The Directory of P450-containing systems on WorldWide Web

Kirill N. Degtyarenko1* and Péter Fábián2

Abstract

To facilitate access to electronic resources for all researchers working in the field of P450 proteins and P450-containing systems, a WorldWide Web server has been established called The Directory of P450-containing Systems at <http://www.icgeb.trieste.it/p450/>. Currently it contains the most up-to-date list of sequences of both the P450 superfamily and proteins mediating electron transfer to P450, i.e. NADPH:P450 reductases, specific NAD(P)H:ferredoxin reductases, cytochrome b5 reductases, ferredoxins and cytochromes b5, and their homologues from different enzyme systems. All the referenced sequences are provided with accession numbers and cross-links to major sequence databanks: PIR, SWISS-PROT, EMBL/GenBank and PRF.

Introduction

P450 enzymes constitute a superfamily of haem-thiolate proteins, widely distributed in bacteria, fungi, plants and animals. The enzymes are involved in the metabolism of a plethora of both exogenous and endogenous compounds. Usually, they act as terminal oxidases in multicomponent electron transfer chains, termed P450-containing monooxygenase systems (Degtyarenko, 1995). All known P450-containing systems share a common structural and functional domain architecture. Apart from P450 itself, these systems can comprise several fundamentally different protein components and domains, all of which can be shared by other multicomponent/multidomain enzyme systems with various functions: FAD flavoprotein or domain, FMN domain, 2Fe-2S ferredoxin, 3Fe-4S ferredoxin, cytochrome b5. Either a FMN (flavodoxin-like) domain, a ferredoxins, or cytochrome b5 serve as the electron transport intermediate between the FAD domain and P450. In turn, on the basis of sequence similarity, FAD flavoproteins can be categorised into three major families:

(i) ferredoxin:NADP+ reductase (FNR) family: NADPH:P450 reductase, cytochrome b5 reductase;
(ii) glutathione reductase (GR) family: putidaredoxin reductase, terpredoxin reductase;
(iii) NADPH:adrenodoxin reductase, which shares local similarity with glutamate synthases and NADH peroxidase.

The rapid growth of sequence information is making the gathering together of a comprehensive and up-to-date data set for extensive protein families increasingly laborious. For example, at least 450 different P450 sequences have been published to date (Nelson, 1995), which represents a doubling of the collection since the 1993 nomenclature update (Nelson et al., 1993). In an effort to facilitate access to electronic resources for all researchers working in the field of P450 proteins and P450 containing systems, we have created The Directory of P450-containing Systems (DPS), a WorldWide Web server available at <http://www.icgeb.trieste.it/p450/>.

DPS contents and structure

Currently DPS presents the following data:

1. ‘Core’ (lists of accession numbers for components of P450-containing systems)
   - Full list of P450 superfamily
   - Animal P450s
   - Fungal P450s
   - Plant P450s
   - Bacterial P450s
   - NADPH:P450 reductases (CPR)
   - Fe-S protein- and cytochrome b5 reductases
   - Fe-S proteins
   - Cytochromes b5
   - Cytochrome b5-like domain containing proteins
   - Ferredoxin:NADP+ reductases and FNR-like domain containing proteins
   - GR-like oxidoreductases
   - Flavodoxins
   - Genes with known exon/intron structure
Table I. Contents of DPS 'core'

<table>
<thead>
<tr>
<th>Group of proteins</th>
<th>No. of genes</th>
<th>No. of species</th>
</tr>
</thead>
<tbody>
<tr>
<td>P450 superfamily, including</td>
<td>434</td>
<td>90</td>
</tr>
<tr>
<td>Animal P450s</td>
<td>323</td>
<td>37</td>
</tr>
<tr>
<td>Fungal P450s</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>Plant P450s</td>
<td>44</td>
<td>17</td>
</tr>
<tr>
<td>Bacterial P450s</td>
<td>32</td>
<td>21</td>
</tr>
<tr>
<td>CPR family (flavodoxin/FNR fusion proteins), including</td>
<td>39</td>
<td>28</td>
</tr>
<tr>
<td>NADPH:P450 reductases</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>Nitric oxide synthases</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>NADPH:sulphite reductases</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>FNR superfamily (except for CPR family), including</td>
<td>99</td>
<td>69</td>
</tr>
<tr>
<td>NADH:cytochrome b5 reductases</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>GR superfamily,</td>
<td>138</td>
<td>83</td>
</tr>
<tr>
<td>including</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P450-specific Fe-S reductases</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adrenodoxin reductases</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Iron-sulphur proteins</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>Flavodoxins</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>Cytochromes b5</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Proteins containing cytochrome b5-like domain</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>781</td>
<td>226</td>
</tr>
</tbody>
</table>

2. Supplementary information:
- Families of structural domains of P450-containing systems
- PRINTS entries for structural domains of P450-containing systems
- List of known three-dimensional structures
- OMIM entries related to P450-containing systems
- Bibliography on P450-containing systems

