

## WEBLEM 1

## SECONDARY STRUCTURE PREDICTION

**Introduction:**

- ➔ Secondary structure prediction is relatively accurate, and is in fact much easier to solve than three-dimensional structure prediction. The accuracy of assigning strand, helix or loops to a certain residue can go up to 80% with the most reliable methods. Typically, these methods use periodicity in the sequence combined with phi and psi angle preferences of certain amino acids types to come to accurate predictions. The real challenge lies in assembling the secondary structure element in a correct topology. Nevertheless, secondary structure prediction may be used to assess the quality of model built with a structure prediction method. Many methods also incorporated secondary structure information during homology detection.

**Secondary structure prediction has three generations:****1. First Generation: Chou and Fasman**

- ➔ CFSSP (Chou & Fasman Secondary Structure Prediction Server) is an online protein secondary structure prediction server. This server predicts regions of secondary structure from the protein sequence such as alpha helix, beta sheet, and turns from the amino acid sequence. The output of predicted secondary structure is also displayed in linear sequential graphical view based on the probability of occurrence of alpha helix, beta sheet, and turns. The method implemented in CFSSP is Chou-Fasman algorithm, which is based on analyses of the relative frequencies of each amino acid in alpha helices, beta sheets, and turns based on known protein structures solved with X-ray crystallography. CFSSP is freely accessible via ExPASy server.

**2. Second Generation: GOR IV**

- ➔ The GOR method is based on information theory and was developed by J.Garnier, D.Osguthorpe and B.Robson (J.Mol.Biol.120,97, 1978). The present version, GOR IV, uses all possible pair frequencies within a window of 17 amino acid residues and is reported by J. Garnier. J.F. Gibrat and B.Robson in Methods in Enzymology, vol 266, p 540-553 (1996). After cross validation on a data base of 267 proteins, the version IV of GOR has a mean accuracy of 64.4% for a three-state prediction (Q3). The program gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil; the second gives the probability values for each secondary structure at each amino acid position. The predicted secondary structure is the one of highest probability compatible with a predicted helix segment of at least four residues and a predicted extended segment of at least two residues.

**3. Third Generation: PSIPRED**

- ➔ 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.

## References:

- ➔ Protein Secondary Structure - an overview | ScienceDirect Topics. (2016). Science Direct. <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/protein-secondary-structure>
- ➔ Edwards, Y. J., and Cottage, A. 2003. Bioinformatics methods to predict protein structure and function. A practical approach. Mol. Biotechnol. 23:139-66.
- ➔ Ashok Kumar, T. (2013). CFSSP: Chou and Fasman Secondary Structure Prediction server. WIDE SPECTRUM: Research Journal. 1(9):15-19.
- ➔ Heringa J. 2002. Computational methods for protein secondary structure prediction using multiple sequence alignments. Curr. Protein Pept. Sci. 1:273-301.
- ➔ Lehnert, U., Xia, Y., Royce, T. E., Goh, C. S., Liu, Y., Senes, A., Yu. H., Zhang, Z. L., Engelman, D. M, and Gerstein M. 2004. Computational analysis of membrane proteins: Genomic occurrence, structure prediction and helix interactions. Q. Rev. Biophys. 37:121-46.
- ➔ Moller, S., Croning, M. D. R., and Apweiler, R. 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17:646-53.

## WEBLEM 1A

### SECONDARY STRUCTURE PREDICTION

#### Aim:

- Secondary Structure prediction for CAD13\_HUMAN (Cadherin) using various tools (CF, GOR IV, PSIPRED)

#### Introduction:

- Secondary Structure prediction has three generations:

##### 1. First Generation: Chou and Fasman

- CFSSP (Chou & Fasman Secondary Structure Prediction Server) is an online protein secondary structure prediction server. This server predicts regions of secondary structure from the protein sequence such as alpha helix, beta sheet, and turns from the amino acid sequence. The output of predicted secondary structure is also displayed in linear sequential graphical view based on the probability of occurrence of alpha helix, beta sheet, and turns.

##### 2. Second Generation: GOR IV

- The GOR method is based on information theory and was developed by J.Garnier, D.Osguthorpe and B.Robson. After cross validation on a data base of 267 proteins, the version IV of GOR has a mean accuracy of 64.4% for a three-state prediction. The program gives two outputs, one eye-friendly giving the sequence and the predicted secondary structure in rows, H=helix, E=extended or beta strand and C=coil; the second gives the probability values for each secondary structure at each amino acid position.

##### 3. Third Generation: PSIPRED

- 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

#### Cadherin:

- Cadherins are a large family of cell-cell adhesion molecules acting in a homotypic, homophilic manner that play an important role in the maintenance of tissue integrity. In the human kidney several members of the cadherin family are expressed in a controlled spatiotemporal pattern. Cadherin-16 also called kidney-specific cadherin, is exclusively expressed in epithelial cells of the adult kidney.

