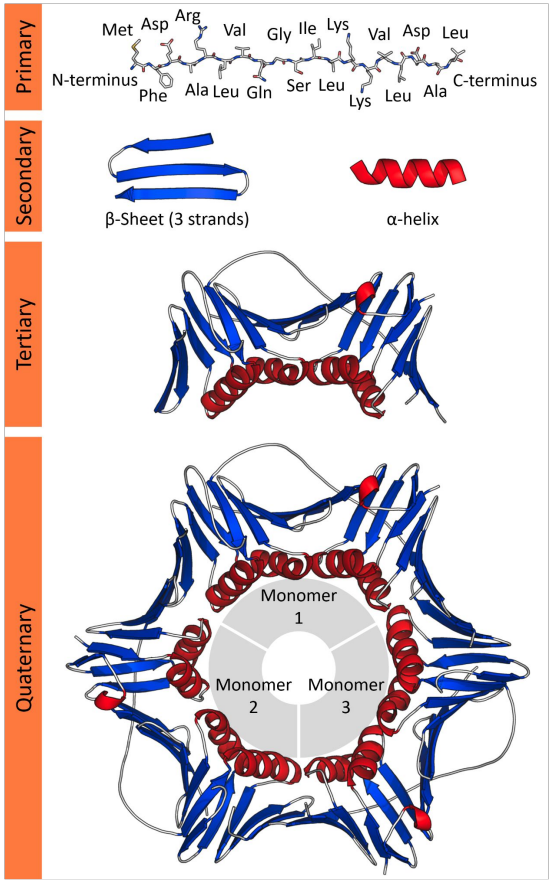
Secondary structure of protein

Some practical aspects

Protein structures

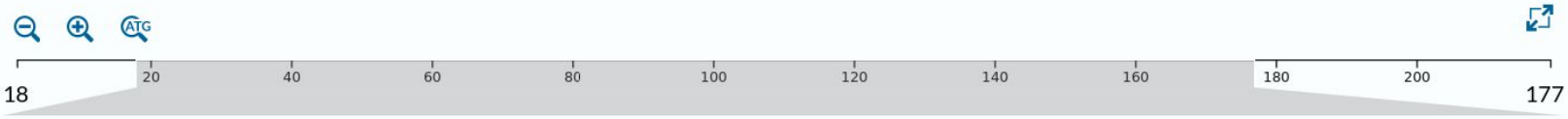
- Get structures from databases
- Visualise
- Compare with each other
- Predict
- Classify



Secondary structure in Uniprot

Features

Showing features for helix[i], turn[i], beta strand[i].



TYPE	ID	POSITION(S)	DESCRIPTION	
-- Select --				
Helix		43-58	Combined Sources	BLAST Add
Helix		59-61	Combined Sources	BLAST Add
Turn		74-78	Combined Sources	BLAST Add
Helix		81-84	Combined Sources	BLAST Add
Helix		88-92	Combined Sources	BLAST Add
Beta strand		107-109	Combined Sources	BLAST Add

biopython

from Bio import SeqIO, PDB

PDB
https://biopython-cn.readthedocs.io/zh_CN/latest/en/chr11.html

11.1 Reading and writing crystal structure files

11.1.1 Reading a PDB file

First we create a `PDBParser` object:

```
>>> from Bio.PDB.PDBParser import PDBParser
>>> p = PDBParser(PERMISSIVE=1)
```

The `PERMISSIVE` flag indicates that a number of common problems (see 11.7.1) associated with PDB files will be ignored (but note that some atoms and/or residues will be missing). If the flag is not present a `PDBConstructionException` will be generated if any problems are detected during the parse operation.

