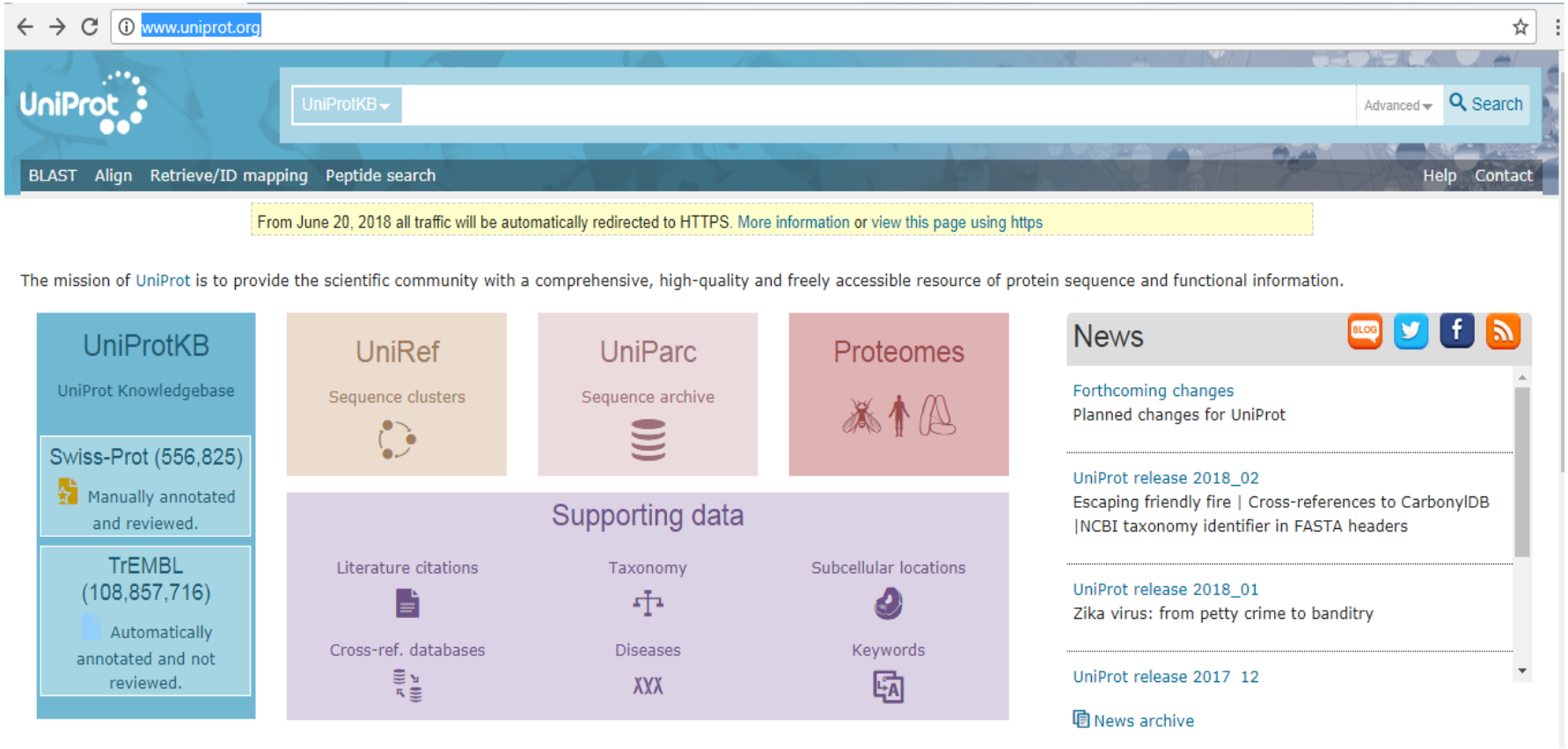
Dear students

Last week we started with the experiment *"Analysis of primary and secondary structure of the protein"*. I feel you all must have completed the theory part in your files from the literature provided to you. Today, I am sending you a presentation for the above mentioned topic. Kindly follow all the steps as shown in the slides. It will be easier for you to understand as the screenshots of the results are provided in each slide.