The 'core' of DPS is systematic information on protein and nucleic acid sequences of the P450 superfamily as well as other components of P450-containing systems and their homologues. Strictly speaking, DPS is more a 'virtual' than a 'real' tool (Harper, 1995), i.e. rather than containing sequence information itself, it is composed mainly of direct links to 'primary' sequence databanks: PIR (George et al., 1994), SWISS-PROT (Bairoch and Boeckmann, 1994), EMBL (Emmert et al., 1994) and PRF/SEQDB (Protein Research Foundation, Osaka). In addition, DPS has cross-references to other databases such as PROSITE (Bairoch and Bucher, 1994), PRINTS (Attwood et al., 1994), ProDom (Sonnhammer and Kahn, 1994), ProtFam at MIPS (George et al., 1994), Entrez Taxonomy Database at NCBI (Benson et al., 1994), ENZYME (Bairoch, 1994), LIGAND (Suyama et al., 1993) and OMIM (Pearson et al., 1994).

Table II. PRINTS entries for structural domains of P450-containing systems

<table>
<thead>
<tr>
<th>PRINTS entry</th>
<th>Number of elements</th>
<th>Description</th>
<th>True positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>P450</td>
<td>5</td>
<td>P450 superfamily</td>
<td>354</td>
</tr>
<tr>
<td>BP450</td>
<td>9</td>
<td>B-class P450</td>
<td>23</td>
</tr>
<tr>
<td>MITP450</td>
<td>9</td>
<td>Mitochondrial P450</td>
<td>36</td>
</tr>
<tr>
<td>CYTOCHROMES</td>
<td>3</td>
<td>Cytochrome b5 superfamily</td>
<td>54</td>
</tr>
<tr>
<td>ADRENODOXIN</td>
<td>3</td>
<td>Adrenodoxin family</td>
<td>15</td>
</tr>
<tr>
<td>3FE4SFRDOXIN</td>
<td>3</td>
<td>3Fe-4S ferredoxins</td>
<td>15</td>
</tr>
<tr>
<td>FLAVDOXIN</td>
<td>4</td>
<td>Flavodoxin superfamily</td>
<td>66</td>
</tr>
<tr>
<td>FPNCR</td>
<td>8</td>
<td>Flavoprotein pyridine nucleotide cytochrome reductases (FNR superfamily)</td>
<td>81</td>
</tr>
<tr>
<td>CYTBDRTASE</td>
<td>6</td>
<td>Cytochrome b5 reductases</td>
<td>47</td>
</tr>
<tr>
<td>FADPR</td>
<td>5</td>
<td>FAD-dependent pyridine nucleotide reductases (GR superfamily)</td>
<td>79</td>
</tr>
<tr>
<td>ADXRDTASE</td>
<td>5</td>
<td>Adrenodoxin reductases</td>
<td>15</td>
</tr>
</tbody>
</table>
Mitochondrial (1) and most bacterial P450-containing systems (2) have three components: an FAD-containing flavoprotein (NADPH or NADH-dependent reductase), an iron-sulphur protein, and P450.

**NADPH** → adrenodoxin reductase → adrenodoxin → P450 (1)

**NADH** → ferredoxin reductase → ferredoxin → P450 (2)

It was shown that in microsomal P450-containing systems cytochrome b₅ can serve as electron donor for P450s:

**NAD(P)H:Fe-S reductases:***

<table>
<thead>
<tr>
<th>Gene</th>
<th>Species</th>
<th>PIR</th>
<th>SwissProt</th>
<th>GenBank/EMBL/DDBJ</th>
<th>PRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. NADPH:adrenodoxin reductase (EC 1.18.1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adxR</td>
<td>bov</td>
<td>A29604</td>
<td>J0751</td>
<td>P91658</td>
<td>M17029</td>
</tr>
<tr>
<td></td>
<td>hum</td>
<td>A16492</td>
<td>A40487</td>
<td>P22570</td>
<td>M85058</td>
</tr>
<tr>
<td></td>
<td>SCE</td>
<td>A40487</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**B. NADH:ferredoxin:NADH reductases (EC 1.18.1.2)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Species</th>
<th>PIR</th>
<th>SwissProt</th>
<th>GenBank/EMBL/DDBJ</th>
<th>PRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>tarpA</td>
<td>Pae</td>
<td>D27654</td>
<td>D42971</td>
<td>P33009</td>
<td>M91440</td>
</tr>
<tr>
<td>forA</td>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>thcD</td>
<td>Rho</td>
<td>U17120</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**C. NADH:cytochrome b₅ reductase (EC 1.6.2.2)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Species</th>
<th>PIR</th>
<th>SwissProt</th>
<th>GenBank/EMBL/DDBJ</th>
<th>PRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ber</td>
<td>bov</td>
<td>A23896</td>
<td>P97514</td>
<td>M83104</td>
<td>1112550A</td>
</tr>
<tr>
<td></td>
<td>hum</td>
<td>US0468</td>
<td>B26616</td>
<td>P00387</td>
<td>M1646</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P28706</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M28710</td>
</tr>
<tr>
<td></td>
<td>rat</td>
<td>JX0110</td>
<td>P00707</td>
<td>007867</td>
<td>771117</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2104333X</td>
</tr>
<tr>
<td></td>
<td>SCE</td>
<td>D49691</td>
<td>P16060</td>
<td>228150</td>
<td>228677</td>
</tr>
</tbody>
</table>