The Structure object is then produced by letting the `PDBParser` object parse a PDB file (the PDB file in this case is called 'pdb1fat.ent', '1fat' is a user defined name for the structure):

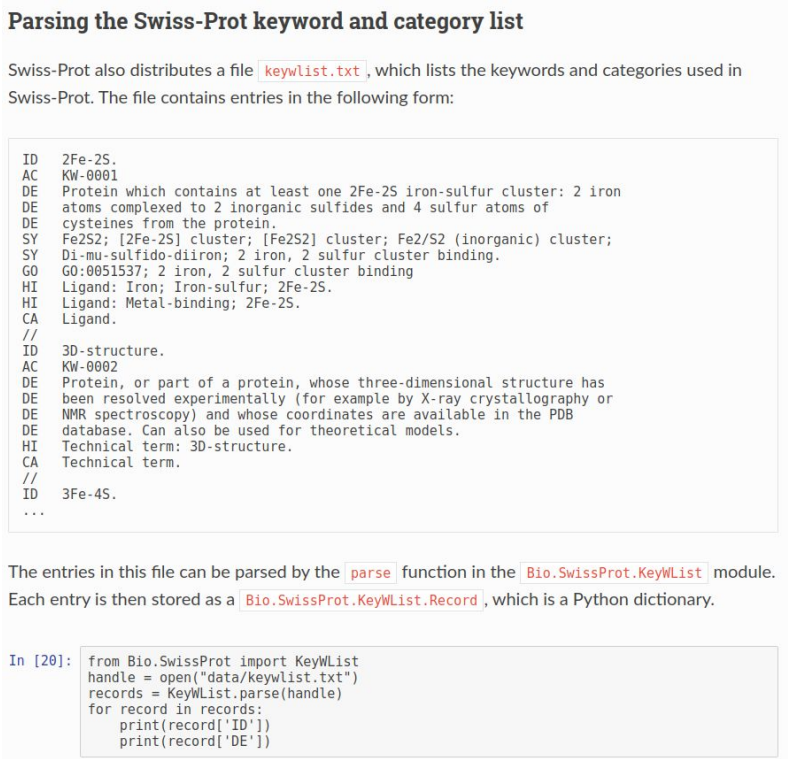
```
>>> structure_id = "1fat"
>>> filename = "pdb1fat.ent"
>>> s = p.get_structure(structure_id, filename)
```

biopython

from Bio import SwissProt

https://biopython-tutorial.readthedocs.io/en/latest/notebooks/10%20-%20Swiss-Prot%20and%20ExPASy.html

Uniprot text file



Parsing the Swiss-Prot keyword and category list

Swiss-Prot also distributes a file `keywlist.txt`, which lists the keywords and categories used in Swiss-Prot. The file contains entries in the following form:

```
ID   2Fe-2S.
AC   KW-0001
DE   Protein which contains at least one 2Fe-2S iron-sulfur cluster: 2 iron
DE   atoms complexed to 2 inorganic sulfides and 4 sulfur atoms of
DE   cysteines from the protein.
SY   Fe2S2; [2Fe-2S] cluster; [Fe2S2] cluster; Fe2/S2 (inorganic) cluster;
SY   Di-mu-sulfido-diiron; 2 iron, 2 sulfur cluster binding.
GO   GO:0051537; 2 iron, 2 sulfur cluster binding
HI   Ligand: Iron; Iron-sulfur; 2Fe-2S.
HI   Ligand: Metal-binding; 2Fe-2S.
CA   Ligand.
//
ID   3D-structure.
AC   KW-0002
DE   Protein, or part of a protein, whose three-dimensional structure has
DE   been resolved experimentally (for example by X-ray crystallography or
DE   NMR spectroscopy) and whose coordinates are available in the PDB
DE   database. Can also be used for theoretical models.
HI   Technical term: 3D-structure.
CA   Technical term.
//
ID   3Fe-4S.
...
```

The entries in this file can be parsed by the `parse` function in the `Bio.SwissProt.KeyWList` module. Each entry is then stored as a `Bio.SwissProt.KeyWList.Record`, which is a Python dictionary.

