicing efficiency of the branchpoint mutant compared with the intron 4 wild type construct. These results suggest that the sequence of intron 4 itself does not contain any potential cryptic branch sites. Once the branchpoint sequence had been mutated there was no alternative to back up the splicing machinery so that the consequence of the natural and the branchpoint mutations was retention of the intron. At the moment, I do not know the reasons why the introduction of the normal branch site into the intronic sequence of the branchpoint mutant increased the efficiency of LCAT pre-mRNA splicing. It is possible that two potential branch sites within a single intron might compete with each other (Zhuang et al., 1989). For the intron wild type and the natural mutant, either of the two potential branchpoint sequences could be screened as the branch site signal, but the one that mostly matches the consensus sequence would be preferentially utilized. In the case of the branchpoint mutant, the inserted branch site becomes the only branchpoint sequence to be used in the pre-mRNA splicing, which may be one of the possible explanations for the increased efficiency of pre-mRNA splicing.

3.4.3 The Replacement of the Branchpoint Sequence CCCCGAC to CCCCAAG Can Partially Rescue the Natural Intron Mutation

The base-pairing interaction between the branchpoint consensus sequence and the sequence AUGAUG in U2 snRNA is one of the early essential events in pre-mRNA splicing (Parker et al., 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989). I have made the compensatory mutation AUGGUG in human U2 snRNA (the wild type human U2 snRNA cloned in pUC13, kindly provided by Dr. AM, Weiner) by overlap-PCR in an attempt to suppress the natural intron mutation (CCCCGAC). After the mutant U2
snRNA and the natural mutant were co-transfected into the HEK-293 cells, I was unable to detect either LCAT activity or spliced RNA from the transfected cells (data not shown). However, the failure of suppression of the natural intron mutation with the compensatory mutation in U2 snRNA could not let me exclude the possibility of a base pairing between these two mutants. One of the possible explanations could be that the expression levels of the mutant U2-gene might be not high enough to compete with the abundance of the endogenous wild type U2-snRNA in the cells. It is also possible that our LCAT activity assay and the RT-PCR are not sensitive enough to detect such small changes of the splicing efficiency due to the potential base-pairing between the two mutants. Therefore, as an alternative approach, I changed the naturally mutated branchpoint sequence from CCCCGAC to CCCCAAC. This alteration would produce a stronger Watson-Crick base-pair between the branchpoint region and the AUGAUG sequence in U2 snRNA compared to the natural mutation (Fig.19A), which might, at least in part, suppress the effect of the natural mutation on the LCAT pre-mRNA splicing by increasing the ability to tolerate a mismatch at the adjacent position. As expected, the base change just one nucleotide upstream of the branchpoint adenosine residue did recover some of the LCAT activity in the medium and produced a clearly spliced RNA band although these changes were relatively small compared with the wild type control (Fig.19B and 19C, and Table 5). Since only a very low levels of correctly spliced RNA could be detected in the double mutant compared with that of the wild type intron, it was important to demonstrate that the significant increase in the LCAT activity as well as the splicing efficiency of the double mutant as compared to the natural mutant was due to the improvement of base-pairing with the AUGAUG sequence in the U2 snRNA (Fig.19A, scheme c) rather than the differences in the efficiency of transient transfection. To address the issue of the controls for the efficiency of transient
transfection, I designed another pair of primers complementary to the sequences of the Neo<sup>R</sup> gene in the pcDNA3.1 vector. If the expression levels of both controls and the intron mutant samples were similar as determined by RT-PCR, it could be demonstrated that there were no differences in the efficiency of the transient transfection of various LCAT minigene constructs into HEK-293 cells. The results as shown in Fig.19C, lane 6-10 have demonstrated that this was the case. Therefore, the correctly spliced transcript of the CCCCCAAC variant (Fig.19C, lane 5), albeit inefficiently formed compared with wild type intron control (Fig.19C, lane 3), results most likely from the specific interaction of the intron variant with U2 snRNA. This observation indicates that the strength of the base pairs between the pre-mRNA and U2 snRNA may be one of the critical factors in the determination of the efficiency of nuclear pre-mRNA splicing.

