PROMISE: a new database of information on prosthetic centres and metal ions in protein active sites

K.N. Degtyarenko, A.C.T. North and J.B.C. Findlay

Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

To whom correspondence should be addressed

The PROMISE (Prosthetic centres and metal ions in protein active sites) database project has been launched to gather together comprehensive sequence, structural, functional and bibliographic information on proteins which possess prosthetic centres, with an emphasis on active site structure and function. PROMISE version 1.0 comprises data on iron-containing proteins.

Background

There are several regularly updated, publicly accessible databases which deal specifically with protein domains. PROSITE (Bairoch et al., 1996) is based on heuristically defined sequence motifs which are predominantly represented as short regular expressions, defining the positions of residues that are invariant or of limited structural class; several PROSITE entries are also represented by sequence profiles. PRINTS (Attwood et al., 1996) also is based on heuristically and objectively defined motifs but uses fingerprints, often of several distinct motifs and which are more robust than PROSITE regular expressions. BLOCKS (Pietrokovski et al., 1996) is essentially based on PROSITE domains but uses an extended representation of short conserved regions (blocks) for the sequences. Both PROSITE and PRINTS entries are annotated and contain bibliographic information and cross-references to sequence and structural databases.

SCOP (acronym for Structure Classification Of Proteins) is a database facilitating understanding of and access to information available for protein three-dimensional (3-D) structures (Murzin et al., 1995). The classification of proteins in SCOP is based on evolutionary relationships and on the characteristics of their 3-D structures. Monodomain proteins are treated as a whole, whereas the domains in multidomain proteins are usually classified individually. The 'place' of a given protein in the classification sequence family/superfamily/fold/class is unambiguously defined. CATH (acronym for Class, Architecture, Topology, Homology) is a hierarchical classification of protein structural relationships derived using a combination of automatic and manual methods (Michie et al., 1995). CATH and SCOP have some differences in definition of their hierarchical levels. In addition, for each entry CATH provides a link to a protein–ligand interaction graphical summary constructed with the LIGPLOT program (Wallace et al., 1995).

The Ligand Chemical Database (Suyama et al., 1993) is designed to provide a linkage between chemical and biological aspects of life in the light of enzymatic reactions and is a combination of the ENZYME and COMPOUND databases. It contains information such as structural formulae of low molecular mass biological ligands: enzyme prosthetic centres, co-enzymes, substrates and products and links to the corresponding protein sequences. A number of electron-transfer proteins, e.g. ferredoxins and cytochromes, are also considered if they are involved in an enzyme reaction. Structural or bibliographic information is not provided.

Thus, there has been no single database which is focused on protein active site structure and on the relationship between prosthetic centre and protein molecule, combining the relevant sequence, structural and physico-chemical information. In an attempt to create such a database, we have constructed PROMISE and have now made available version 1.0, which comprises data on iron-containing proteins. An important characteristic of PROMISE is that it has been designed from the outset for interrogation via the WorldWide Web and incorporates hypertext links to other relevant databases.

Structure of PROMISE entries

Entries in the PROMISE database follow a hierarchy: at the top level are major groups (such as haem proteins and iron–sulphur proteins); these consist of classes (such as cytochromes, haem peroxidases and Fe$_2$S$_2$ proteins), which in turn consist of families (such as Class II cytochromes $_c$, animal haem peroxidases and adrenodoxin-type ferredoxins). The classification is not rigid and some intermediate or alternative levels could easily be introduced.

A typical group entry (e.g. iron–sulphur proteins) contains:
- Classification by function
- Classification by prosthetic group type and/or coordination and relevant class entries in the PROMISE database (e.g. Fe$_2$S$_2$ proteins)
- Classification by type and number of prosthetic groups
- Relevant entries in motif databases: PRINTS, PROSITE and BLOCKS (with hypertext links to them)
- Relevant family entries in the PROMISE database
- Classification by type of HET (hetero) groups in the PDB
- References (with hypertext links, where available)

Selection of the link to the Fe$_2$S$_2$ protein class displays:
- Table summarizing properties of the prosthetic group typical for the class: its chemical structure and formal oxidation states
- Short description of protein class
- Relevant family entries in the PROMISE database (e.g. adrenodoxin)
- Relevant entries in motif databases: PRINTS, PROSITE and BLOCKS (with hypertext links to them)
- References (with hypertext links, where available)

Selection of the link to the adrenodoxin family gives (Figure 1):
- Table summarizing properties of prosthetic group(s): chemical structure, coordination, amino acid ligands of metal ion(s), formal oxidation states
Adrenodoxin, putidaredoxin and terpredoxin are soluble Fe\(_2\)S\(_2\) iron–sulphur proteins that act as single electron carriers. In mitochondrial monooxygenase systems, adrenodoxin transfers an electron from NADPH: adrenodoxin reductase to membrane-bound P450 [1, 2]. In bacteria, putidaredoxin and terpredoxin serve as electron carriers between corresponding NADH-dependent ferredoxin reductases and soluble P450 [3, 4]. The exact functions of other members of this family are not known.

