Structural Organization of the Human *Dihydropyrimidine Dehydrogenase* Gene

Martin R. Johnson, Kangshang Wang, Silke Tillmanns, Nicolas Albin, and Robert B. Diasio

Department of Pharmacology and Toxicology, University of Alabama, Birmingham, Alabama 35294

Abstract

Deficiency of the pyrimidine catabolic enzyme, dihydropyrimidine dehydrogenase (DPD), has been shown to be responsible for a pharmacogenetic syndrome in which administration of 5-fluorouracil is associated with severe and potentially life-threatening toxicity. Following the recent availability of the cDNA for DPD, there were initial reports of several molecular defects (point mutations, deletions due to exon skipping) that were suggested as a potential molecular basis for DPD deficiency, even before the complete physical structure of the *DPD* gene was known. To understand the mechanism responsible for DPD deficiency, we have determined the genomic structure and organization of the human *DPD* gene.

The gene is approximately 150 kb in length, and it consists of 23 exons, ranging in size from 69 to 1404 bp. The sequences of intronic regions flanking the exon boundaries have been determined. The physical map of the *DPD* gene should permit development of rapid assays to detect point mutations or small deletions in the *DPD* gene associated with 5-fluorouracil toxicity.

Introduction

5-FUra, synthesized in 1957 and studied extensively by Heidelberger et al. (1), today continues to be widely used in the management of several common malignancies, including cancers of the colon, breast, and skin. The anticancer effects and toxicity of 5-FUra are directly related to anabolism of 5-FUra into its nucleotides, which can then produce cytotoxicity through inhibition of thymidylate synthase activity or incorporation into RNA and/or DNA (2).

In humans, more than 85% of administered 5-FUra is degraded through the catabolic pathway (3). DPD (EC 1.3.1.2; also known as dihydropyrimidine dehydrogenase, dihydroliririmidine dehydrogenase, and uracil reductase) catalyzes the initial and rate-limiting step in pyrimidine catabolism, the reduction of uracil or thymine to 5,6-dihydropyrimidinol or 5,6-dihydropyrimidine (4). Thus, DPD regulates the availability of 5-FUra for anabolism, thereby affecting its pharmacokinetics, toxicity, and efficacy (5).

Several recent studies have described a pharmacogenetic disorder in which individuals with absent or significantly decreased DPD activity develop life-threatening toxicity following exposure to 5-FUra (6–10). Administration of standard doses of 5-FUra to DPD-deficient patients has resulted in severe life-threatening toxicity, including mucositis, granulo-cytopenia, neuropathy, and even death. Since the initial reports several years ago, there have been an increasing number of cases reported suggesting that this disorder may be more frequent than initially thought (7, 11). Population studies in 124 healthy volunteers (7) and 185 cancer patients (11) have demonstrated that approximately 3% of the individuals tested were partially DPD deficient, having enzyme activity below the 95% distribution range of DPD enzyme activity (below 0.064 nmol/min/mg). The recent availability of the DPD cDNA (12, 13), has allowed investigators to examine, at the molecular level, patients who were previously phenotyped (by enzyme assay) as DPD deficient. These studies have demonstrated polymorphisms (14) and exon skipping (15, 16) as possible bases for this pharmacogenetic syndrome.

These observations have raised the possibility that there are multiple causes for DPD deficiency. An understanding of the structure and organization of the *DPD* gene should enhance our ability to identify mutations and alternatively spliced regions of this gene. Here, we report the genomic structure and organization for the human *DPD* gene.

Materials and Methods

Library Screening. Clones representing the human genomic locus of the *DPD* gene were isolated from a titered P1 human genomic library (Genome System Inc., St. Louis, MO). The library was screened using PCR and the following primers: sense, TAGGAAACAGCTAGCTACCTTG, and antisense, CTGGTACGCAAGATCTACAGCTATG. Specific cycling conditions were evaluated prior to screening the P1 library using purified genomic DNA. The conditions used to screen the library were 10 μM each primer, 60 mM Tris-HCl (pH 8.5), 15 mM (NH₄)₂SO₄, 2.5 mM MgCl₂, 2.5 mM each dNTP, and 1 unit of AmpliTaq DNA polymerase in a total volume of 50 μl. The samples were amplified in an MJ model PTC-100 thermal cycler (MJ Research, Inc., Watertown, MA) programmed for a temperature-step cycle of 1 min at 95°C, 1 min at 58°C, and 1 min at 72°C. This cycle was repeated for a total of 30 cycles.