3.4.4 Individual Nucleotides in the Branchpoint Sequence of Intron 4 Are Important in Determining the Efficiency of LCAT Pre-mRNA Splicing

The above mentioned conclusion was further substantiated by the mutational analysis of the functional significance of other specific nucleotides in the branchpoint consensus sequence. In addition to the mutations at the fourth position of the branchpoint sequence, the alteration of nucleotides, especially around the branchpoint A, i.e., IVS4-19a, -19g, and -21c has a dramatic effect on the splicing of the intron probably because these mutations greatly reduce potential base-pairing with the U2 snRNA. And if the branchpoint adenosine residue was not flanked by a base-paired residue on either its 3' site or 5' site, such as in the case of IVS4-19a, -19g and -21c, the branch nucleotide would not be properly bulged, and thus affecting the splicing. Recently, Pascolo <i>et al.</i> have demonstrated that U2 snRNA does indeed base pair with the nucleotides preceding and following the branchpoint adenosine residue that is
constrained into a bulged conformation (Pascolo and Seraphin, 1997). The mutants IVS4-23g, -21t, and -23a also have significant effects on the splicing efficiency, probably through the same mechanism. Interestingly, in contrast to these downstream mutations, the transversion of the last nucleotide of the BPS from C→A (IVS4-25a) or G (IVS4-25g) significantly enhanced the efficiency of LCAT pre-mRNA splicing up to 2.7-fold as compared to the wild type control. However, the substitution of the sixth position in the branchpoint sequence appeared to have less effect on the splicing, which is consistent with the consensus sequence in which the nucleotide could be either a purine or pyrimidine. These observations suggest that the potential base-pairing between the first four nucleotides in the BPS and the U2 snRNA are probably one of the major determinants in the splicing of nuclear pre-mRNA, and this might also explain why the branch site in mammalian introns could be more degenerate. The reasons for the markedly increased splicing efficiency in the two mutants, IVS4-25a and -25g, are not clear. However, these findings raise the possibility that a DNA polymorphism involving the branchpoint sequence might affect the efficiency of mRNA splicing and thus have significant clinical implications. A common C→T polymorphism that resides in the last position of the branchpoint consensus sequence of intron 9 of the LDL receptor gene has been reported (Webb et al., 1996). However, this polymorphism has been observed to have no effect either on splicing in vitro or on plasma lipid concentrations in the population. This is not surprising because the last position of the consensus sequence for the mammalian branchpoint sequence could be either C or T.

3.4.5 The Invariant Branchpoint Sequence (TACTAAC) of Yeast does not increase the efficiency of LCAT pre-mRNA splicing
Another unexpected observation made in this study is that the substitution of the normal BPS CCCUGAC with the yeast branch site UACUAAC sequence did not increase the efficiency of LCAT pre-mRNA splicing but modestly decreased it. This result is surprising because it conflicts with a previous study (Zhuang et al., 1989) in which the data have shown that the optimal branchpoint sequence is the yeast UACUAAC sequence for mammalian introns. Although the UACUAAC has perfect complementarity with the sequence AUGAUG in U2 snRNA, the results of my transfection experiments seem not to support the concept that the perfect match improves the efficiency of pre-mRNA splicing in vivo. Rather, they are consistent with the results that mutation of the last position of the BPS markedly increases the splicing efficiency even though the changes would reduce the match within the consensus sequence. This would, at least in part, explain why splicing is less dependent on base pairing of the pre-mRNA with U2 snRNA in mammals. For the mammalian introns, it is likely that other protein-RNA and RNA-RNA interactions are required to maintain the pre-mRNA and the U2 snRNA in the correct conformation in the absence of one or two of these base pairs.

In conclusion, by taking advantage of our in vivo assay for measuring the splicing efficiency, I am able to demonstrate that the branchpoint consensus sequence of intron 4 is essential for the splicing of LCAT pre-mRNA. With the increasing amount of intron sequence data available, more intron mutations far from the splice sites will likely be found to be the underlying causes for the human genetic disorders.
CHAPTER 4