Primary Structure Analysis of Proteins

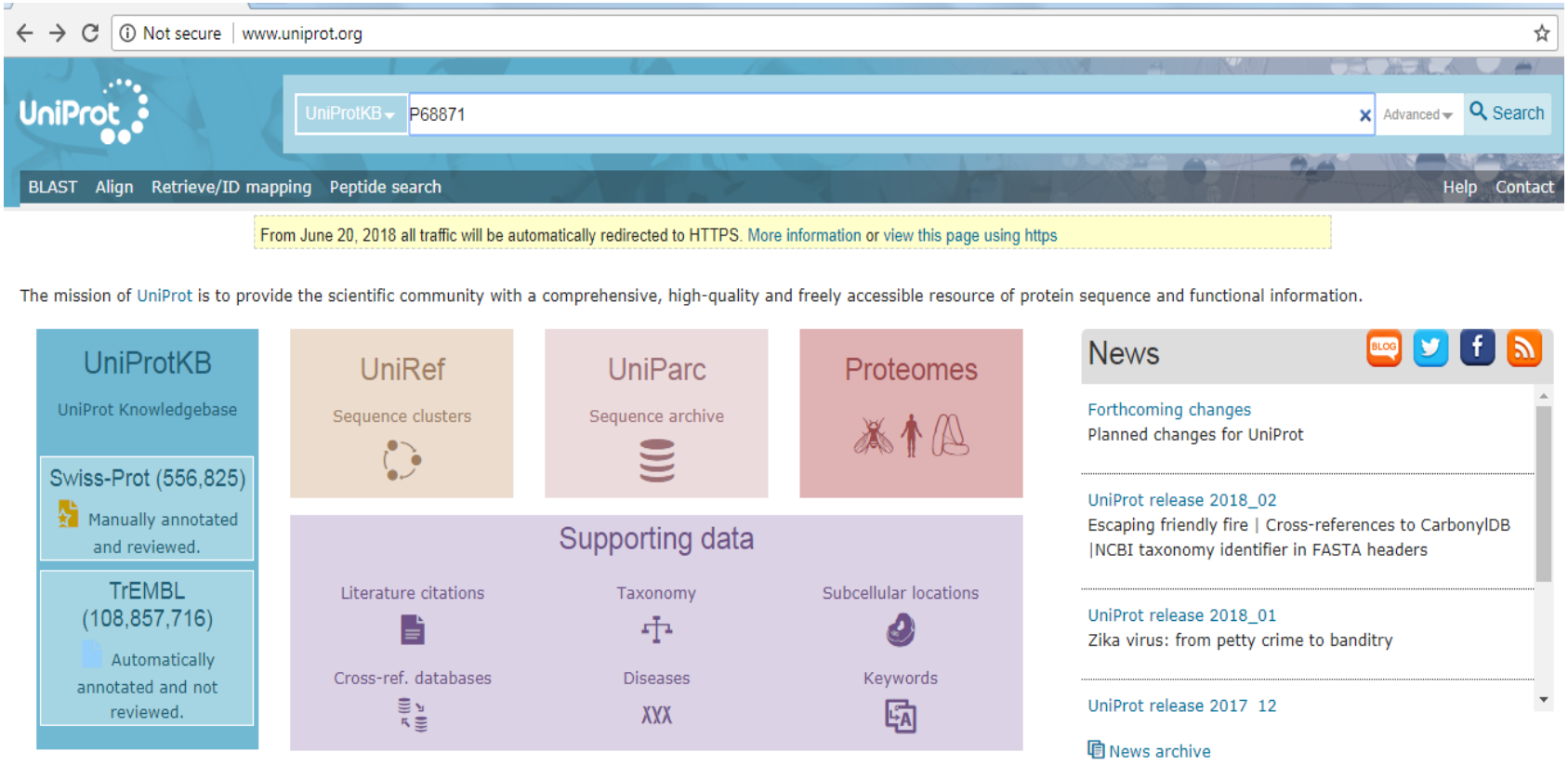
1st Step

Go to UniProt homepage by typing