#### Methodology:

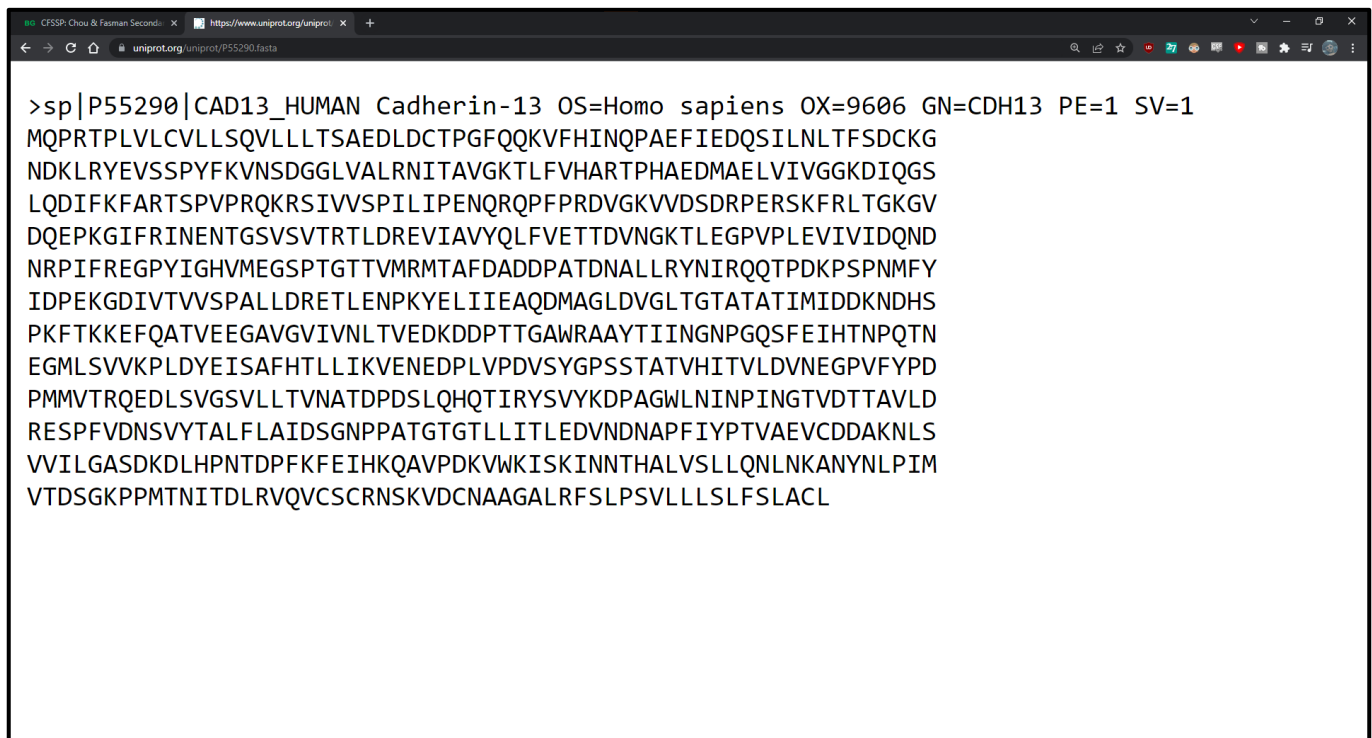
- Open Homepage of (CF, GOR IV, PSIPRED)
- Enter the query Cadherin Fasta sequence in the search bar.
- Open the Result page.
- Interpret the Results.