Remarks and Future Directions
4.1 Implications of the studies

In the studies reported above, I demonstrated that the branchpoint consensus sequence of intron 4 is essential for accurate and efficient splicing of the LCAT pre-mRNA. As expected, mutations of the branchpoint adenosine residue in the branch site abolish RNA splicing, and the substitution of an uracil residue just two bases upstream of the putative branchpoint adenosine also completely inhibits pre-mRNA splicing. However, the introduction of the wild-type BPS into the intronic sequences of the natural (IVS4-22c) and the branchpoint (IVS4-20t) mutants only six bases upstream of the mutated BPS can totally restore the RNA splicing, suggesting that the BPS is the only branch site used in the intron for splicing. When these two important sites in the BPS of the LCAT intron 4 were mutated, there were no alternative and/or cryptic branchpoint signals available to support the splicing machinery, and this may be the reason that splicing of intron 4 of the human LCAT primary transcripts is vulnerable to the mutations in the branchpoint region. In this respect, the splicing of LCAT intron 4 is very similar to that of yeast introns. The result that the natural mutation can be partially rescued by making a single nucleotide change (G→A) adjacent to the mutated thymine within the branchpoint consensus sequence supports our assumption that the loss of the thymine residue might interfere with the binding of U2 snRNA to the branchpoint region, and thus affect LCAT pre-mRNA splicing. These conclusions are further supported by the observations that other single-base changes within the BPS also significantly affect the efficiency of splicing and thus enzyme activity. Mutations, particularly around the branchpoint adenosine residue that decrease the potential base pairing interactions between U2 snRNA and the BPS markedly reduce the splicing efficiency.

Site-directed mutagenesis studies of the branchpoint consensus sequence not only provide further evidence that the branch site is crucial in splicing of human nuclear
mRNA precursors but also shed insight into the understanding of the mechanisms of selection of the branchpoint during pre-mRNA splicing. In this regard, intron 4 of the LCAT gene serves as a suitable model to study the functional significance of the branchpoint sequence in pre-mRNA splicing due to its vulnerability to any nucleotide changes within the consensus sequence. By comparing the short sequences (from 70 to 88 base pairs) of LCAT intron 4 from 17 different species (Data from Genbank), it is found that the potential branch sites are highly conserved according to the consensus sequence YNYURAY among these species. The putative branchpoint adenosine residue is located between -20 to -30 relative to the 3’ splice site. These observations suggest that mutations of these branchpoint sequences might have similar effects on the splicing of the pre-mRNA as shown in the studies for human LCAT gene. I believe that there must be many more branchpoint mutations causing human inherited disease that have not been identified. Therefore, our conclusion that the search for mutations should include consideration of regions other than the promoter region, coding sequences and intron-exon junctions of the gene will help to define more human genetic diseases.

4.2 Implications of DNA polymorphism in the branchpoint consensus sequence

One of the interesting findings made in this study is that the changes at the last position of the branchpoint consensus sequence of intron 4 of the LCAT gene surprisingly increase the efficiency of splicing up to approximately three times compared with the wild type intron. As the expression of the wild type intron is relatively low as compared to the LCAT cDNA positive control, I was able to observe an increase in the efficiency of RNA splicing in the two intron mutants, i.e., IVS4-25a and IVS4-25g. It has
been shown that the length of an intron can affect the splicing efficiency (Wieringa et al., 1984). Reducing the size of an intron would decrease the efficiency of splicing accordingly. The minimal size of the intron for efficient splicing is about 80 nucleotides (Wieringa et al., 1984). Since intron 4 of the LCAT gene is only 83 bases in size, it is not surprising to see why the expression level of LCAT minigene is not as high as that of LCAT cDNA. However, this interesting observation suggests that DNA polymorphism involving the branchpoint sequence might affect the efficiency of RNA splicing.

Polymorphism in introns of certain genes have been shown to be associated with a particular phenotype (Lacoviello et al., 1998; Kuivenhoven et al., 1998). As discussed in Section 3.4.4, a common C→T polymorphism in the LDL receptor gene that also resides at the last position of the branchpoint consensus sequence has been described (Webb et al., 1996). It is not surprising that this polymorphism has no effect on splicing in vitro or on plasma lipid concentrations in the population because it could be either C or T according to the branchpoint consensus sequence YNYURAY (Y=pyrimidine; R=purine; N=nucleotide; and the underlined Y is the last nucleotide of the consensus) in mammals. Another example of polymorphism within the consensus branch site was reported in the gene encoding coagulation factor VIII (Kogan and Gitschier, 1990). Although no alteration in gene expression is expected as a result of the polymorphism (CTCTG/AAC, G/A indicates the position of the polymorphism; the underlined A represents the putative branchpoint) as it conserves the purine normally present at this location, this polymorphism is indeed useful for genetic prediction in some cases with hemophilia A. However, the branch site polymorphism itself is of interest. What would happen if the polymorphism within the branchpoint consensus sequence involves a nucleotide transversion? In this case, it would be expected that such a polymorphism would affect RNA splicing and/or gene expression. Therefore, DNA polymorphisms
close to or within branchpoint consensus sequence may regulate the net rate of gene expression through changes in the efficiency of pre-mRNA splicing. Although very few of these polymorphisms (Webb et al., 1996; Kogan and Gitschier, 1990; Fetzer et al., 1999) have been described so far, my results involving the point mutations in the branchpoint consensus sequence in intron 4 of the LCAT gene suggest that such a branch site polymorphism could have significant clinical implications.

