INTRODUCTORY REMARKS ON HUMAN CYPs

David R. Nelson*
Department of Biochemistry, Tennessee Health Science Center,
University of Tennessee, Suite G01, 858 Madison Avenue,
Memphis, TN 38163

This issue of Drug Metabolism Reviews is centered on the human P450 metabolism database compiled by Slobodan Rendic. With more than 3900 entries and over 1000 references, this database is a resource for studying the enzymology of the 57 known human P450 proteins. Presently, the database has entries on 42 of these P450s. Eleven of the enzymes have more than 100 entries each and CYP3A4 has 920 entries. The database has information related to substrates, inhibitors, and inducers. It is searchable by the drug name and the type of reaction catalyzed is given. This is one of several databases devoted to cytochrome P450, but it is unique in its emphasis on enzymology and metabolism. When future meta-databases are created by merger and hyper linking, this one will surely form a significant part of any online P450 encyclopedia.

The set of human P450 genes and proteins is finite. At present there are 57 named P450s in humans (1) and this number is not expected to vary much in the future. The sequence of the human genome is nearly complete. The most recent figures at NCBI from Feb 11, 2001 show 32.5% finished, 61% draft, and 6.5% not represented yet in the genomic sequence data (2). These figures are 4 months old and they are certainly closer to completion now. It is satisfying to have a complete inventory of human P450 genes, but it must be understood that this is only a catalog. The more interesting properties of the P450s are their structure/function relationships and the phenotypes they produce when deleted, inhibited, or mutated. This information only comes through experimental work and it is hard to obtain.

Fortunately, experimental work on the human P450s has been accumulating for several decades. The extensive Table 1 of Rendic (3) contains over 3900

____________________
*E-mail: dnelson@utmem.edu
entries listing the results of this work. The number of entries per P450 is shown in Table 1. Currently, 42 of the 57 human P450s are represented in this table. Most of the 15 missing P450s (CYP2A7 (4), 2R1, 2S1 (5), 2U1, 2W1, 4A20 (or 4Z1), 4A22 (6), 4F11 (7), 4F22, 4V2, 4X1, 20, 26B1 (8,9), 26C1, 27C1) have been described only in the last year or two. These are largely known only through their sequences (1) and almost nothing is available yet on their function. This will change as each of these genes are expressed and tested for activity on a number of candidate substrates.

There are some striking P450s that are highly represented in the table. These are the experimental greatest hits of P450 mediated drug metabolism. Each of these eleven P450s has over 100 entries in Rendic’s Table 1 “Summary of Information on Human CYP Enzymes: Human P450 Metabolism Data”, in this issue. Cyp3A4 has 920 entries! More detailed information on these is given in his Tables 5, 6, and 7. The value of Rendic’s Table 1 will satisfy two different groups. Those working on the top 11 P450s will be aided by the gathering of the abundant literature references for their favorite enzymes into one place, along with the readily accessible information on substrates, inhibitors, and inducers of these enzymes. The searchable web based form of the table will be helpful for those looking for a particular drug and what P450s act on it or are inhibited or induced by it. The second group of P450 researchers will be those interested in the P450s with relatively few entries in the table. Some P450s have only one or two entries.

Table 1. Entries in Rendic’s Table 1 for 57 Human P450s

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1A1</td>
<td>239</td>
<td>3A4</td>
<td>920</td>
<td>8A1</td>
<td>6</td>
</tr>
<tr>
<td>1A2</td>
<td>420</td>
<td>3A5</td>
<td>35</td>
<td>8B1</td>
<td>1</td>
</tr>
<tr>
<td>1B1</td>
<td>110</td>
<td>3A7</td>
<td>37</td>
<td>11A1</td>
<td>5</td>
</tr>
<tr>
<td>2A6</td>
<td>134</td>
<td>3A43</td>
<td>1</td>
<td>11B1</td>
<td>16</td>
</tr>
<tr>
<td>2A7</td>
<td>0</td>
<td>4A11</td>
<td>11</td>
<td>11B2</td>
<td>9</td>
</tr>
<tr>
<td>2A13</td>
<td>5</td>
<td>4A20</td>
<td>0</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>2B6</td>
<td>171</td>
<td>4A22</td>
<td>0</td>
<td>19</td>
<td>86</td>
</tr>
<tr>
<td>2C8</td>
<td>125</td>
<td>4B1</td>
<td>12</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>2C9</td>
<td>332</td>
<td>4F2</td>
<td>13</td>
<td>21</td>
<td>4</td>
</tr>
<tr>
<td>2C18</td>
<td>32</td>
<td>4F3</td>
<td>15</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>2C19</td>
<td>231</td>
<td>4F8</td>
<td>6</td>
<td>26A1</td>
<td>3</td>
</tr>
<tr>
<td>2D6</td>
<td>415</td>
<td>4F11</td>
<td>0</td>
<td>26B1</td>
<td>0</td>
</tr>
<tr>
<td>2E1</td>
<td>218</td>
<td>4F12</td>
<td>8</td>
<td>26C1</td>
<td>0</td>
</tr>
<tr>
<td>2F1</td>
<td>8</td>
<td>4F22</td>
<td>0</td>
<td>27A1</td>
<td>16</td>
</tr>
<tr>
<td>2J2</td>
<td>5</td>
<td>4V2</td>
<td>0</td>
<td>27B1</td>
<td>2</td>
</tr>
<tr>
<td>2R1</td>
<td>0</td>
<td>4X1</td>
<td>0</td>
<td>27C1</td>
<td>0</td>
</tr>
<tr>
<td>2S1</td>
<td>0</td>
<td>5A1</td>
<td>13</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td>2U1</td>
<td>0</td>
<td>7A1</td>
<td>7</td>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td>2W1</td>
<td>0</td>
<td>7B1</td>
<td>4</td>
<td>51</td>
<td>8</td>
</tr>
</tbody>
</table>
INTRODUCTION TO HUMAN CYPs

with a similar number of references. These will be quick to locate and this will be helpful to those who are interested in the less abundantly studied P450 enzymes.