In [20]:
```
from Bio.SwissProt import KeyWList
handle = open("data/keywlist.txt")
records = KeyWList.parse(handle)
for record in records:
    print(record['ID'])
    print(record['DE'])
```

Secondary structure prediction/extraction from 3D structure

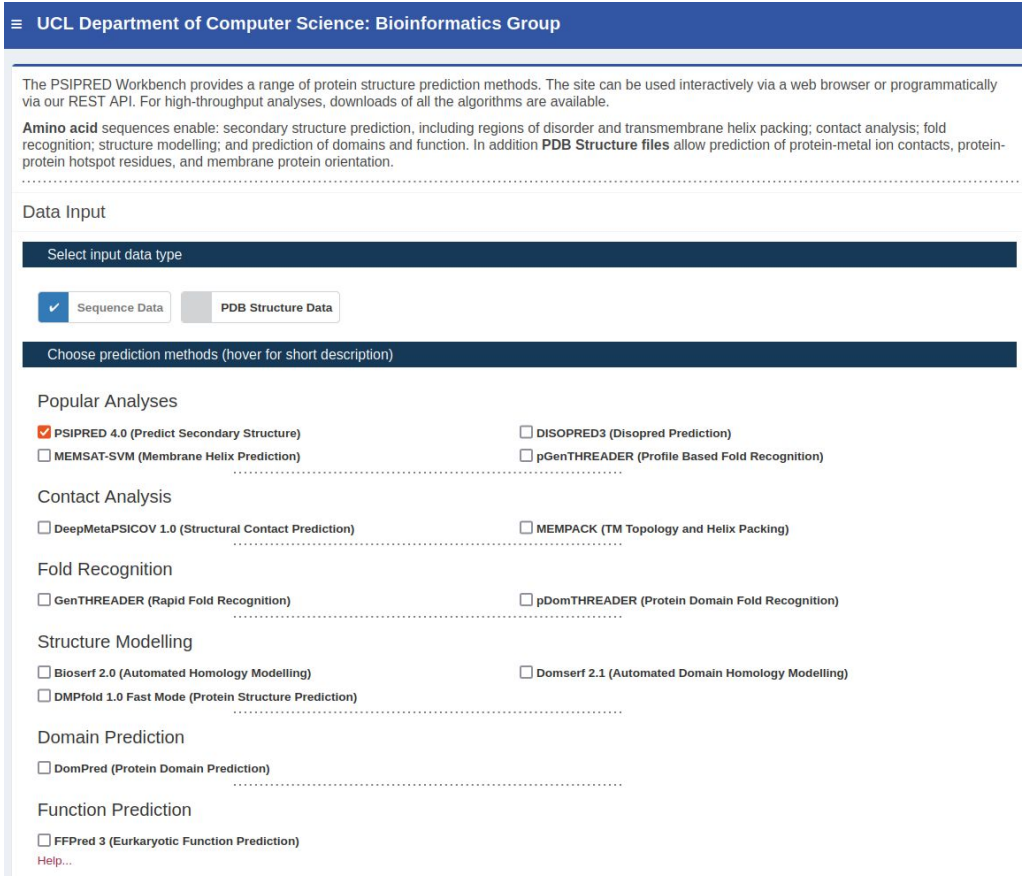
http://bioinf.cs.ucl.ac.uk/psipred/

DSSP algorithm

https://swift.cmbi.umcn.nl/gv/dssp/

Web server
https://www3.cmbi.umcn.nl/xssp/

PSIPRED


UCL Department of Computer Science: Bioinformatics Group
The PSIPRED Workbench provides a range of protein structure prediction methods. The site can be used interactively via a web browser or programmatically via our REST API. For high-throughput analyses, downloads of all the algorithms are available.
Amino acid sequences enable: secondary structure prediction, including regions of disorder and transmembrane helix packing; contact analysis; fold recognition; structure modelling; and prediction of domains and function. In addition PDB Structure files allow prediction of protein-metal ion contacts, protein-protein hotspot residues, and membrane protein orientation.
Data Input
Select input data type
Sequence Data
PDB Structure Data
Choose prediction methods (hover for short description)
Popular Analyses
PSIPRED 4.0 (Predict Secondary Structure)
DISOPRED3 (Disopred Prediction)
MEMSAT-SVM (Membrane Helix Prediction)
pGenTHREADER (Profile Based Fold Recognition)
Contact Analysis
DeepMetaPSICOV 1.0 (Structural Contact Prediction)
MEMPACK (TM Topology and Helix Packing)
Fold Recognition
GenTHREADER (Rapid Fold Recognition)
pDomTHREADER (Protein Domain Fold Recognition)
Structure Modelling
Bioserf 2.0 (Automated Homology Modelling)
Domserf 2.1 (Automated Domain Homology Modelling)
DMPfold 1.0 Fast Mode (Protein Structure Prediction)
Domain Prediction
DomPred (Protein Domain Prediction)
Function Prediction
FFPred 3 (Eurkaryotic Function Prediction)
Help...

