Clinical Investigation and Reports

Splice-Site Mutations in Atherosclerosis Candidate Genes Relating Individual Information to Phenotype

Yskert von Kodolitsch, MD; Reed E. Pyeritz, MD, PhD; Peter K. Rogan, PhD

Background—Nucleotide variants in several genes for lipid and methionine metabolism influence the risk of premature atherosclerosis. Ten percent of single nucleotide substitutions in these genes involve mRNA splice sites. The effects of some of these changes on splicing and on phenotypic severity are not inherently obvious.

Methods and Results—Using an information theory-based model, we measured the individual information content (R_i) of splice sites adjacent to 289 mutations (including 31 splice-site mutations) in the atherosclerosis candidate genes APOAI, APOB, APOCII, APOE, CBS, CETP, LCAT, LIPA, LDLR, and LPL. The predictions of information analysis were then corroborated by published mRNA analyses. The R_i values of mutant sites were consistent with either complete (n=17) or partial (n=5) inactivation of these sites. Seven mutations were predicted to activate cryptic splice sites. Predicted inactive mutant sites were associated with either “average” or “severe” dyslipidemia and commensurate reductions in protein levels or activity, whereas mutations expected to exhibit residual splicing had average or “mild” effects on lipid and protein expression.

Conclusions—Information analysis of splice-junction variants in atherosclerosis candidate genes distinguishes inactive from leaky splice sites and identifies activated cryptic sites. Predicted changes in splicing were related to phenotypic severity. (Circulation. 1999;100:693-699.)

Key Words: atherosclerosis ■ genetics ■ lipids ■ risk factors ■ RNA

Methods

Selection of Mutations

Information theory–based models that account for all of the nucleotide variations in splice sites have been used to predict the activities of natural and mutant sites and identify cryptic splice sites. The information content (in bits) of a member of any sequence family describes the degree to which that member contributes to the conservation of the entire family. The effects of all nucleotide changes are detectable, given that information is cumulative over all positions in a splice site.

In previous studies, lipid and protein levels or enzymatic activities of gene products were used to indirectly assess the effects of mutations that predispose a person to atherosclerosis. We used information theory–based models of mRNA splicing to relate the severity of splicing mutations to atherosclerosis phenotypes.

Received December 31, 1998; revision received May 10, 1999; accepted May 25, 1999.

From the Department of Cardiology (Y.v.K.), University Hospital Eppendorf, Hamburg, Germany; Departments of Human Genetics, Medicine, and Pediatrics (R.E.P.), Allegheny General Hospital, Pittsburgh, Pa; and Section of Molecular Genetics and Molecular Medicine (P.K.R.), Children’s Mercy Hospital, Kansas City, Mo.

Correspondence to Peter K. Rogan, PhD, Section of Molecular Genetics and Molecular Medicine, Children’s Mercy Hospital and Clinics, 2401 Gillham Road, Kansas City, Mo 64108. E-mail progan@cmh.edu

© 1999 American Heart Association, Inc.

Circulation is available at http://www.circulationaha.org

693
Computation of Individual Information

The individual information contents (R_i) of natural and mutant splice sites and coding sequence variants in atherosclerosis candidate genes were compared. The computer programs Scan, MakeWalker, and Lister were used to compute R_i values and display results.

The information contents of potential splice sites were determined for sequences up to 150 nucleotides upstream and downstream of each nucleotide substitution. R_i values could not be directly assessed for the LDLR and CBS genes because the terminal nucleotides of the corresponding splice junctions (positions −16 to −25 for acceptor sites, and +6 for donor sites) were unavailable. To model incomplete sites, the possible range of R_i values was defined by computing R_i for the best and worst sequence combinations. These ranges were computed for the missing sequence data from the corresponding coordinates of the donor and acceptor weight matrix elements that produced the maximal and minimal R_i values. Interpretation of variants was possible only if the R_i intervals of the natural and corresponding mutant sites did not overlap. Inactivating mutations were distinguishable from those that resulted in leaky splicing if the R_i interval of the mutant site was consistently either less or greater than 2.4 bits, respectively. Six of 171 variants with incomplete acceptor splice sites could not be interpreted because the R_i intervals of the variant and corresponding natural sequences overlapped.