The table will also serve as a kind of progress meter for human P450 characterization. The 15 P450s that do not have entries in the table today should dwindle over time towards zero as the new P450s get assigned functions. The top 11 are heavily weighted towards hepatic enzymes, while many of the lesser characterized P450s are non-hepatic. This emphasis on hepatic P450s will probably change as the importance of the lesser enzymes becomes known. Comparative genomics is already identifying highly conserved vertebrate P450s that are candidates for some significant, if unknown functions.

The CYP2R1 has clear orthologs in Fugu (84% over 69 amino acids from EXXR to the PERF motif in LPC25839 at the Fugu blast server (10), not in Genbank) and in Xenopus (BF024911 79% over the last 61 amino acids). These regions of high identity make CYP2R1 stand out in this category. The CYP4V2 is another highly conserved P450 found in trout (CYP4V1 AF046012 69% to 4V2 from the I-helix to the heme), Xenopus (CYP4V4 no accession number, 58% to 4V2) and even in sea urchin (AZ191715, AZ198827, AZ147906, AZ176297). The AZ176297 is 72% identical with mouse Cyp4v3 over 37 amino acids between the PERF motif and the heme signature region. The CYP2U1 matches a Fugu sequence from scaffold_10678 (10) with 70% identity over 54 amino acids between the PKG motif and the heme-signature region. As these sequences are conserved, so the functions will probably be conserved. These P450s would be predicted to have important endogenous substrates that have been maintained over vertebrate and even deuterostome evolution in the case of the sea urchin 4V related sequences.

The tendency in recent years has been to adopt a new name for any research area that is broad and covers a whole organism’s repertoire. This has led from the reasonable genome, to proteome, metabolome (all metabolites), transcriptome (all transcripts), and recently interactome (for all two hybrid protein interactions). These are concepts that are now being filled in for the model organisms. The awareness of the importance of all these -omes, that are really just big databases, has been to increase the number and interconnectedness of these databases. Each new database builds in links to the older more established databases. A key to this process is nomenclature. A unified nomenclature provides the means to connect these databases. Rendic’s Table 1 is keyed to the standardized CYP nomenclature, which makes finding information about a human P450 rapid and convenient. It might seem that the process of a unifying nomenclature is implemented everywhere now, but that is not completely true. The automated programs that sort ESTs in Unigene often cannot distinguish closely related sequences from one another, so they are lumped together. Enzymes that have other names, like thromboxane synthase (CYP5A) or prostacyclin synthase (CYP8A) are often not linked with their CYP names. Researchers still publish papers on P450s without the CYP name in the title or abstract, making later retrieval from public databases like PubMed difficult. Databases like Rendic’s Table 1, that are based on the CYP
nomenclature, reduce these problems and enhance access to useful information that might be hard to find otherwise.

The beauty of databases on the web is their ability to provide detailed information on specific topics. No single database will cover all the aspects of a given subject. That is why multiple databases that focus on limited areas are highly valuable. Rendic’s database (11) is committed to human cytochrome P450 metabolism and enzymology. The human P450 allele database (12) is focused on polymorphisms in human P450s. My own P450 homepage is dedicated to nomenclature for all P450s and evolution of the P450 superfamily (13). The *Drosophila* P450s are the subject of a special database at Rene Feyereisen’s lab in France (14) and *Arabidopsis* P450s are covered at the Center for Molecular Plant Physiology in Denmark (15). The Cytochrome P450 Database in Moscow (16) is a more global effort, linking the nomenclature with sequence data, biophysical, and structural properties. All of these databases form a kind of mesh or P450 network. I have links to all of them and others on my homepage under the label Cytochrome P450 Web Matrix (17). Other databases that might be anticipated are a P450 structure and modeling database, showing known structures and models of related P450s based on the most suitable known crystal structure. This might be linked to a comprehensive P450 mutation database, mapping mutations to structural models and listing phenotypes of the mutations. Such databases will require individuals dedicated to creating and maintaining the hundreds to thousands of files needed to present such data. There will be a corresponding need to fund these projects that have been lacking in the past.

Our justifiable interest in humans has created the wealth of information indexed in Rendic’s Table 1. However, another use of this data will be in understanding other organisms. Especially other mammals for which genome sequences will become available, like the mouse and the rat. There are already 82 P450 sequences known in the mouse (18) and more are expected. The obvious place to go to predict enzymatic properties of many of these enzymes will be Table 1. This can work both ways since many P450 enzymes have been characterized in rodents first. One obvious possibility is to create a sister database for mouse and rat and for each model organism. Clearly, this is beyond the ability of one individual, but the pattern has been set. I can envision a database like GenBank in format. Each entry would have fields, more detailed than the present GenBank format. One field would offer the gene sequence with intron–exon boundaries mapped out and the protein sequence given, with alternative splicing indicated. Another field would offer sequence alignments with multiple choices, most closely related sequences, mammalian P450 sequences, P450s from individual species. Another field would be for substrates, inhibitors, inducers, and enzymatic properties, much like the information in Table 1. Then there would be crystal structures, polymorphic variants, genome map locations, syntenic regions in other genomes, etc. This is the direction we are moving. It will require massive integration of data from numerous individual databases and some uniformity in structure that would permit linkage between databases. Some of these databases are already here.
Rendic’s Table 1 is one of them. More are sure to come and the existing ones will need to be expanded to become more complete. The process is not unlike the development of a brain, with more and more connections made, the ability to understand increases.

REFERENCES
