

Proteins import format

Proteins are imported into the system from the XML format of an international database of protein Uniprot. This format is exhaustively described protein, but for the needs of the system retrieves only required minimum of information. One file can contain multiple proteins.

Example of import file format

```
<uniprot xmlns="http://uniprot.org/uniprot"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://uniprot.org/uniprot http://www.uniprot.org/support/docs/uniprot.xsd">
  <entry dataset="Swiss-Prot" created="2000-05-30" modified="2012-02-22" version="123">
    <accession>O43520</accession>
    <accession>Q9BTP8</accession>
    <name>AT8B1_HUMAN</name>
    <protein>
      <recommendedName>
        <fullName>Probable phospholipid-transporting ATPaseIC</fullName>
        <ecNumber>3.6.3.1</ecNumber>
      </recommendedName>
      <alternativeName>
        <fullName>ATPase class I type 8B member 1</fullName>
      </alternativeName>
      <alternativeName>
        <fullName>Familial intrahepatic cholestasis type 1</fullName>
      </alternativeName>
    </protein>
    <gene>
      <name type="primary">ATP8B1</name>
      <name type="synonym">ATPIC</name>
      <name type="synonym">FIC1</name>
      <name type="synonym">PFIC</name>
    </gene>
  </entry>
</uniprot>
```

```
<organism>

  <name type="scientific">Homo sapiens</name>

  <name type="common">Human</name>

  <dbReference type="NCBI Taxonomy" id="9606" />

  <lineage>

    <taxon>Eukaryota</taxon>

  </lineage>

</organism>

<reference key="1">

  <citation type="journal article" date="1998"

name="Nat. Genet." volume="18" first="219" last="223">

  <title>A gene encoding a P-type ATPase mutated in two forms

of hereditary cholestasis.</title>

  <authorList>

    <person name="Bull L.N." />

  </authorList>

  <dbReference type="MEDLINE" id="98160178" />

</citation>

<scope>NUCLEOTIDE SEQUENCE [MRNA]</scope>

<scope>VARIANTS PFIC1 SER-288; VAL-308; 645-ILE--ILE-699 DEL

AND ARG-892</scope>

<scope>VARIANTS BRIC1 THR-661 AND 795-GLY--ARG-797 DEL</scope>

<scope>VARIANT THR-1152</scope>

<source>

  <tissue>Intestine</tissue>

  <tissue>Liver</tissue>

</source>

</reference>

<comment type="tissue specificity">

  <text>Found in most tissues except brain and skeletal muscle.
```

Most abundant in pancreas and small intestine.</text>

</comment>

<dbReference type="EC" id="3.6.3.1" />

<proteinExistence type="evidence at protein level" />

<sequence length="1251" mass="143695"

checksum="770FEF3946CB579F" modified="2009-05-05" version="3">

MSTERDSETTFDEDSQPNDVVPYSDDETEDELDDQGSVAPEQNRVNRVREAEENREPFRK</sequence>

</entry>

</uniprot>

Tags description

Table 1 Tags description for proteins import

Тэг	Атрибуты	Описание
uniprot		The required root element.
entry		The tag describes a single protein.
name		Uniprot id of the protein.
accession		Accession number of protein in Uniprot. The first element become the main AC of the protein, the rest - the alternative AC.
protein		The container for the name of the protein.
recommendedName>fullName		The main name of the protein.
alternativeName>fullName		The alternative names of the protein.
gene>name		The gene name of the protein.
	type	Type the gene name: PRIMARY – main, SYNONYM - an alternative.