Criteria for Grading Phenotypic Severity

For each splicing mutation, reference distributions of protein expression, plasma lipids, or plasma homocystine were derived from individuals with other mutations in the corresponding gene from previously published reports (data not shown) (Table 1). Individuals with mutations in the same gene were also distinguished according to genotype because the diagnosis, in some instances, depended on the mode of inheritance. Protein levels and enzymatic and/or binding activity were measured by the same methods for each splice-site mutation and the corresponding reference population; these values are given as the reported range of the percentage of normal protein levels. The mean and 95% confidence intervals (±2SD) of lipid values for each reference population were either derived from published reports or computed from values in published patient cohorts.

### Table 1. Reference Parameters for Atherosclerosis Candidate Gene Mutations

<table>
<thead>
<tr>
<th>Gene</th>
<th>Disorder</th>
<th>Trait</th>
<th>Missense Mutation</th>
<th>Splicing Mutation</th>
<th>TG</th>
<th>TC</th>
<th>LDL-c</th>
<th>HDL-c</th>
<th>TG</th>
<th>TC</th>
<th>LDL-c</th>
<th>HDL-c</th>
<th>Disease Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>APOA1</td>
<td>APO A-Ⅱ deficiency</td>
<td>rec</td>
<td>&lt;1</td>
<td>...</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>129±36.2</td>
<td>231±31</td>
<td>...</td>
<td>57±11</td>
<td>...</td>
</tr>
<tr>
<td>APOB</td>
<td>Hypobetalipoproteinemia</td>
<td>cod</td>
<td>0 (31–38)</td>
<td>...</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>49±2 (77±61)</td>
<td>73±23</td>
<td>5±2</td>
<td>61±26</td>
<td>−</td>
</tr>
<tr>
<td>apo B-100</td>
<td>Familial defective</td>
<td>dom</td>
<td>...</td>
<td>...</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>3337±2570 (257±85)</td>
<td>363±253</td>
<td>...</td>
<td>...</td>
<td>N</td>
</tr>
<tr>
<td>APOCII</td>
<td>APO C-Ⅱ deficiency</td>
<td>rec</td>
<td>0</td>
<td>...</td>
<td>+, ++</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>694±104</td>
<td>440±25</td>
<td>...</td>
<td>37±3</td>
<td>+</td>
</tr>
<tr>
<td>APOE</td>
<td>APOE deficiency</td>
<td>cod</td>
<td>0</td>
<td>...</td>
<td>+, ++</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>694±104</td>
<td>440±25</td>
<td>...</td>
<td>37±3</td>
<td>+</td>
</tr>
<tr>
<td>CBS</td>
<td>CBS deficiency</td>
<td>rec</td>
<td>...</td>
<td>0–40</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>84±25</td>
<td>−</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CETP</td>
<td>CETP deficiency</td>
<td>dom</td>
<td>0.6 (56)</td>
<td>0 (64)</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>101±51 (148±75)</td>
<td>378±170</td>
<td>625±160</td>
<td>34±10</td>
<td>+</td>
</tr>
<tr>
<td>LCAT</td>
<td>LCAT deficiency, fish eye disease</td>
<td>rec</td>
<td>2–43</td>
<td>0–9</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>304±247</td>
<td>19±10</td>
<td>11.3±2.7</td>
<td>(+)</td>
<td></td>
</tr>
<tr>
<td>LDLR</td>
<td>Familial hypercholesterolemia</td>
<td>dom</td>
<td>...</td>
<td>&lt;2–30</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>208±93</td>
<td>319±78</td>
<td>151±31</td>
<td>43±15</td>
<td>+</td>
</tr>
<tr>
<td>LIPA</td>
<td>Cholesterol ester storage disease</td>
<td>rec</td>
<td>...</td>
<td>0–36</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>2398±1595</td>
<td>304±247</td>
<td>19±10</td>
<td>11.3±2.7</td>
<td>(+)</td>
</tr>
<tr>
<td>LPL</td>
<td>LPL deficiency</td>
<td>rec</td>
<td>...</td>
<td>0 (20–62)</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>N, N</td>
<td>2398±1595</td>
<td>304±247</td>
<td>19±10</td>
<td>11.3±2.7</td>
<td>(+)</td>
</tr>
</tbody>
</table>

HLP indicates hyperlipoproteinemia; dom, dominant; cod, codominant; rec, recessive; TG, plasma triglycerides; TC, plasma total cholesterol; LDL-c, LDL-cholesterol; HDL-c, HDL-cholesterol; +, increased, and, in some cases, high levels; ++, very high levels; N, normal levels; −, low levels; −−, very low levels; and . . ., no information assessed.