http://www.uniprot.org/ in the web browser



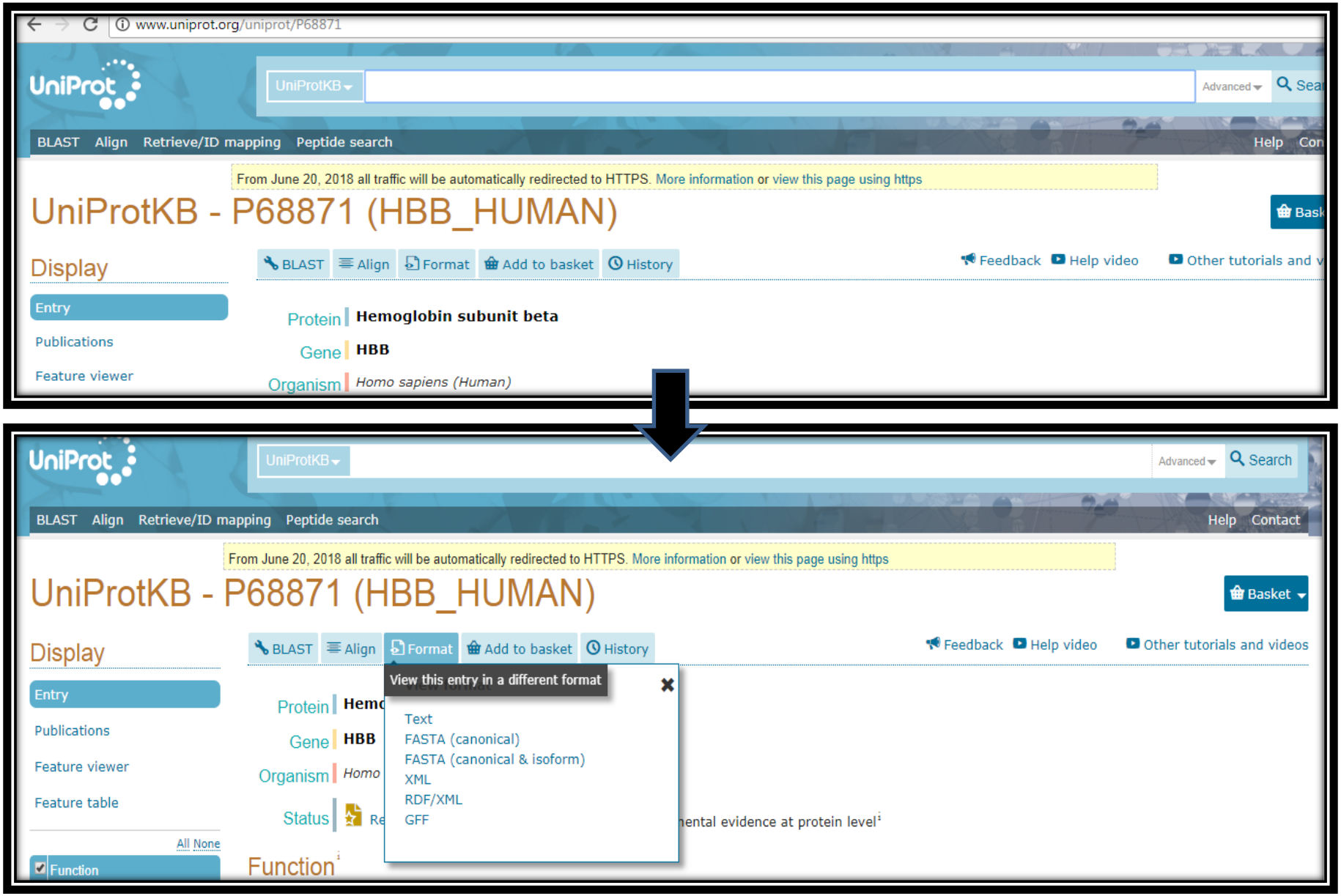
2nd Step

Type “P68871” in the search box and click “Go”



