

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

← → ↻ ⓘ www.uniprot.org ☆ ⋮

UniProt

UniProtKB ▾ Advanced ▾ 🔍 Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

From June 20, 2018 all traffic will be automatically redirected to HTTPS. [More information](#) or [view this page using https](#)

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB
UniProt Knowledgebase

Swiss-Prot (556,825)
Manually annotated and reviewed.

TrEMBL (108,857,716)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

Literature citations
Cross-ref. databases

Taxonomy
Diseases
XXX

Subcellular locations
Keywords

News

BLOG Twitter Facebook RSS

Forthcoming changes
Planned changes for UniProt

UniProt release 2018_02
Escaping friendly fire | Cross-references to CarbonylDB
| NCBI taxonomy identifier in FASTA headers

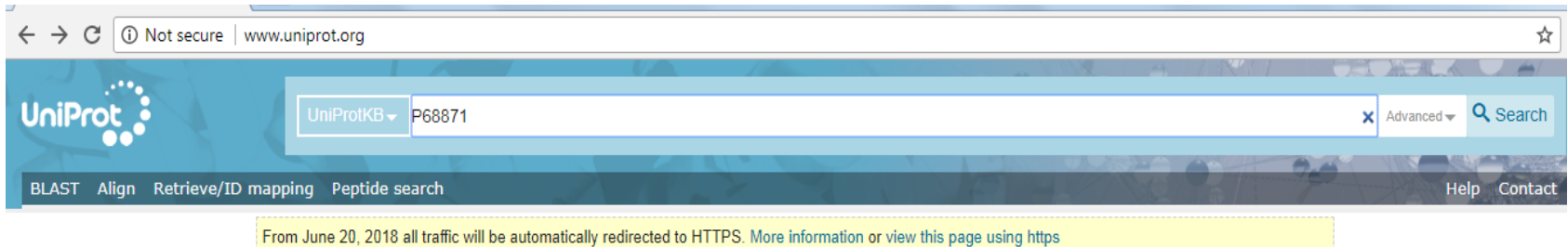
UniProt release 2018_01
Zika virus: from petty crime to banditry

UniProt release 2017_12

News archive

2nd Step

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



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

The image shows the main navigation area of the UniProt website. It features several colored tiles: a blue tile for "UniProtKB" (UniProt Knowledgebase) with "Swiss-Prot (556,825)" and "Manually annotated and reviewed."; a light blue tile for "TrEMBL (108,857,716)" with "Automatically annotated and not reviewed."; a tan tile for "UniRef" (Sequence clusters); a light purple tile for "UniParc" (Sequence archive); a red tile for "Proteomes" with icons of a fly, a person, and a protein structure; and a large purple tile for "Supporting data" containing links for "Literature citations", "Cross-ref. databases", "Taxonomy", "Diseases", "Subcellular locations", and "Keywords". To the right is a "News" section with social media icons (BLOG, Twitter, Facebook, RSS) and a list of news items: "Forthcoming changes", "Planned changes for UniProt", "UniProt release 2018_02" (Escaping friendly fire | Cross-references to CarbonylDB | NCBI taxonomy identifier in FASTA headers), "UniProt release 2018_01" (Zika virus: from petty crime to banditry), and "UniProt release 2017_12". A "News archive" link is at the bottom.

3rd Step

Click on “FASTA” format to retrieve the sequence

The screenshot shows the UniProtKB entry page for P68871 (HBB_HUMAN). The browser address bar displays 'www.uniprot.org/uniprot/P68871'. The UniProt logo is in the top left. A search bar with 'UniProtKB' is in the top right. Below the search bar is a navigation menu with 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. A yellow banner at the top states: 'From June 20, 2018 all traffic will be automatically redirected to HTTPS. More information or view this page using https'. The main title is 'UniProtKB - P68871 (HBB_HUMAN)'. Below the title is a 'Display' section with buttons for 'BLAST', 'Align', 'Format', 'Add to basket', and 'History'. The 'Format' button is highlighted with a blue border. To the right of the 'Display' section are links for 'Feedback', 'Help video', and 'Other tutorials and v'. Below the 'Display' section are tabs for 'Entry', 'Publications', and 'Feature viewer'. The 'Entry' tab is selected, showing 'Protein | Hemoglobin subunit beta', 'Gene | HBB', and 'Organism | Homo sapiens (Human)'. A large black arrow points from the 'Format' button in this screenshot to the 'Format' button in the next screenshot.