*Lipid values are in mg/dL±1 SD for homozygous-affected individuals, and values in parentheses refer to heterozygotes.
### TABLE 2. Information Analysis of Splice-Site Variants

<table>
<thead>
<tr>
<th>Mutation No.</th>
<th>Gene (Accession)</th>
<th>Mutation, Coordinate</th>
<th>Primary Splice Site Coordinate</th>
<th>Secondary Splice Site Coordinate</th>
<th>Primary Splice Site Use (%)</th>
<th>Phenotype</th>
<th>Genotype‡</th>
<th>Activity (Mass)§</th>
<th>Dyslipidemia</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CETP (M32995)</td>
<td>IVS 10, donor, T→G, 290</td>
<td>289†</td>
<td>1254</td>
<td>8.9→8.9</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LPL (S71696)</td>
<td>IVS 1, donor, G→C, 14</td>
<td>10</td>
<td>3</td>
<td>9.7→3.0</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>LDLR (L00337)</td>
<td>IVS 3, donor, G→A, 139</td>
<td>Best, 139</td>
<td>Worst, 139</td>
<td>5.8→7.0</td>
<td>Reduced</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LDLR (L00337)</td>
<td>IVS 3, donor, G→T, 139</td>
<td>Best, 139</td>
<td>Worst, 139</td>
<td>5.8→2.0</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LDLR (L00338)</td>
<td>IVS 4, donor, T→C, 398</td>
<td>Best, 397</td>
<td>Worst, 397</td>
<td>5.2→2.2</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>LIPA (X75494)</td>
<td>IVS 6, acceptor, A→G, 394</td>
<td>395</td>
<td>13.5→5.1</td>
<td>+ (12)</td>
<td>Reduced</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LIPA (X75495)</td>
<td>IVS 8, donor, G→A, 416</td>
<td>417</td>
<td>8.8→5.7</td>
<td>(32)</td>
<td>4.1</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LPL (M76722)</td>
<td>IVS 6, acceptor, C→A, 196</td>
<td>198</td>
<td>6.7→2.5</td>
<td>(24)</td>
<td>2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>APOA-II (X04988)</td>
<td>IVS 3, donor, G→A, 1882</td>
<td>1882</td>
<td>9.5→3.3</td>
<td>&lt;0.3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>APOB (M19811)</td>
<td>IVS 5, donor, G→T, 411</td>
<td>411</td>
<td>9.0→1.2</td>
<td>...</td>
<td>+</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>APOC-II (M10612)</td>
<td>IVS 2, donor, G→C, 2995</td>
<td>2995</td>
<td>9.0→0.8</td>
<td>(0.1)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>LPL (S71696)</td>
<td>IVS 2, acceptor, G→A, 40</td>
<td>40</td>
<td>7.6→0.1</td>
<td>23</td>
<td>0</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>LDLR (L00343)</td>
<td>IVS 9, donor, G→A, 188</td>
<td>Best, 188</td>
<td>Worst, 188</td>
<td>11.3→1.5</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>LDLR (L00344)</td>
<td>IVS 10, donor, G→A, 248</td>
<td>Best, 244</td>
<td>Worst, 244</td>
<td>5.8→16.2</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>LDLR (L00345)</td>
<td>IVS 11, donor, G→T, 135</td>
<td>Best, 135</td>
<td>Worst, 135</td>
<td>8.4→0.6</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>LDLR (L00346)</td>
<td>IVS 12, donor, T→C, 157</td>
<td>Best, 156</td>
<td>Worst, 156</td>
<td>7.3→16.2</td>
<td>33</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>LPL (M76722)</td>
<td>IVS 6, acceptor, C→T, 196</td>
<td>198</td>
<td>6.7→5.1</td>
<td>(62)</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Experimentally verified loss of primary splice site function**