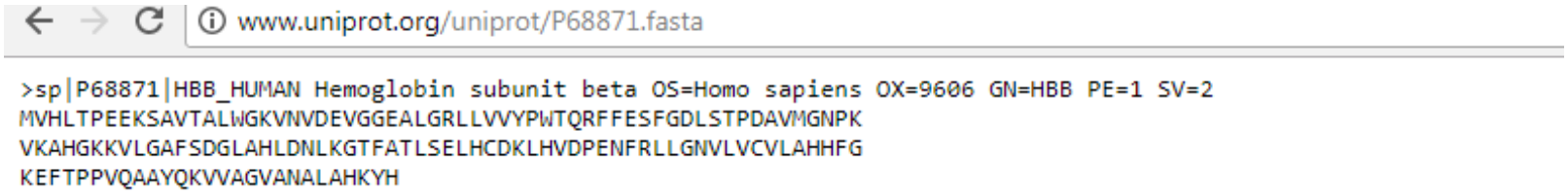
3rd Step

Click on “FASTA” format to retrieve the sequence



4th Step

Copy the sequence and paste it in any text document

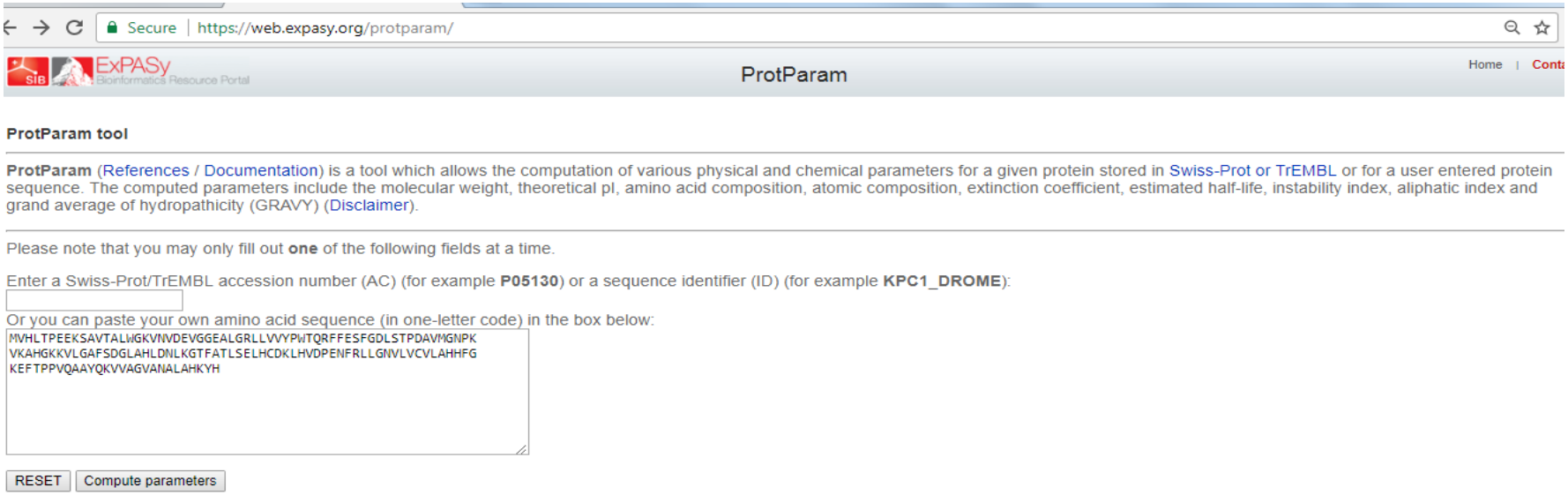


5th Step

Now open a new web browser and type https://www.expasy.ch/tools/protparam.html

6th Step

Paste the retrieved sequence in the text box given



7th Step

Click on “Compute parameters”.

8th Step

Analyze the result obtained.

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDAVMGNPK

        70         80         90        100        110        120
VKAHGKKVLG AFSDGLAHLD NLKGTFATLS ELHCDKLHVD PENFRLLGNV LVCVLAHHFG

       130        140
KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

References and documentation are available.