The screenshot shows the UniProtKB entry page for P68871 (HBB_HUMAN) with the 'Format' dropdown menu open. The browser address bar displays 'www.uniprot.org/uniprot/P68871'. The UniProt logo is in the top left. A search bar with 'UniProtKB' is in the top right. Below the search bar is a navigation menu with 'BLAST', 'Align', 'Retrieve/ID mapping', and 'Peptide search'. A yellow banner at the top states: 'From June 20, 2018 all traffic will be automatically redirected to HTTPS. More information or view this page using https'. The main title is 'UniProtKB - P68871 (HBB_HUMAN)'. Below the title is a 'Display' section with buttons for 'BLAST', 'Align', 'Format', 'Add to basket', and 'History'. The 'Format' button is highlighted with a blue border, and its dropdown menu is open, showing a list of formats: 'Text', 'FASTA (canonical)', 'FASTA (canonical & isoform)', 'XML', 'RDF/XML', and 'GFF'. The 'FASTA (canonical)' option is selected. To the right of the 'Display' section are links for 'Feedback', 'Help video', and 'Other tutorials and videos'. Below the 'Display' section are tabs for 'Entry', 'Publications', and 'Feature viewer'. The 'Entry' tab is selected, showing 'Protein | Hemoglobin subunit beta', 'Gene | HBB', and 'Organism | Homo sapiens (Human)'. A large black arrow points from the 'Format' button in the previous screenshot to the 'Format' button in this screenshot.

4th Step

Copy the sequence and paste it in any text document

← → ↻ ⓘ www.uniprot.org/uniprot/P68871.fasta

```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMGNPK
VKAHGKKVLGAFSDGLAHLNLIKGTFAITSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH
```