Submission details

Protein Sequence

>sp|P29459|IL12A_HUMAN Interleukin-12 subunit alpha OS=Homo sapiens OX=9606 GN=IL12A PE=1 SV=2
MCPARSLLLVATLVLLDHLSLARNLPVATPDPGMFPCLHHSQNLLRAVSNMLQKARQTLE
FYPCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMM

Help...
If you wish to test these services follow this link to retrieve a test fasta sequence.

Job name

sp|P29459|IL12A_HUMAN

Email (optional)

Email (optional)

Reset Submit

Sequence Plot

Show psipred | Show memsat | Show aatypes

```
          10        20        30        40        50
1   M C P A R S L L L V A T L V L L D H L S L A R N L P V A T P D P G M F P C L H H S Q N L L R A V S N   50
51  M L Q K A R Q T L E F Y P C T S E E I D H E D I T K D K T S T V E A C L P L E L T K N E S C L N S R   100
101 E T S F I T N G S C L A S R K T S F M M A L C L S S I Y E D L K M Y Q V E F K T M N A K L L M D P K   150
151 R Q I F L D Q N M L A V I D E L M Q A L N F N S E T V P Q K S S L E E P D F Y K T K I K L C I L L H   200
201 A F R I R A V T I D R V M S Y L N A S                                                               219
          10        20        30        40        50
```

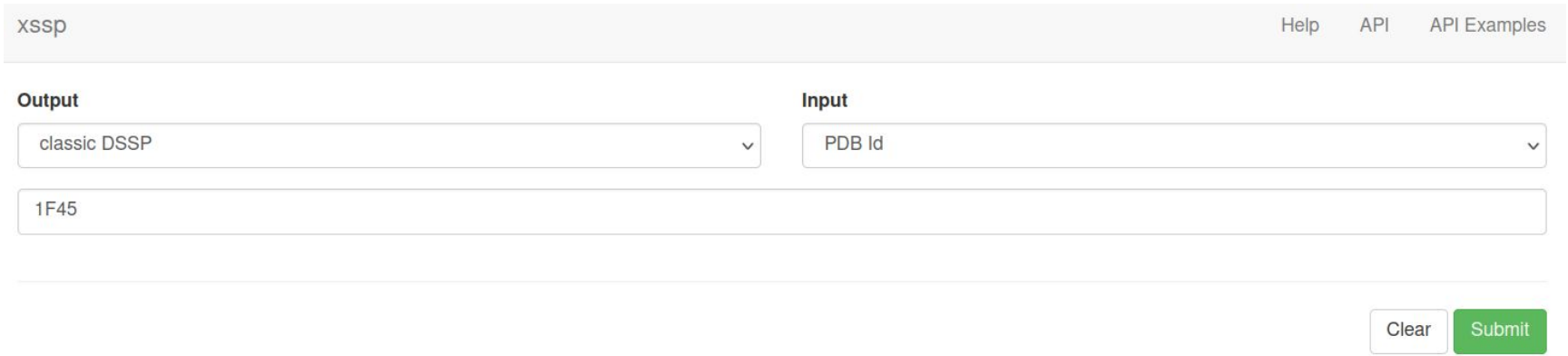
Strand | Helix | Coil | Disordered

Disordered, protein binding | Putative Domain Boundary | Membrane Interaction | Transmembrane Helix