Number of amino acids: 147

Molecular weight: 15998.41

Theoretical pI: 6.74

Amino acid composition: CSV format

Ala (A)	15	10.2%
Arg (R)	3	2.0%
Asn (N)	6	4.1%
Asp (D)	7	4.8%
Cys (C)	2	1.4%
Gln (Q)	3	2.0%
Glu (E)	8	5.4%
Gly (G)	13	8.8%
His (H)	9	6.1%
Ile (I)	0	0.0%
Leu (L)	18	12.2%
Lys (K)	11	7.5%
Met (M)	2	1.4%

Total number of negatively charged residues (Asp + Glu): 15
Total number of positively charged residues (Arg + Lys): 14

Atomic composition:

Carbon	C	729
Hydrogen	H	1128
Nitrogen	N	196
Oxygen	O	202
Sulfur	S	4

Formula: $C_{729}H_{1128}N_{196}O_{202}S_4$
Total number of atoms: 2259

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient	15595
Abs 0.1% (=1 g/l)	0.975, assuming all pairs of Cys residues form cystines

Ext. coefficient	15470
Abs 0.1% (=1 g/l)	0.967, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 6.16
This classifies the protein as stable.

Aliphatic index: 93.47

Grand average of hydropathicity (GRAVY): 0.014

Secondary Structure Prediction of Proteins

1st Step

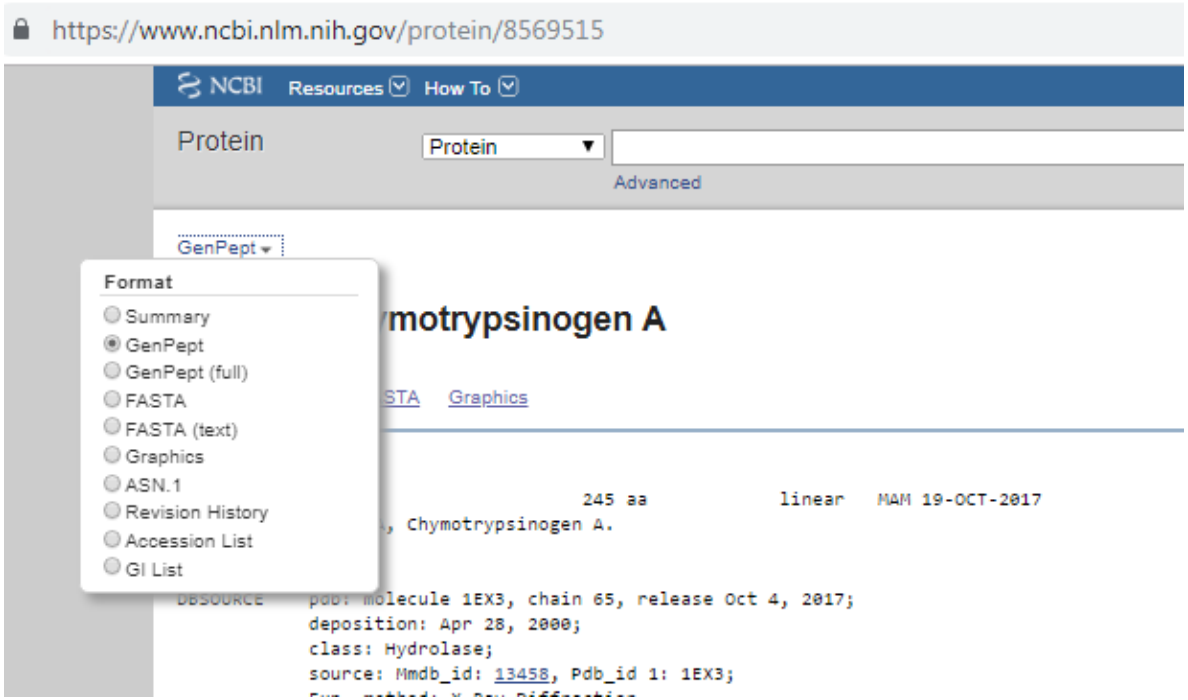
Go to NCBI homepage by typing http://www.ncbi.nlm.nih.gov in the web browser