4.3 Future directions

4.3.1 In vivo studies of splicing

In order to gain further insight into the molecular basis for defective splicing, an in vivo tissue culture model system that consists of transiently or stably transfected cells expressing easily quantifiable levels of wild type and mutant transcripts as well as the protein products would be ideal. In this respect, our LCAT expression and assay system provide a reporter gene model to detect the effects of various mutations in the 5' and 3' splice sites as well as the branchpoint sequence on the efficiency of RNA splicing. Furthermore, a LCAT reporter gene system could also be used to study intron mutations from other human genes that affect splicing of nuclear messenger RNA precursors as well as the effects of polymorphisms in human introns that have potential clinical significance on the efficiency of RNA splicing.

As mentioned above in Section 4.2, increasing the size of LCAT intron 4 would be expected to increase splicing efficiency. More importantly, a LCAT minigene containing the natural or branchpoint mutation of intron 4 could be used as a means to further test the exon-definition model. For this purpose, we could first construct the LCAT minigene containing both intron 4 and intron 5, then insert a large piece of another intronic fragment into intron 4 sequence upstream of either the natural or the
branchpoint mutation to see whether the splicing pattern would be changed, e.g. intron retention vs. exon skipping. Mutation of splice sites in genes with small introns usually has a different splicing phenotype than the same mutation in genes with large introns. Since the LCAT intron 4 is a small intron, the same mutation in the branchpoint sequence of the larger introns might have different impact on the efficiency of splicing. If the hypothesis is correct, we would expect to see the change of splicing pattern from intron retention to exon skipping according to the exon-definition model (Berget, 1995).

4.3.2 In vitro studies of splicing

Although our in vivo transfection experiments provide direct evidence that mutations in the branchpoint sequence have significant effects on the pre-mRNA splicing, it is difficult to determine which step is responsible for the defective splicing. For example, splicing intermediates and the lariat intron products can rarely be detected in vivo, and only the final splicing products can be analyzed.

The mutation of the branch point in yeast not only abolishes the splicing of the introns in vivo, but also inhibits the splicing in vitro (Newman et al., 1985). Data from in vitro splicing reactions showed that no branches were made, as evidenced by the absence of lariat RNAs. Similarly, there was no cleavage at the 5' splice junction, since the branchpoint mutant produced no detectable amount of correctly spliced mRNA. The human LCAT intron 4 natural and the branchpoint mutants have a similar effect on splicing in vivo as that of the yeast intron. Do they have similar effects on splicing in vitro as well? The answer to this question will depend on the results from experiments of in vitro splicing reactions.

The development of cell-free in vitro splicing systems has led to the elucidation of a two-step transesterification reaction mechanism. Unlike the in vivo studies, the in vitro
systems allow analysis of the reaction intermediates and products; therefore, the effects of various intron mutations on the different steps of the splicing reaction can be tested. Radioactively labeled pre-mRNA substrates are usually generated by in vitro transcription of cloned templates using bacteriophage SP6 or T7 RNA polymerase and processed in HeLa nuclear extracts. The RNA products of splicing reactions are analyzed in denaturing polyacrylamide gels and detected by autoradiography. The efficiency of splicing can be then determined by quantitation of radioactive bands through densitometry of autoradiograms. By using this approach, it could be determined whether the mutations in the branchpoint consensus sequence affect the formation of the lariat structure according to the sizes of different fragments. Furthermore, we can confirm the actual branchpoint nucleotide of an intron by performing primer extension on in vitro splicing intermediates and products. The results obtained from these experiments would complement the in vivo studies and provide important clues for understanding the mechanism of accurate branch-site selection during nuclear pre-mRNA splicing.