**Predicted loss of primary splice-site function**

**Experimentally verified cryptic splicing**

**Predicted cryptic splicing**
Table 2. Continued

<table>
<thead>
<tr>
<th>Mutation No.</th>
<th>Gene (Accession)</th>
<th>Mutation, Coordinate</th>
<th>Primary Splice Site Coordinate</th>
<th>Secondary Splice Site Coordinate</th>
<th>Primary Splice Site Use (% use†)</th>
<th>Genotype‡</th>
<th>Activity (Mass)$</th>
<th>Dyslipidemia§</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$R_{\text{natural}}$ $\rightarrow$ $R_{\text{mutant}}$</td>
<td>$R_{\text{natural}}$ $\rightarrow$ $R_{\text{mutant}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Predicted cryptic splicing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>LDLR (L00342)</td>
<td>Exon 8, donor, G→A, 141</td>
<td>Best, 142</td>
<td>7.0→4.0</td>
<td>6.5→3.4</td>
<td>Donor, 73</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
<tr>
<td>24</td>
<td>LDLR (L00350)</td>
<td>Exon 16, donor, G→A, 93</td>
<td>Worst, 142</td>
<td>7.7→4.6</td>
<td>7.2→4.1</td>
<td>Donor, 89</td>
<td>5.3→5.3</td>
<td>3.7→3.7</td>
<td>++</td>
</tr>
<tr>
<td><strong>Predicted unaltered splice-site function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>LCAT (X04981)</td>
<td>IVS 4, acceptor, T→C, 2327</td>
<td>2348</td>
<td>9.7→9.6</td>
<td></td>
<td>Donor, 13</td>
<td>0.6→0.6</td>
<td>0.3→0.3</td>
<td>++</td>
</tr>
<tr>
<td>26</td>
<td>LDLR (L00342)</td>
<td>Exon 7, acceptor, T→C, 19</td>
<td>Best, 26</td>
<td>19.2→19.2</td>
<td>6.4→6.4</td>
<td>Donor, 9</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
<tr>
<td>27</td>
<td>LDLR (L00338)</td>
<td>Exon 4, donor, G→A, 93</td>
<td>Worst, 26</td>
<td>5.2→7.3</td>
<td>4.7→6.7</td>
<td>Donor, 13</td>
<td>0.6→0.6</td>
<td>0.3→0.3</td>
<td>++</td>
</tr>
<tr>
<td>28</td>
<td>LDLR (L00340)</td>
<td>Exon 6, donor, G→A, 136</td>
<td>Best, 139</td>
<td>9.1→10.2</td>
<td>8.6→9.6</td>
<td>Donor, 9</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
<tr>
<td>29</td>
<td>LDLR (L00342)</td>
<td>Exon 8, acceptor, A→G, 27</td>
<td>Worst, 26</td>
<td>19.2→20.1</td>
<td>6.4→7.4</td>
<td>Donor, 9</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
<tr>
<td>30</td>
<td>LDLR (L00342)</td>
<td>Exon 8, acceptor, A→T, 27</td>
<td>Best, 26</td>
<td>19.2→17.9</td>
<td>6.4→5.2</td>
<td>Donor, 9</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
<tr>
<td>31</td>
<td>LPL (AF050163)</td>
<td>Exon 4, acceptor, G→A, 950</td>
<td>949</td>
<td>11.1→10.2</td>
<td></td>
<td>Donor, 9</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
<tr>
<td>32</td>
<td>LDLR (L00344)</td>
<td>IVS 9, acceptor, C→T, 6</td>
<td>35**</td>
<td>10.7→10.7</td>
<td>10.7</td>
<td>Donor, 39</td>
<td>3.4→3.4</td>
<td>3.0→3.0</td>
<td>++</td>
</tr>
</tbody>
</table>

*R$_{\text{natural}}$ and R$_{\text{mutant}}$ are R$_i$ values of the natural/primary and mutant splice sites.
†Splice-site use predicted to be leaky (+), unaltered (−), or none (−); for leaky mutations, calculated percent of correctly spliced mRNA is shown (in parentheses).
‡Genotype of mutation designated as homozygous (1), heterozygous (12), or compound heterozygous (23).
§Protein activity and mass are, respectively, the percent normal enzymatic activity/binding and the level of expression.
Dyslipidemia is graded as mild (††), average (1), severe (‡‡‡); phenotype not available (…).
Cryptic splicing verified in intron 13 (accession M32997, position 1254); R$_i$=8.9 bits.
Best and worst are the maximum and minimum possible R$_i$ values for incomplete sites.
**Additional cryptic sites with lower R$_i$ values are detected.
††Plasma triglyceride is +4.5SD; HDL-cholesterol and total cholesterol are, respectively, +4.5SD and +3.6SD above average.