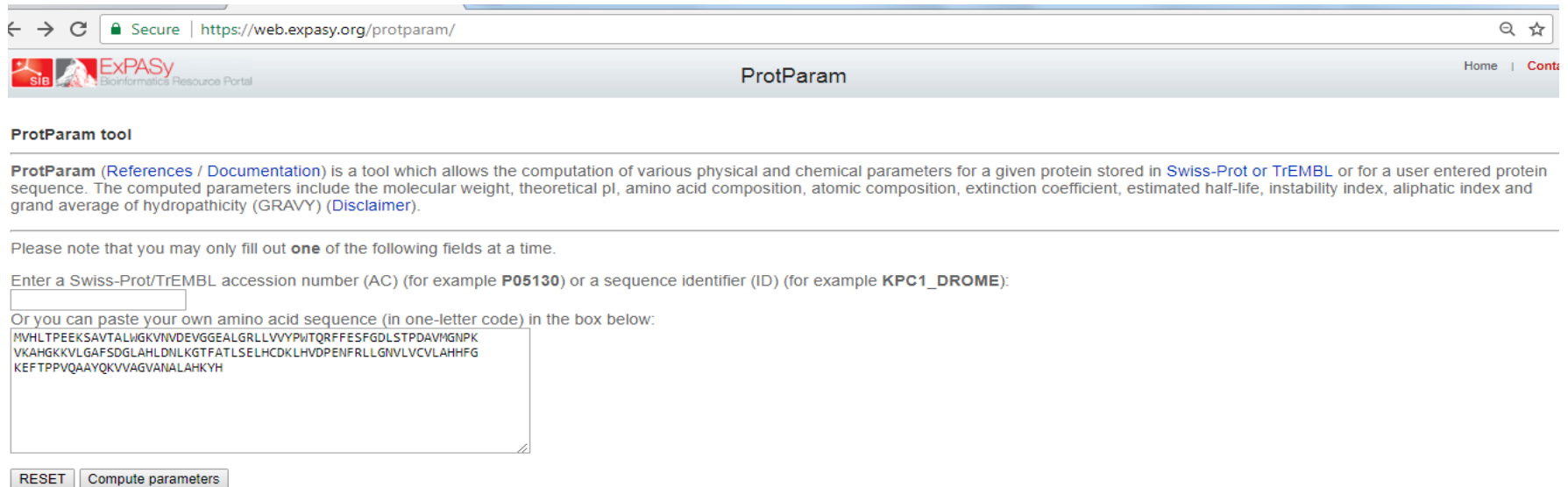
5th Step

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



The screenshot shows a web browser window with the URL <https://web.expasy.org/protparam/>. The page title is "ProtParam" and the logo for SIB and Expasy is visible. The main heading is "ProtParam tool". Below this, there is a description of the tool: "ProtParam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer)." A note states: "Please note that you may only fill out **one** of the following fields at a time." There are two input fields: "Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):" and "Or you can paste your own amino acid sequence (in one-letter code) in the box below:". The second field contains the amino acid sequence: MVHLTPEEKSAVTALWGKVIVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMVGNPK VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLVHDPENFRLLGNVLVCLAHHFG KEFTPPVQAAYQKVVAGVANALAHKYH. At the bottom, there are two buttons: "RESET" and "Compute parameters".

7th Step

Click on “Compute parameters”.

8th Step

Analyze the result obtained.