4.4 Conclusions

Mutations within the branchpoint sequence causing human genetic disease have not been identified as frequently as mutations within splice sites. This is probably because they are not readily detectable and are most likely to result in a less severe clinical phenotype. One of the possible explanations is that the branchpoint sequence appears to exhibit a weak consensus sequence and in some cases the mutations of the branchpoint sequence activate the use of the nearby cryptic branch site.

To date only a small number of systems have been employed to analyze the consequences of the naturally occurring mRNA splicing mutations in vivo. My work
provides one of the most extensive studies on the functional significance of the branchpoint sequence *in vivo*. Branch site mutations have been neglected as a cause of disease. The results from my work presented here have demonstrated that such mutations, at least in certain introns, could have significant consequences in affecting RNA splicing and gene expression. As shown previously, the absence of cryptic or alternative branchpoint sequences in these introns may be essentially vulnerable to such mutations and cause human disease. The findings made in this study have also suggested that DNA polymorphisms involving the branchpoint consensus sequence might affect the efficiency of RNA splicing and thus have significant clinical implications.


Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlange, S., Buchler, C., Porsch-Ozcurumez, M., Kaminski, W. E., Hahmann, H. W.,


associated with the fish eye syndrome. Lecithin-cholesterol acyltransferase (Thr123—Ile) and lecithin-cholesterol acyltransferase (Thr347—Met). *J. Clin. Invest.* **89**, 499-506.


Kramer, A., Mulhauser, F., Wersig, C., Groning, K., and Bilbe, G. (1995) Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is
homologous to the essential splicing factor PRP21p of Saccharomyces cerevisiae. 

Kramer, A. (1996) The structure and function of proteins involved in mammalian pre-

Krawczak, M., Reiss, J., and Cooper, D. N. (1992) The mutational spectrum of single 
base-pair substitutions in mRNA splice junctions of human genes: causes and 

Kuivenhoven, J. A., Stalenhoef, A. F., Hill, J. S., Demacker, P. N., Errami, A., Kastelein, 
J. J., and Pritchard PH. (1995a) Two novel molecular defects in the LCAT gene are 

Kuivenhoven, J. A., van Voorst tot Voorst, E. J., Wiebusch, H., Marcovina, S. M., Funke, 

Kuivenhoven, J. A., Weibusch, H., Pritchard, P. H., Funke, H., Benne, R., Assmann, G., 
and Kastelein, J. J. (1996) An intronic mutation in a lariat branchpoint sequence is a 
364.

(1997) The molecular pathology of lecithin:cholesterol acyltransferase (LCAT) 

Kuivenhoven, J. A., Jukema, J. W., Zwinderman, A. H., de Knijff, P., McPherson, R., 
the cholesteryl ester transfer protein gene in the progression of coronary 

the differential binding of U1 small nuclear ribonucleoprotein particle. *Science.* **251**, 
1045-1050.

assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc. 


the conserved, intron-contained TACTAAC box as an essential splicing signal sequence 


Query, C. C., Bentley, R. C., and Keene, J. D. (1989b) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell. 57, 89-101.


adhesion molecule L1 mRNA in a family with X-linked hydrocephalus. Nat. Genet. 2,
107-112.


Involvement of U1 small nuclear ribonucleoproteins (snRNP) in 5' splice site-U1 snRNP

Ruby, S. W., and Abelson, J. (1988) An early hierarchic role of U1 small nuclear


Ruskin, B., Greene, J. M., Green, M. R. (1985) Cryptic branch point activation allows

Ruskin, B., and Green, M. R. (1985a) Role of the 3' splice site consensus sequence in

Ruskin, B., and Green, M. R. (1985b) An RNA processing activity that debranches RNA


Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J. C., Deleuze, J. F.,
caused by mutations in the gene encoding ATP-binding cassette transporter 1. Nat.
Genet. 22, 352-255.

of the Yeast Saccharomyces: Gene Expression, Volume II (Jones, E. W., Pringle, J. R.,
Harbor, New York.

Sakai, N., Vaisman, B. L., Koch, C. A., Hoyt, R. F, Jr., Meyn, S. M., Talley, G. D., Paiz,
mouse lecithin:cholesterol acyltransferase (LCAT) gene. Generation of a new animal

requirements determined by deletion analysis of an intervening sequence. Proc. Natl.
Acad. Sci. USA. 82, 2885-2889.


145