The 3-D structure of putidaredoxin has been solved by NMR [5]. Despite a low sequence similarity between putidaredoxin and plant-type ferredoxins, putidaredoxin retains a similar folding topology to structurally characterized plant-type ferredoxins. The fold belongs to the \(\alpha + \beta\) class, with 3 \(\alpha\)-helices and 6 \(\beta\)-strands forming a barrel-like structure, and an extruded loop containing three of the four cysteinyl residues of the iron–sulphur cluster:

Adrenodoxin-type ferredoxins in motif databases

<table>
<thead>
<tr>
<th>PRINTS ID</th>
<th>PRINTS AC</th>
<th>PROSITE/BLOCKS ID</th>
<th>PROSITE AC</th>
<th>BLOCKS AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADRENODOXIN</td>
<td>PR00355</td>
<td>ADX</td>
<td>PS00814</td>
<td>BL00814</td>
</tr>
</tbody>
</table>
Adrenodoxin-type ferredoxins in 3-D databases

Putidaredoxin contains single [Fe₂S₂] prosthetic group.

<table>
<thead>
<tr>
<th>PDB</th>
<th>scop</th>
<th>BSM</th>
<th>HEADER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1put</td>
<td>1put</td>
<td>1put</td>
<td>Putidaredoxin (oxidised); <em>Pseudomonas putida</em>, strain ATCC 17453</td>
</tr>
</tbody>
</table>

¹ Macromolecular Structures abstract. Full text is available to BioMedNet Members

References


Bibliography on structural studies of adrenodoxin-type ferredoxins

Fig. 1. A typical family entry from PROMISE database. Underlined text indicates the hypertext links.

- Description of protein family
- Relevant entries in motif databases: PRINTS, PROSITE and BLOCKS (with hypertext links to them)
- Relevant entries in databases of 3-D structures: PDB, SCOP, CATH and Macromolecular Structures Database (Hendrickson and Wuthrich, 1991–96) with hypertext links
- References (with hypertext links, where available)

Every family entry contains a hypertext link to the relevant structural bibliography entry (e.g. bibliography on structural studies of fungal, plant and bacterial haem peroxidases), containing references on crystallographic and spectroscopic studies.

Comparison of PROMISE with existing databases

As in PROSITE and PRINTS, a group of proteins sharing a certain feature is analysed and the structural feature is described; a concise description and bibliography are presented; links to a variety of on-line databases are collected. As in SCOP, the data are organized hierarchically, giving an idea about the ‘place’ or genealogy of certain (groups of) metallo-proteins or other complex proteins.

In contrast to SCOP or CATH, the proteins are categorized on the basis of their active/binding site structure. Although there are many families of proteins sharing both fold and active site structure, there are also evolutionarily unrelated but functionally analogous systems which have similar active site structures (e.g. bacterial-type ferredoxins and high potential iron proteins (HiPIPs); animal-type peroxidases and plant/fungal/bacterial-type peroxidases; the P450 and nitric oxide synthase haem domains); yet again, there are proteins which share the same fold but have distinct active site structures (cf. cytochrome b₅₆₂ and cytochrome c’; ribonucleotide reductase-like proteins, ferritins and cytochrome b₈). The latter are grouped together in SCOP and CATH. The former are grouped together in PROMISE, in which a description of the protein active site and prosthetic centre is presented.

The classification of proteins by type of prosthetic centre in PROMISE differs from that which is inherent in Brookhaven Protein Data Bank (PDB) entries (Bernstein et al., 1977), which rest solely on the type of hetero (HET) groups. For instance, there are no distinctions in PDB between haem groups b and c, between iron coordinated by four cysteines (as in rubredoxin) and any other iron ion. The classification of PDB entries extracted by DbBrowser (Michie et al., 1996) makes use of this simplistic definition of ligand types.

The information on protein families accumulated so far in the PROSITE and PRINTS databases has proved to be very helpful during the work on PROMISE. However, a number of important ‘gaps’ in domain databases have been discovered. For example: there are no entries in PRINTS corresponding to protein families consisting of only a few members, such as cytochrome b₅₆₂, desulfoferredoxin, prismane; the CYTOCHROME_C signature in PROSITE does not discriminate the structurally and functionally distinct Class I, II, III and IV cytochromes c, cytochrome c₁ and cytochrome f.

Thus, while PROMISE interrelates closely to a number of existing structural and sequence databases, it brings together for the first time a comprehensive description of prosthetic
centre binding sites, together with ready access, via the
hypertext links, to the other relevant data resources. It can
thus provide an invaluable resource to aid any investigation of
binding site structure from the viewpoint of protein engineering,
protein design and interactions of protein with ligands or
drugs. Although version 1.0 is limited to iron-containing
proteins, entries to other protein groups are currently being
prepared in our laboratory and will be incorporated in PROM-
ISE as they are completed.

Availability
PROMISE is available on the WorldWide Web from URL:
<http://bioinf.leeds.ac.uk/promise/>. Use of Netscape version
2.0 or newer is recommended.

Acknowledgements
This work has been supported by Astra Charnwood, Glaxo Wellcome, Pfizer,
Zeneca and the BBSRC under the UK DTI/BBSRC Biotechnology LINK
Programme.

References
Attwood,T.K., Beck, M.E., Bleasby,A.J., Degtyarenko,K. and Parry-Smith,D.J.
189–196.
Bernstein,F.C., Koetzle,T.F., Williams,G.I.B., Meyer,E.F., Jr, Brice,M.D.,
chbiology/mms.htm.)
Michie,A.D., Hutchinson,E.G., Laskowski,R.A., Orengo,C.A. and
21, 191.
247, 536–540.
197–200.
9, 9–15.
127–134.

Received December 10, 1996; accepted December 30, 1996