Identification of Exon-Intron Boundaries. Approximately 10% of the exon-intron boundaries were identified by direct cycle sequencing (New England Biolabs) of the P1 genomic clone (17) using primers designed from DPD cDNA. The remaining boundaries were determined using suppression PCR, as follows (Clontech Laboratories, Inc., Palo Alto, CA). Genomic DNA was digested with EcoRV, ScaI, DraI, Pvull, and SspI separately. An adaptor was ligated to the ends of the DNA fragments, and a small aliquot of each digested-ligated DNA was used as a template for PCR using adaptor primers and gene-specific primers designed from DPD cDNA (18). Amplified products were subcloned into the pCRII vector (Invitrogen) and sequenced using the universal T7 and M13 primers bordering the multiple cloning site. DNA sequencing was accomplished using the dyeoxy chain termination procedure (Sequenase Version 2, United States Biochemical Corp.). Exon-intron boundaries were identified by the presence of consensus splice junctions at sites where the sequence of the genomic product differed from the published DPD cDNA sequence. The computer programs MacVector and AssemblyLIGN were used for comparison of DNA sequences.

Sizes of introns were determined by exon-exon PCR amplification products using LA PCR (Takara Shuzo Co., Kyoto, Japan) with a buffer supplied by the manufacturer. The final reaction volume was 50 μl and contained 10 μM each primer and 50 ng of genomic DNA as a template. Cycling conditions were as follows: denaturation at 98°C for 20 s and annealing and extension at 68°C for 15 min. Following the first 14 cycles, an autosegment extension of 15 s per
cycle (at the annealing and extension step temperature of 68°C) was added for the remaining 16 cycles (19, 20). Most runs included an initial 15-min hold at 94°C and a final 10-min hold at 72°C. All samples were run using a manual "hot start" technique in which Mg2+ was withheld until samples had been incubated at 94°C. A positive control that amplified a region of the human β-globin cluster (21.5 kb) was run with each reaction and used the following primers: sense, ACATGATFAGCAAAAGGGCCTAGCTFGGACTCAGA, and antisense, TGCACCTGCTCFGTGAUATGACTATCCCACAGTC.

Primer Extension. Oligonucleotide (CCAGTGACAAACCCTCCTAGG) was phosphorylated with [γ-32P]ATP and hybridized to 30 μg of total RNA prepared from freshly isolated peripheral blood mononuclear cells (7). Reverse transcription was performed in 16.7 mM KC1, 13.3 mM MgCl2, 23.3 mM Tris-HCl (pH 8.3), 0.33 mM dNTPs, and 0.133 mg/ml actinomycin D with 200 units of Moloney murine leukemia virus RNaseH and an "adaptor primer" was then ligated to both ends of the double-stranded cDNA using T4 DNA ligase. PCR amplification proceeded in 10 mM Tris-HCl (pH 8.3), 1 mM EDTA, and 250 mM KCl at 60°C for 1 h, followed by incubation at room temperature for 1.5 h. Reverse transcription was performed in 16.7 mM KCl, 13.3 mM MgCl2, 23.3 mM Tris-HCl (pH 8.3), 13.3 mM DTT, 0.33 mM dNTPs, and 0.133 mg/ml actinomycin D with 200 units of Moloney murine leukemia virus RNaseH reverse transcriptase (Life Technologies, Inc.). The extension products were fractionated on an 8.0 M urea-6% polyacrylamide gel and analyzed by autoradiography (21).

5'-RACE. Total RNA was prepared from freshly isolated peripheral blood mononuclear cells as described above. 5'-RACE was performed according to the manufacturer's instructions (Clontech). Briefly, cDNA was generated using a modified lock-docking oligo(dT) primer that contained two degenerate nucleotides at the 3' end (22). Second-strand synthesis was performed using a combination of Escherichia coli DNA polymerase I, RNase H, and E. coli DNA ligase provided by the manufacturer. T4 DNA polymerase was then used to create blunt ends on the double-stranded cDNA. An "adaptor primer" was then ligated to both ends of the double-stranded cDNA using T4 DNA ligase. PCR was then performed using an "anchor primer" complementary to the adaptor primer and a specific DPD cDNA primer located in exon 2. The amplified products were purified, subcloned, and sequenced as described above.

Results and Discussion

In the initial phase of this study, we screened a tittered P1 human genomic library with specific primers (designed within exon 22 of the DPD DNA sequence). Clones representing the human genomic locus of the DPD gene were identified by amplification of a 135-bp product. Three independent P1 clones were isolated, and the boundaries for exons 21–23 were identified by direct thermal cycle sequencing using exon-specific primers. Unfortunately, PCR and hybridization analysis of the three P1 clones demonstrated that they did not contain exons 1–20. The relatively large size of the introns isolated for the DPD gene from the P1 screening (23), along with previous reports describing difficulties in screening phage libraries (15), suggested that the average insert size of most commercial λ phage libraries (14–20 kb) would be too small to be useful for this particular gene. We therefore elected to use a recently described technique known as suppression PCR (18, 24, 25). This method had a particular advantage over traditional library screening with regard to the DPD gene in that only the boundary of each splice site junction is amplified from genomic DNA and subcloned. Following amplification of each splice site, the PCR products were purified on low melting point agarose and subcloned into the pcRll vector for sequence analysis.

