February 3, 2012

Went to the Orientations of Proteins in Membranes database, went to the "Beta-barrel transmembrane superfamilies" page, and saved the source code as "opm beta barrel tm.html"

Downloaded FASTA format sequences for each of the 98 listed beta barrels, using the following code:

import re

from urllib import urlretrieve

pdbids = set()

with open('opm beta barrel tm.html', 'r') as f:

for line in f:

search = re.search('protein.php\?pdbid=(....)', line)

if search is not None:

pdbids.add(search.group(1))

for pdbid in pdbids:

urlretrieve('http://www.rcsb.org/pdb/download/' +\

'downloadFile.do?fileFormat=FASTA&compression=NO&' +\

'structureId=' + pdbid.upper(),

'bbtm sequences/{}.fasta'.format(pdbid))

I renamed the folder they went to "98 bbtm sequences", so it will be ovbious when it's out of date.