**MIPS**

MIPS protein superfamily: 120.0
MIPS protein domain family: 00050

**Entrez taxonomy**

| bov = bovine | hum = human | Pae = Pseudomonas species | Psp = Pseudomonas sp | Rat = rat | Pse = Ps. aeruginosa | Ser = S. cerevisiae (baker's yeast) |

**Species abbreviations**

bov = bovine
hum = human
Pae = Pseudomonas species
Psp = Pseudomonas sp
Rat = rat
Pse = Ps. aeruginosa
Ser = S. cerevisiae (baker's yeast)

Fig. 1. A typical page from DPS (boxed). Underlined text indicates the hypertext links to the number of databases (listed on the top and along the left margin).
positioned in the CYP1A1
hum field and therefore the user
will not treat them as different genes in spite of their
distinct ’trivial’ names [P-450 6, P(1)-450, P-1–450, P-
450c, P4501A1]. In the EMBL/GenBank database, the
multi-exon genes are sometimes presented as a series of
entries. For example, entries M30795 through M30804
represent the human CYP19 gene where each entry
corresponds to one exon; six entries (M58932 — M58937)
span the rat CPR gene consisting of 16 exons. A special
page in DPS is dedicated to genes with a known exon/intron
structure.

Search in DPS
To provide simple and fast textual searching in the whole
P450 directory, glimpse version 3.0 (Manber and Wu,
1994) was installed.

Tools for updating DPS
To shorten the time-consuming process of searching for
novel related sequences, we have created a small program
in perl (Wall and Schwartz, 1991). The program works in
the following way:

• It connects to selected WorldWide Web servers (currently to expasy.hcuge.ch, www.genome.ad.jp
and www.ebi.ac.uk). These servers offer the possibility
of searching in the SWISS-PROT, PRF and
EMBNew databanks, respectively
• The program searches the above databanks using 30
pre-specified keywords
• It then retrieves the results of each search and
extracts the accession number for each sequence found
• After comparing the accession numbers with those
already included in DPS, the program extracts the
missing ones and inserts them into a hypertext file.

Finally, the new sequences are classified manually and
inserted into DPS. The above procedure can easily be
employed for the semi-automatic update of any other
sequence library.

Fingerprinting the structural domains of P450-containing
systems
Structural domains can be characterized in terms of a
number of separate sequence motifs. The groups of motifs,
constituting the signature for a particular (super)fami-
y of sequences (also referred to as ‘fingerprint’s), provide a tool
for the effective search of distantly homologous sequences
(Attwood et al., 1994). For all the components/domains of
P450 systems, the corresponding fingerprints were con-
structed, which can be used to identify the new members of
these protein superfamilies, even if the overall amino acid
sequence identities are at or below the limit of significance
(Table II). The fingerprint entries are integrated in the
PRINTS database and are available both via the www
server at <http://www.biochem.ucl.ac.uk/bcm/dbbrowser/> and via the anonymous ftp elsewhere (Attwood et
al., 1994). Extra fingerprints will be added in the incoming
releases of PRINTS to allow the further discrimination
between protein families, e.g. P450 families.

Acknowledgements
The authors thank Dr T.K.Attwood for her inestimable help in the
creation of the fingerprints. This project was launched during the authors’
fellowship at the Protein Structure and Function Laboratory of the
International Centre for Genetic Engineering and Biotechnology, Trieste,
Italy.

References
22, 3590–3596.
3626–3627.
Degtyarenko,K.N. (1995) Structural domains of P450-containing mono-
xygenase systems. Protein Engineering, 8, 737–747.
European Bioinformatics Institute (EBI) databases. Nucleic Acids
George,D.G., Barker,W.C, Mewes,H.-W., Pfeiffer,F. and Tsugita,A.
(1994) The PIR-International Protein Sequence Database Nucleic
Genet., 11, 223–228.
Nelson,D.R., Kamataki,T., Waxman,D.J., Guengerich,F.P., Estab-
took,R.W., Feyereisen,R., Gonzalez,F.J., Coon,M.J., Gunsalus,I.C,
update on new sequences, gene mapping, accession numbers, early
trivial names of enzymes, and nomenclature. DNA Cell Biol., 12,1–51.
International symposium on Cytochrome P450 Biodiversity, Woods
file systems. In Proceedings of the USENIX Winter Technical
Conference, San Francisco, pp. 23–32. Glimpse is available by
anonymous ftp at ftp.cs.arizona.edu.
Pearson,P., Franchonano,C., Foster,P., Bocchini,C., Li,P and McKu-
protein families, e.g. P450 families. Protein Science,
3, 892–893.
amino acid sequence motifs among enzymes: the enzyme-reaction
Wall,L. and Schwartz,R (1991) Programming Perl. O'Reilly and
Associates, Inc.

Received on January 2, 1996; accepted on March 8, 1996