Previous studies have mapped the DPD gene to chromosome 1p22 (26). We now report the organization and structure of the DPD gene, which consists of 23 exons spanning a region approximately 150 kb in length. The size and sequence of intronic regions flanking the exon boundaries has also been determined and is summarized in Table 1. All intron-exon boundaries were found to conform to the canonical GT-AG rule. A physical map encompassing the entire coding region of the DPD gene is shown in Fig. 1. Exons are numbered 1–23, with exon 15 (69 bp) being the smallest and exon 23 (1404 bp) being the largest. The translation start site (ATG) was located in exon 1. Exon 23 contained 168 nucleotides of coding sequence, followed by the translation stop codon (TAA) and a 1236-nucleotide 3' untranslated region. The human liver DPD cDNA sequence published by Yokota et al. (12) did not contain the complete 3' untranslated region. Sub-

---

### Table 1 Intron-exon boundaries of the human DPD gene

<table>
<thead>
<tr>
<th>Exon</th>
<th>Exon size (bp)</th>
<th>3' intron splice site</th>
<th>Exon sequence (5' to 3')</th>
<th>5' intron splice site</th>
<th>Intron length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140 ±a</td>
<td>...</td>
<td>GTCCAGAG</td>
<td>CTATGGCT</td>
<td>3,554</td>
</tr>
<tr>
<td>2</td>
<td>111</td>
<td>...</td>
<td>CTCTTTTT</td>
<td>CTATGGCT</td>
<td>10,610</td>
</tr>
<tr>
<td>3</td>
<td>83</td>
<td>...</td>
<td>CTTTGTGT</td>
<td>AAAGCAAG</td>
<td>5,055</td>
</tr>
<tr>
<td>4</td>
<td>88</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>AAAGCAAG</td>
<td>12,613</td>
</tr>
<tr>
<td>5</td>
<td>166</td>
<td>...</td>
<td>ACTATTG</td>
<td>TTGGCAAG</td>
<td>&gt;20,000</td>
</tr>
<tr>
<td>6</td>
<td>191</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>7,762</td>
</tr>
<tr>
<td>7</td>
<td>82</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>5,643</td>
</tr>
<tr>
<td>8</td>
<td>88</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>8,414</td>
</tr>
<tr>
<td>9</td>
<td>108</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>2,236</td>
</tr>
<tr>
<td>10</td>
<td>170</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>&gt;20,000</td>
</tr>
<tr>
<td>11</td>
<td>211</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>4,610</td>
</tr>
<tr>
<td>12</td>
<td>185</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>5,296</td>
</tr>
<tr>
<td>13</td>
<td>216</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>4,323</td>
</tr>
<tr>
<td>14</td>
<td>165</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>4,972</td>
</tr>
<tr>
<td>15</td>
<td>69</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>9,080</td>
</tr>
<tr>
<td>16</td>
<td>91</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>6,902</td>
</tr>
<tr>
<td>17</td>
<td>114</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>843</td>
</tr>
<tr>
<td>18</td>
<td>121</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>2,948</td>
</tr>
<tr>
<td>19</td>
<td>142</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>4,972</td>
</tr>
<tr>
<td>20</td>
<td>180</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>4,894</td>
</tr>
<tr>
<td>21</td>
<td>144</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>5,107</td>
</tr>
<tr>
<td>22</td>
<td>141</td>
<td>...</td>
<td>TTGGCAAG</td>
<td>TGAGGAAG</td>
<td>2,753</td>
</tr>
</tbody>
</table>

---

a Transcription start site not determined.
sequent studies in our laboratory identified the complete human lymphocytic DPD cDNA sequence (GenBank accession no. U20938), which contained this region and was used in our comparison of genomic and cDNA sequence searching for consensus splice junctions where the sequence of the genomic product differed from the lymphocytic DPD cDNA sequence.

Primer extension experiments were undertaken to determine the precise size of exon 1. Freshly isolated RNA from peripheral blood mononuclear cells was used for this analysis. The band indicated by the arrow (Fig. 2) is the reverse-transcribed product, whereas the other bands represent known DNA sequences that were included for determination of molecular size. Comparison of the nucleotides in the known sequence with the mobility of the DPD mRNA lane demonstrates that the distance from the primer to the 5' end of the message coding for DPD is 104 nucleotides. This result suggested that there were approximately 20–25 additional nucleotides in the mRNA sequence that had not been identified previously by cDNA cloning (12, 13). We used 5'-RACE to clone and sequence these additional nucleotides located upstream from the published DPD cDNA sequence. PCR with an anchor primer and an internally nested primer located 193 bp from the 5' end of the published human DPD cDNA sequence resulted in the amplification of a cDNA approximately 214 bp in length. Following subcloning, 10 clones were selected and sequenced. All of the inserts examined contained the human DPD cDNA sequence reported previously; however, 8 of the 10 clones extended further upstream for an additional 21 nucleotides. These data suggest that exon 1 is 140 nucleotides long and contains 101 nucleotides of 5' untranslated region, followed by the initiating methionine and 36 nucleotides of coding sequence.