ProtParam

User-provided sequence:

```
10      20      30      40      50      60
MVHLTPEEK AVTALWGKVN VDEGGGALG RLLVVPWTQ RFFESFGDLS TPDAMVGNPK

70      80      90      100     110     120
VKAHGKKVLG AFSDDLHLN NLKGTFFATLS ELHCCKLHVD 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

The screenshot shows the NCBI protein page for Chymotrypsinogen A (PDB ID: 1EX3). The URL is <https://www.ncbi.nlm.nih.gov/protein/8569515>. The page features a search bar with 'Protein' selected and a search button. A dropdown menu is open, showing the 'Format' section with the following options: Summary, GenPept (selected), GenPept (full), FASTA, FASTA (text), Graphics, ASN.1, Revision History, Accession List, and GI List. The main content area displays the protein name 'Chymotrypsinogen A' and its properties: 245 aa, linear, MAM 19-OCT-2017. The source information is: `DBSOURCE pdb: molecule 1EX3, chain 65, release Oct 4, 2017; deposition: Apr 28, 2000; class: Hydrolase; source: Mmdb_id: 13458, Pdb_id 1: 1EX3;`

4th Step

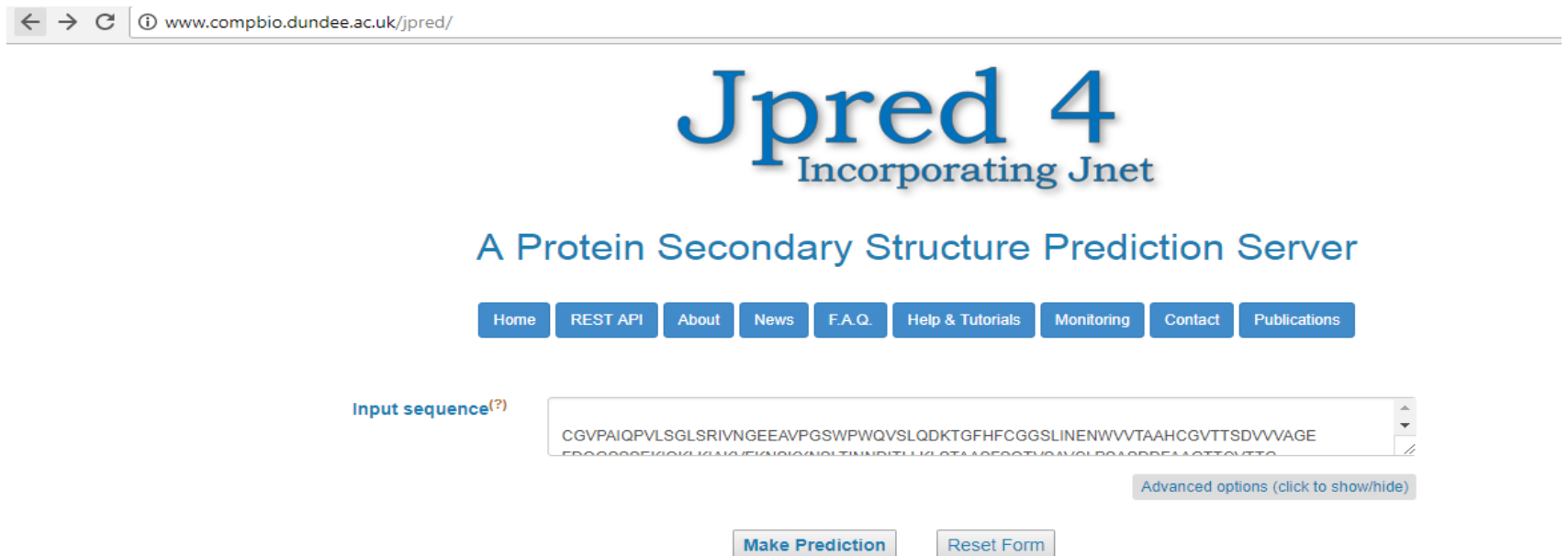
Select only protein sequence leaving the headers

5th Step

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

6th Step

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