2nd Step

Search “Protein” for “1EX3A” and press “GO”

3rd Step

Click on ‘FASTA’ from the format section



4th Step

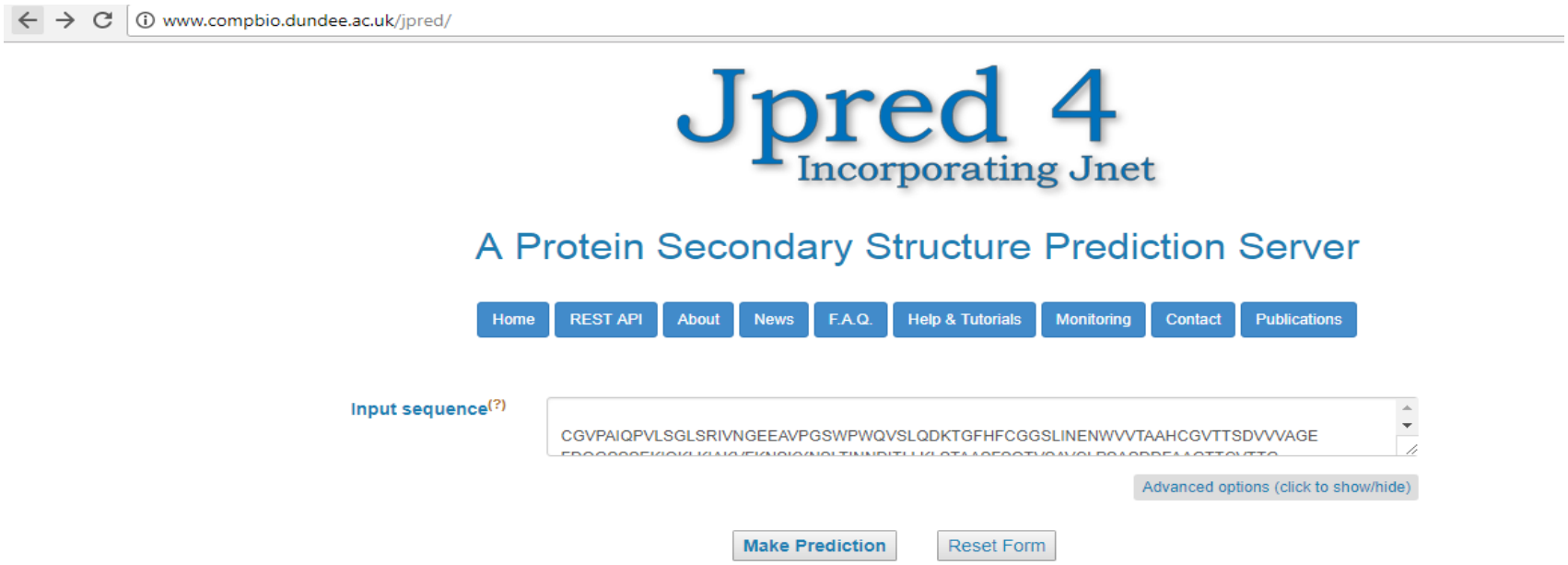
Select only protein sequence leaving the headers

5th Step

Go to www.compbio.dundee.ac.uk/jped

6th Step

In the sequence text box paste the amino acid sequence and click on “Make Prediction”.