Intron sizes were determined by PCR amplification using genomic DNA as a template with gene-specific primer pairs from adjacent exons. The use of nested PCR allowed verification of the sizes of the introns. Only PCRs that gave a single well-defined product (Fig. 3) were used for estimation of intron sizes. Introns 5 and 10 were the largest, with sizes greater than 20 kb, whereas introns 9 and 17 were the smallest.

The recent examination of two unrelated DPD-deficient patients has suggested exon skipping (15, 16) as a possible mechanism for DPD deficiency. These studies demonstrated a 165-bp deletion in the mRNA of both patients. Genomic DNA analysis of these patients revealed a G to A mutation in the GT 5' splicing recognition sequence of the intron preceding the mutation. It is now clear from the current study that exon 14 represents the skipped exon and that the G to A point mutation in the invariant GT splice donor site is located at the 5' end of intron 14. However, our examination of this region of the DPD gene in two additional well-characterized DPD-deficient patients has demonstrated that neither patient exhibits this particular G to A mutation.

The aim of this study was to determine the structure and organization of the DPD gene. These data could then be used as a foundation for rapidly scanning the DPD gene for mutations in genomic DNAs in an attempt to define the molecular basis of DPD enzyme deficiency. The data presented in this study should allow the development of genetic assays that could clarify the role of the DPD gene in deficient patients. The clinical diagnosis of DPD deficiency is difficult because the appearance of life-threatening toxicity secondary to exposure to FUra is the first symptom of this pharmacogenetic syndrome. Until now, only phenotypic DPD enzyme assays have been available to diagnose patients as DPD deficient. These assays are expensive and labor intensive and are not available in most cancer treatment centers. Thus, there is a great need for a more direct diagnostic method,
preferably available to the patient prior to receiving FUrA. The character-
ization of the genomic structure of the DPD gene provides the basis for the development of simple genetic tests. To date, all DPD-
deficient patients have been analyzed without a complete understand-
ing of the DPD gene structure and organization. The stability of DNA in comparison to RNA along with the ease with which DNA samples may be extracted and purified represents a clear diagnostic advantage. In addition, the development of genomic DNA based assays would allow analysis of archive paraffin-embedded samples from deceased DPD-deficient patients. These samples would allow researchers to build pedigrees on families (including determination of heterozy-
gotes), who, until now, were impossible to study. Although the size and complexity of the DPD gene makes scanning the gene for mutations a challenge, we believe that recent advances in the fields of molecular biology (such as single-stranded conformational polymor-
phism) may be used for identifying additional mutations in the DPD gene.

References
5. Diasio, R. B., and Lu, Z-H. Dihydropyrimidine dehydrogenase activity and fluorou-
6. Diasio, R. B., Beavers, T. L., and Carpenter, J. T. Familial deficiency of dihydro-
7. Lu, Z. Zhang, R., and Diasio, R. B. Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients and clinical implication in 5-fluorouracil chemother-
8. Harris, B. E., Carpenter, J. T., and Diasio, R. B. Severe 5-fluorouracil toxicity secondary to dihydropyrimidine dehydrogenase deficiency: a potentially more com-
10. Takimoto, C. H., Lu, Z. H., Zhang, R., Liang, M. D., Larson, L. V., Cantilena, L. R., Grem, J. L., Allegro, C. J., Diasio, R. B., and Chu, E. Severe neurotoxicity following 5-fluorouracil-based chemotherapy in a patient with dihydropyrimidine dehydrogen-
11. Milano, G., and Etienne, M. C. Potential importance of dihydropyrimidine dehy-
13. Johnson, M. R., Albin, N., Wang, K., and Diasio, R. B. cDNA cloning of dihydro-
15. Meinsma, R., Salguero, P. F., Van Kuilenburg, A. B. P., Van Gennip, A. H., and 
Gonzalez, F. J. Human polymorphism in drug metabolism: mutation in the dihydro-
17. Pierce, J. C., Sternber, N., and Sauer, B. A mouse genomic library in the bacterioph-
24. Maiti, S., Doskow, J., Li, S., Nhim, R. P., Lindsey, J. S., and Wilkinson, M. F. The 
GFP homeobox gene. Androgen-dependent and -independent promoters and tissue-
Structural Organization of the Human Dihydropyrimidine Dehydrogenase Gene

Martin R. Johnson, Kangshang Wang, Silke Tillmanns, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/57/9/1660

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.