The screenshot shows a web browser window with the address bar displaying www.compbio.dundee.ac.uk/jpred/. The main heading is "Jpred 4 Incorporating Jnet". Below this is the subtitle "A Protein Secondary Structure Prediction Server". A navigation menu contains buttons for Home, REST API, About, News, F.A.Q., Help & Tutorials, Monitoring, Contact, and Publications. The "Input sequence" field is highlighted in orange and contains the amino acid sequence: CGVPAIQPVL SGLSRVNGEEAVPGSWPQVSLQDKTGFHFCGGSLINENWVWVTAAHCGVTTSDVWVAGE. Below the input field is a button labeled "Advanced options (click to show/hide)". At the bottom of the form are two buttons: "Make Prediction" and "Reset Form".

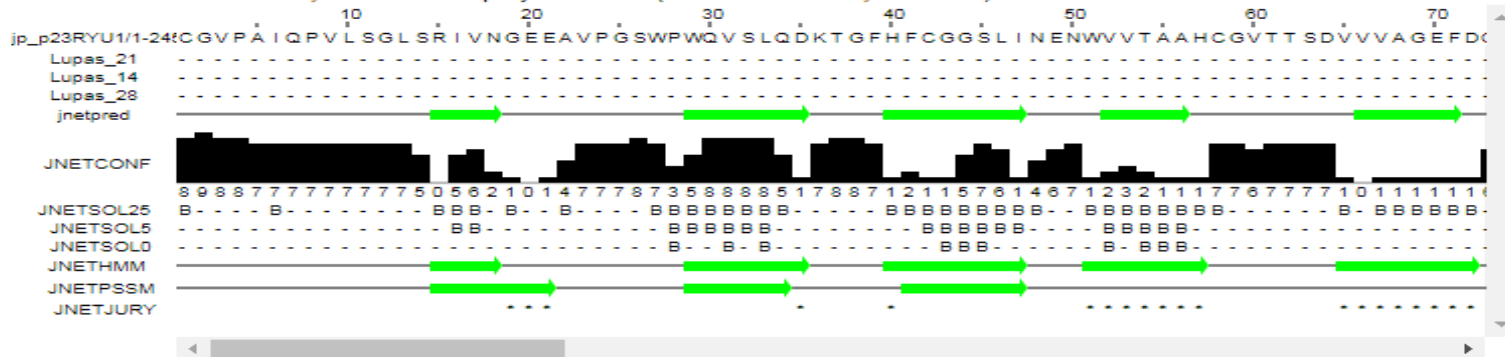
7th Step

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.

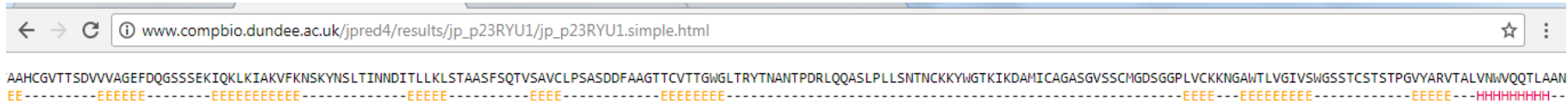
- View results summary in SVG - displayed below (details on acronyms used):



- 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 result viewer
- Get all files in TAR.GZ archive
- View results using in-browser viewer (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.

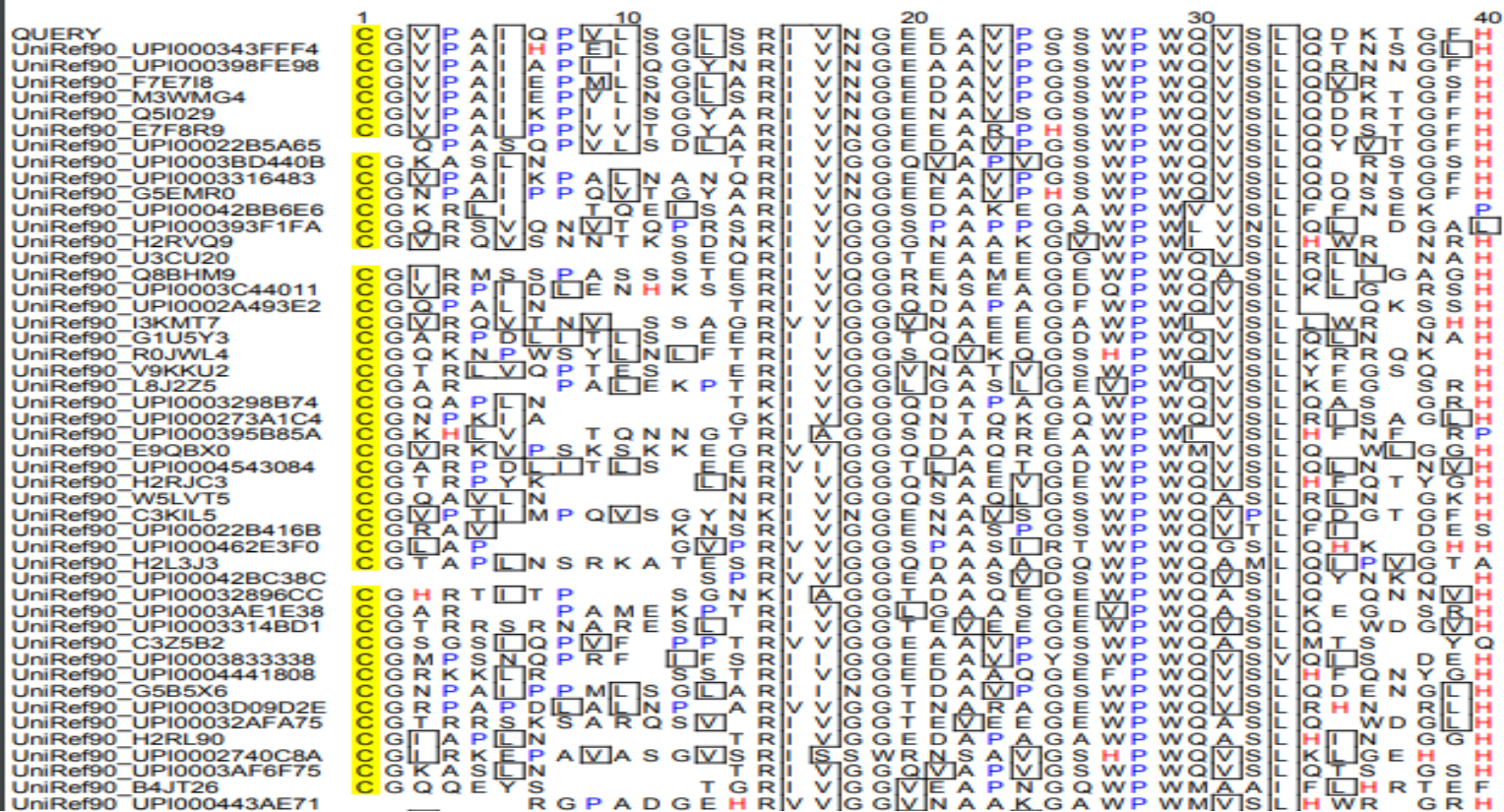
View simple result in HTML










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