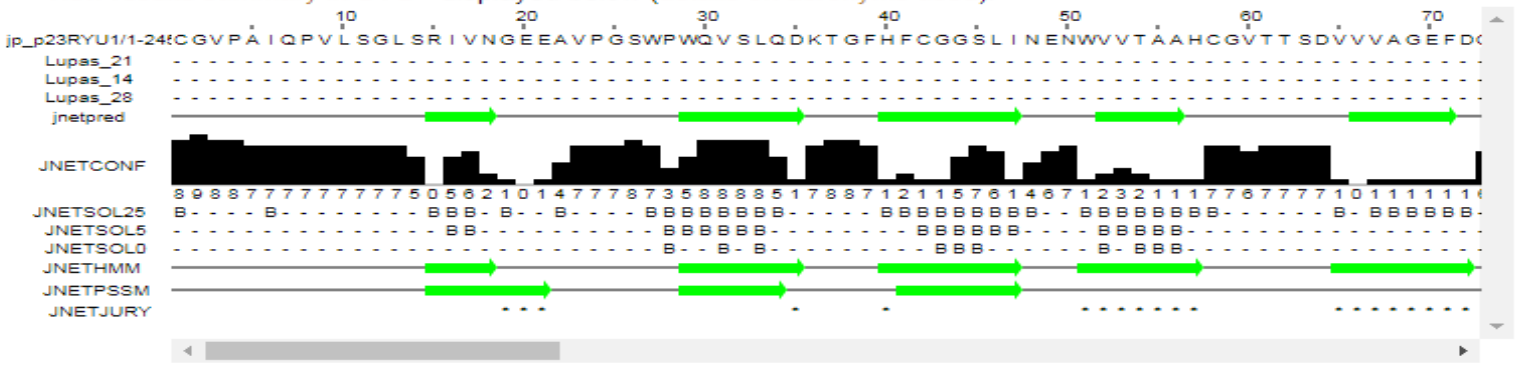
7th Step

Click continue and analyze the result

Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

- View results summary in SVG - displayed below (details on acronyms used):

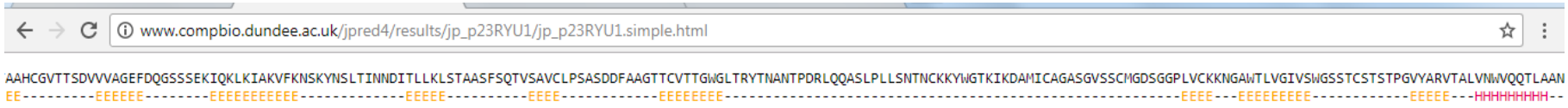


- View full results in HTML
- View simple results in HTML
- View results in PDF
- View results in Jalview (Link to a separate page with the Jalview Java Desktop application)
- View everything in a resu
- Get all files in TAR.GZ ar
- View results using in-bro

linked above)

- Lupas_21, Lupas_14, Lupas_28
  Coiled-coil predictions for the sequence. These are binary predictions for each location.
- Jnet Burial
  Prediction of Solvent Accessibility. levels are
  - *0 - Exposed*
  - *3 - 25% or more S.A. accessible*
  - *6 - 5% or more S.A. accessible*
  - *9 - Buried (<5% exposed)*
- *JNetPRED*
  The consensus prediction - helices are marked as red tubes, and sheets as dark green arrows.
- *JNetCONF*
  The confidence estimate for the prediction. High values mean high confidence. prediction - helices are marked as red tubes, and sheets as dark green arrows.
- *JNetALIGN*
  Alignment based prediction - helices are marked as red tubes, and sheets as dark green arrows.
- *JNetHMM*
  HMM profile based prediction - helices are marked as red tubes, and sheets as dark green arrows.
- *JNETPSSM*
  PSSM based prediction - helices are marked as red tubes, and sheets as dark green arrows.
- *JNETJURY*
  A '' in this annotation indicates that the JNETJURY was invoked to rationalise significantly different primary predictions.*