Stringent criteria for dyslipidemic phenotypes were defined from reference mean values based on populations of individuals with mutations in the same gene. Each lipid parameter (total plasma triglycerides, cholesterol, and LDL and HDL cholesterol) with values within ±2SD of the respective reference population was defined as “average.” Individuals with significantly increased lipid levels (>2SD above the mean in the affected reference population) were defined as having “severe” dyslipidemia (eg, Table 2, mutations 1, 13, 16, 22, and 31 [mutation numbers throughout the article are those found in Table 2]). Those carrying mutations that decreased lipid levels >2SD below the average in the reference population (eg, mutation 21) were also designated as severe. Dyslipidemia was defined as “mild” in patients with mutations producing HDL-cholesterol levels >2SD above the average for the reference population (mutations 8 and 20).

Protein levels and/or activities were similarly graded relative to reference values for populations with mutations in the same gene. Protein levels and/or activities were defined as average if they were within the reported range of the percentage of normal for each respective reference population. Reduction of protein levels and/or activities below or above this reported range were graded as severe or mild, respectively.

Clinical severity of atherosclerosis was assessed using age-related expression of angina pectoris or myocardial infarction. The ages of onset of coronary symptoms for 2 or more unrelated patients with identical LDLR (mutation 3) and APOE (mutation 20) mutations were related to the corresponding age distributions for reference populations with type II (LDLR)$^{44}$ or type III (APOE)$^{45}$ familial hyperlipoproteinemia. Clinical signs of atherosclerosis were graded as average for ages of onset within 2SD of the mean age for the reference populations; mild or severe atherosclerosis occurred with average ages of onset above or below the 2SD bound, respectively.

Phenotypic analyses were limited to patients with fasting plasma lipid and homocysteine levels. Because outlier phenotypes were derived using stringent statistical criteria, some patients originally described as severely or mildly affected were categorized as having...
Criteria for Predicting Splice-Site Function

The effects of nucleotide substitutions were predicted from $R_i$ values based on the following, previously validated criteria:

1. Substitutions resulting in $R_i$ $<$ 2.4 bits would completely inactivate primary splice sites and result in severely affected phenotypes.

2. Substitutions with decreased $R_i$ values that were $\geq$ 2.4 bits would reduce splicing, thereby producing a milder phenotype. The residual amount of correctly spliced mRNA at leaky sites was computed as the minimum fold change in binding affinity, $2^{\Delta R}$ ($\Delta R$ is the difference between the $R_i$ value of the natural site and that of the variant site). The result was expressed as the maximum percent of normal mRNA.

3. Substitutions would activate cryptic splice sites with $R_i$ values exceeding that of the adjacent natural site. Observations of exon skipping or cryptic splicing were used to validate $R_i$-based predictions whenever these data were available.

4. Nucleotide changes that did not significantly alter $R_i$ would have no effect on splicing. The significance of the change in $R_i$ ($\Delta R_i$) was assessed with a paired $t$-test using a cutoff at 5% significance.

Results

Prediction and Validation of Splicing Defects

We predicted either complete loss of primary splice site function (mutations 1 to 5, 9 to 16, 18, 19, 21, and 22) or leaky splicing (mutations 6 to 8, 17, 20, 23, and 24) on the basis of reductions in $R_i$ values of the corresponding mutant sites. We predicted that mutations 18 to 24 would activate cryptic splice sites. mRNA analyses corroborated these predictions for primary splice sites (mutations 1 to 8 and 18 to 20) and activated cryptic splice sites (mutations 1, 18 to 20, and 32). The $R_i$ value of a polymorphic intron variant within an acceptor site was unchanged (mutation 26), which was expected on the basis of previous studies of polymorphic splice sites.

Of the 258 missense mutations analyzed, we only predicted 2 (0.8%) would decrease $R_i$ values and affect splicing at adjacent donor sites (mutations 23 and 24). Five missense substitutions produced insignificant changes in the $R_i$ values in adjacent splice sites (mutations 27 to 31); the remainder had no detectable effect on splicing. Despite the fact that the splice-site sequences for the $LDLR$ and $CBS$ genes were incomplete, the $R_i$ analyses were interpretable for 165 of the 171 missense changes adjacent to these sites.