Extracellular | Re-entrant Helix | Cytoplasmic | Signal Peptide

Get PNG | Get SVG

DSSP online server



```
  91 21.9   TOTAL NUMBER OF HYDROGEN BONDS IN ANTIPARALLEL BRIDGES, SAME NUMBER PER 100 RESIDUES
   2  0.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-5), SAME NUMBER PER 100 RESIDUES
   2  0.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-4), SAME NUMBER PER 100 RESIDUES
   2  0.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-3), SAME NUMBER PER 100 RESIDUES
   1  0.2   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-2), SAME NUMBER PER 100 RESIDUES
   1  0.2   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I-1), SAME NUMBER PER 100 RESIDUES
   0  0.0   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+0), SAME NUMBER PER 100 RESIDUES
   0  0.0   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+1), SAME NUMBER PER 100 RESIDUES
  34  8.2   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+2), SAME NUMBER PER 100 RESIDUES
  31  7.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+3), SAME NUMBER PER 100 RESIDUES
  85 20.5   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+4), SAME NUMBER PER 100 RESIDUES
   1  0.2   TOTAL NUMBER OF HYDROGEN BONDS OF TYPE O(I)-->H-N(I+5), SAME NUMBER PER 100 RESIDUES
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30     *** H
 0  0  0  1  0  0  0  0  0  0  0  0  0  1  1  0  0  0  0  0  0  0  0  0  0  2  0  0    RESIDUES PER ALPHA HELIX         .
 3  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    PARALLEL BRIDGES PER LADDER      .
 3  5  5  1  1  0  1  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    ANTIPARALLEL BRIDGES PER LADDER  .
 3  1  1  1  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0    LADDERS PER SHEET                .
#  RESIDUE AA STRUCTURE BP1 BP2  ACC     N-H-->O    O-->H-N    N-H-->O    O-->H-N    TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
  1    1 A I              0   0   99      0, 0.0     2,-0.3     0, 0.0    10,-0.2   0.000 360.0 360.0 360.0 115.7   15.3   27.3   15.4
  2    2 A W  E     -A   10   0A  78      8,-2.1     8,-2.5     0, 0.0     2,-0.6  -0.945 360.0-108.0-144.1 164.7   18.4   25.8   17.0
  3    3 A E  E     +A    9   0A 106     -2,-0.3     6,-0.2     6,-0.2     3,-0.1  -0.809  28.3 172.4 -95.4 122.0   20.5   25.3   20.1
  4    4 A L        -     0   0   31      4,-1.7    -1,-0.1    -2,-0.6     3,-0.1   0.662  67.8 -17.0 -85.3-112.1   23.8   27.2   20.4
  5    5 A K  S >  S-     0   0   91      1,-0.2     3,-1.5     2,-0.1     2,-0.3   0.043 105.0 -45.8 -84.2-165.8   25.1   26.6   23.9
  6    6 A K  T 3  S-     0   0  167      1,-0.3    -1,-0.2    -3,-0.1     3,-0.1  -0.509 127.5  -4.0 -66.9 121.9   23.2   25.3   27.0
  7    7 A D  T 3  S+     0   0   51     -2,-0.3    66,-1.7     1,-0.2     2,-0.4   0.754 112.0 109.3  67.8  23.6   19.9   27.1   27.4
  8    8 A V  E <   - b   0  73A  17     -3,-1.5    -4,-1.7    64,-0.2     2,-0.3  -0.997  44.3-173.6-134.8 140.9   20.5   29.4   24.4
  9    9 A Y  E     -Ab   3  74A  70     64,-1.9    66,-2.8    -2,-0.4     2,-0.4  -0.931  15.7-142.2-133.4 156.6   18.9   29.5   21.0
 10   10 A V  E     -Ab   2  75A   0     -8,-2.5    -8,-2.1    -2,-0.3     2,-0.5  -0.934   7.7-160.6-120.2 139.3   19.2   31.2   17.7
 11   11 A V  E     - b   0  76A   0     64,-2.0    66,-3.0    -2,-0.4     2,-0.4  -0.974  12.6-147.1-122.6 115.2   16.4   32.2   15.5
 12   12 A E  E     - b   0  77A  81     -2,-0.5     2,-0.4    64,-0.2    66,-0.2  -0.683  23.6-178.5 -80.5 125.5   17.2   32.8   11.8
 13   13 A L  E     - b   0  78A   6     64,-2.4    66,-1.3    -2,-0.4     2,-0.5  -0.979  36.1-129.8-135.9 131.8   15.2   35.5   10.2
```

The DSSP code

The output of DSSP is explained extensively under

- H = α-helix
- B = residue in isolated β-bridge
- E = extended strand, participates in β ladder
- G = 3-helix (3_{10} helix)
- I = 5 helix (π-helix)
- T = hydrogen bonded turn
- S = bend