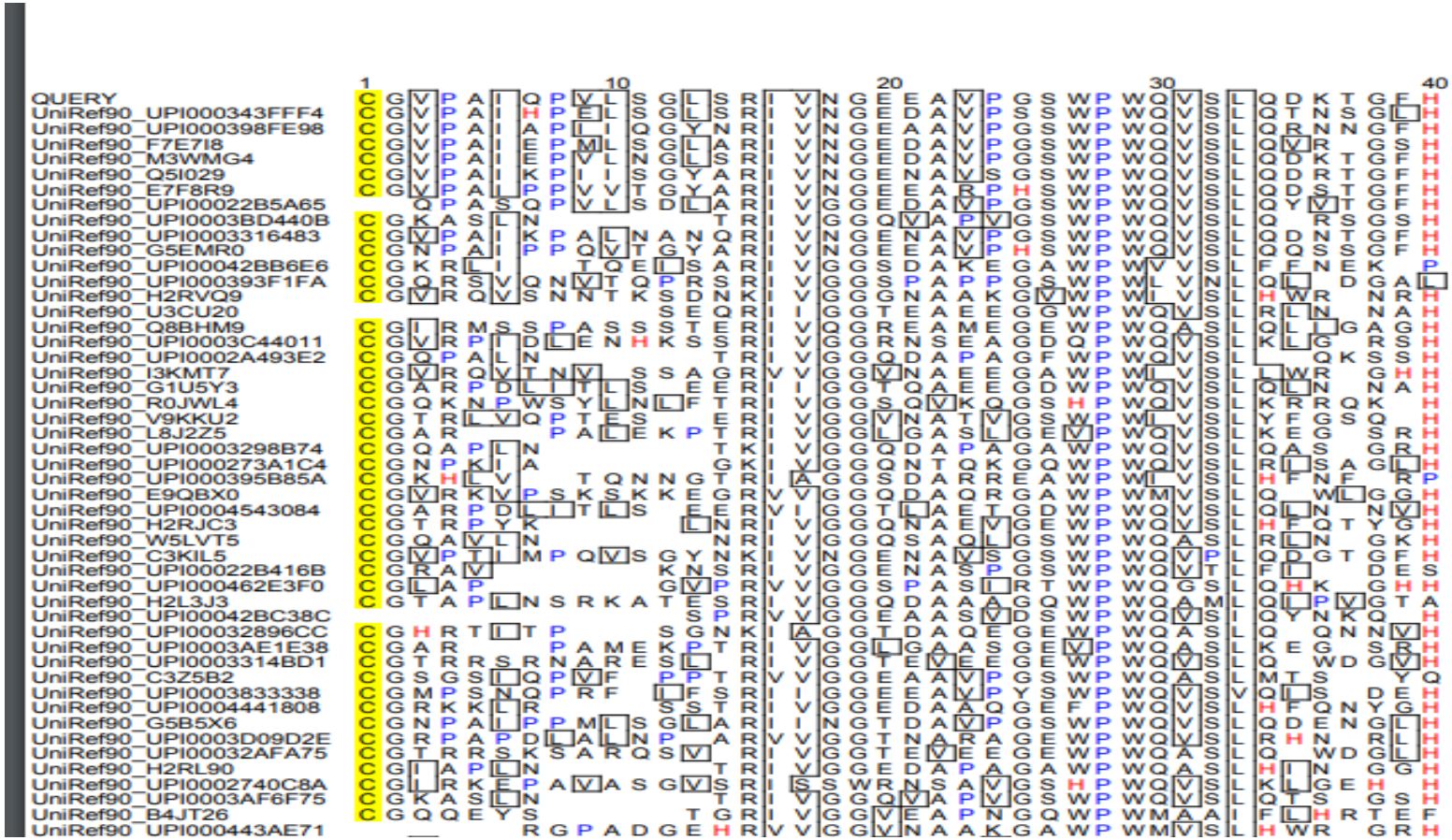
View simple result in HTML



The prediction is the definition of each amino acid residue into either of the secondary structures :

1) alpha helix ('H'),
2) beta sheet ('E') or
3) random coil ('-')

View result in PDF



Pictorial definitions used in POLYVIEW-2D for protein representation

Legend	Description
1 ———	*Amino acid residue numeration*
	Protein secondary structure
	H - α and other helices (view 1)
	H - α and other helices (view 2)
	E - β-strand or bridge
	C - coil
	Relative solvent accessibility (RSA)
 0 1 2 3 4 5 6 7 8 9	0 - completely buried (0-9% RSA), 9 - fully exposed (90-100% RSA)
	Physical-chemical properties
 HAPNC	H - hydrophobic: A,C,F,G,I,L,M,P,V A - amphipathic: H,W,Y P - polar: N,Q,S,T N/C - charged: D,E - neg; R,K - pos
	Confidence level of prediction
 0 1 2 3 4 5 6 7 8 9	0 - the lowest level, 9 - the highest level
GLCFEPFERL	*Transmembrane domain*