Splicing mutations at the presumed branchpoint sequence reduced (mutation 32) or completely abolished (mutation 25) the use of the corresponding acceptor site but did not alter the information content. This was not surprising because this signal is not detectable by information analysis or by other methods. However, one of these branchpoint mutations was also predicted to activate cryptic splicing (mutation 32).

Relationship of $R_i$ to Phenotype

To determine whether disease phenotype was related to predicted splice-site use, the degrees of dyslipidemia, protein expression, and clinical severity were graded and compared with the $R_i$ values of mutant splice sites. Severe (mutations 1, 13, 16, 21, and 22) and average (mutations 2 to 4, 9 to 12, 14, 18, and 19) dyslipidemia were present in individuals with primary splice-site mutations with $R_i$ values $<$ 2.4 bits, but dyslipidemia was not present in individuals with mutant sites $\geq$ 2.4 bits. Conversely, mutant sites with $R_i$ values $\geq$ 2.4 bits were found in patients with mild (mutations 8 and 20) or average (mutations 7, 23, and 24) dyslipidemia but not in those with severe dyslipidemia.

Average (mutations 2, 9, 11, 16, 19, 21, and 22), severe (mutation 1), and mild (mutation 18, which is borderline average) decreases in protein levels or enzymatic activity were found in individuals harboring mutant splice sites $<$ 2.4 bits (Table 3). Mutations that resulted in sites with $R_i$ $\geq$ 2.4 bits, however, exhibited only mild (mutations 8 and 20) or average (mutation 7) reductions in protein expression.

On the basis of age at onset of coronary symptoms, the severity of clinical atherosclerosis for one patient who had $R_i$ $<$ 2.4 bits was graded as average (mutation 3); it was graded mild for another individual carrying a mutation with $R_i$ $\geq$ 2.4 bits (mutation 20). Corresponding clinical data were not available for the other mutations.

Discussion

The decrease in information content at mutant splice sites in atherosclerosis candidate genes is related to diminished splice-site use. Predicted reductions in the proportions of normal mRNA were confirmed for 11 splice-site mutations. The decrease in $R_i$ also tended to be related to the degree of reduced protein expression and to phenotypic severity (Table 3).

Missense mutations can, in some cases, simultaneously affect both translation and splicing, and the splicing effects can be detected by information analysis. We found 2 missense changes within codons adjacent to splice junctions (mutations 23 and 24) that were predicted to impair splicing. By contrast, 5 other similarly located missense mutations slightly increased (mutations 27 to 29) or decreased (mutations 30 and 31) $R_i$ values without affecting splicing.

Incomplete genomic sequences for splice sites in the $LDLR$ and $CBS$ genes were analyzed by comparing the ranges of possible $R_i$ values for the corresponding natural and variant sites. The results were interpretable for all except 6 mutations in which the $R_i$ intervals of natural and variant acceptor sites overlapped. Thus, the 5’ and 3’ terminal intronic sequences (positions $-$16 to $-$25 of acceptor and $+$6 of donor sites) can sometimes make a significant contribution to the overall information content of a
splice site. The decrease in R, values for splicing mutations was related to the severity of dyslipidemia (Table 3). Splicing mutations that resulted in milder dyslipidemia were predicted to be leaky (≥2.4 bits), whereas severely affected individuals carried mutations that presumably inactivated splice sites and abolished protein expression (<2.4 bits). The phenotypes at the outlying lipid values were concordant with these predictions, although corresponding R, values for mutations in individuals with average dyslipidemia were not predictive. More refined models, incorporating lipid phenotypes as quantitative traits, will require analysis of additional patients and mutations.

Acknowledgments

We thank Drs Thomas D. Schneider and Christoph A. Nienaber for their suggestions. Support was provided by grants from the Public Health Service (CA74683), the American Cancer Society (DHP-132), Merck Genome Research Foundation (to P.K.R.), and Deutsche Forschungsgemeinschaft (KO 1828/1-2; to Y.v.K).

References


Splice-Site Mutations in Atherosclerosis Candidate Genes: Relating Individual Information to Phenotype
Yskert von Kodolitsch, Reed E. Pyeritz and Peter K. Rogan

Circulation. 1999;100:693-699
doi: 10.1161/01.CIR.100.7.693

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/100/7/693

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org/subscriptions/