## Observations:

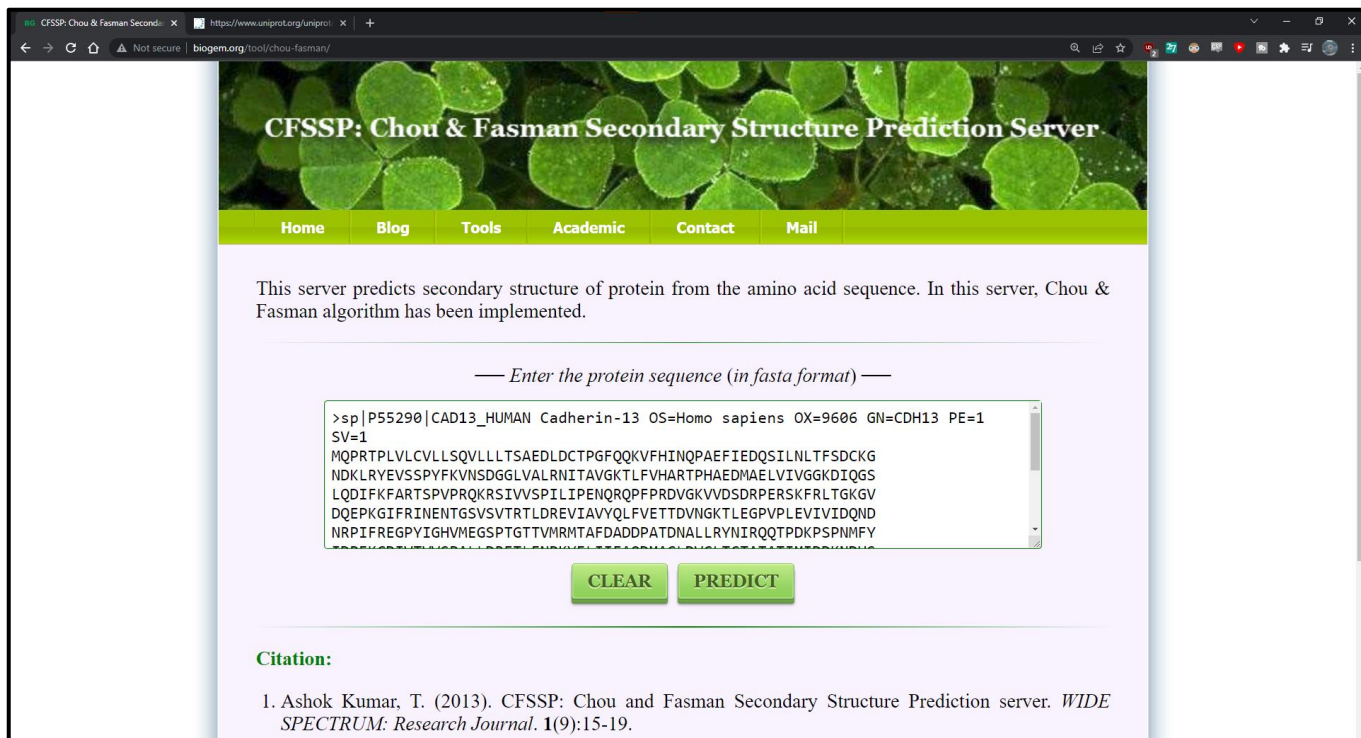
### → First Generation: Chou and Fasman



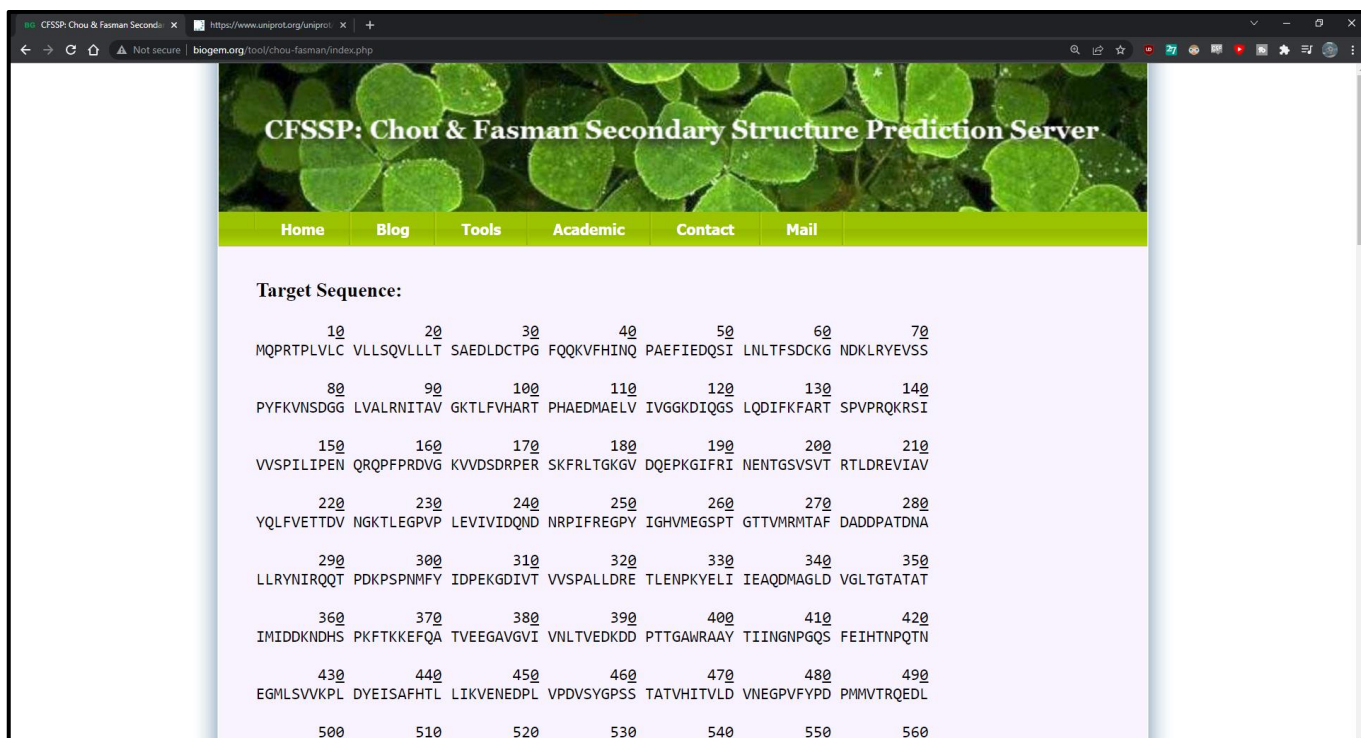
**Fig1: Homepage of Chou and fasman**



**Fig2. FASTA sequence of CAD13\_HUMAN Cadherin**



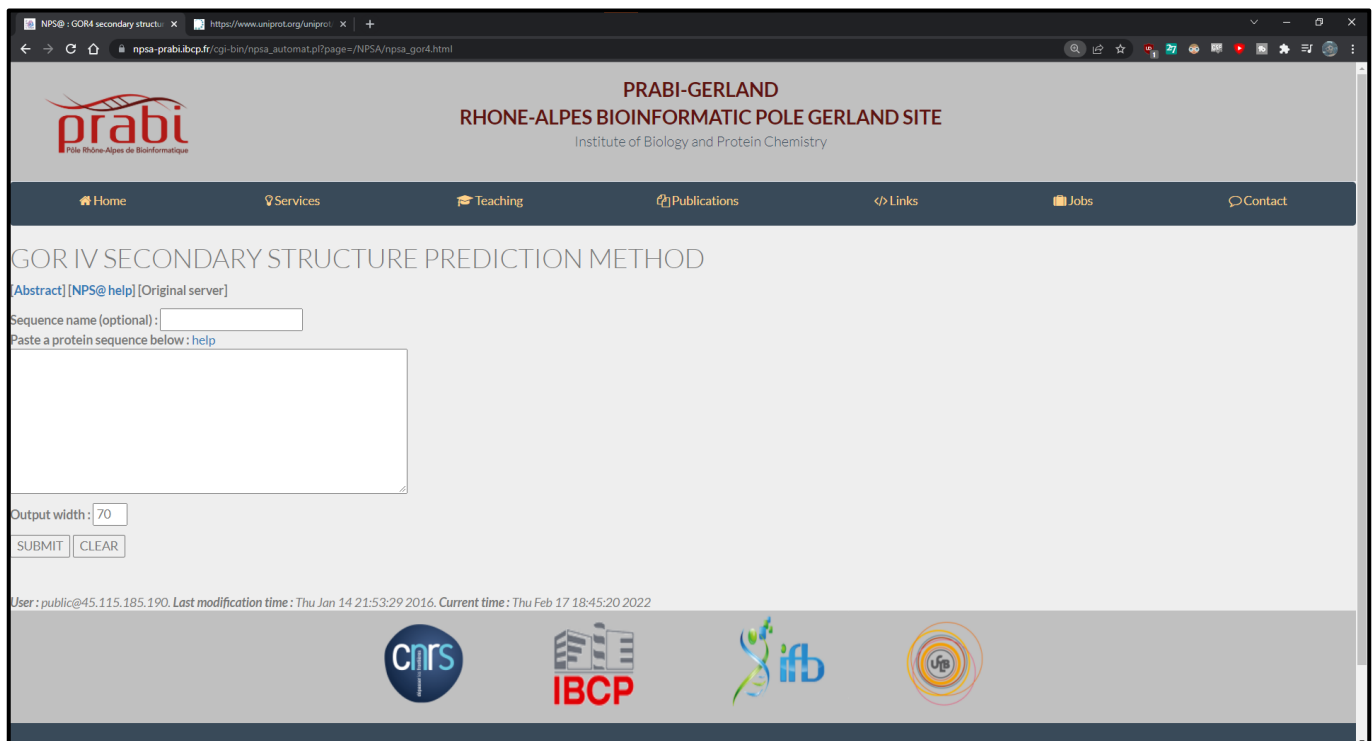
**Fig3. Search bar with FASTA sequence of CAD13\_HUMAN Cadherin**



**Fig4. Result page of Cadherin showing target sequence**



### → Secondary Generation: GOR IV



```

>sp|P55290|CAD13_HUMAN Cadherin-13 OS=Homo sapiens OX=9606 GN=CDH13 PE=1 SV=1
MQPRTPLVLCVLLSQVLLLSAEDLDCTPGFQQKVFHINQPAEFIEDQSILNLTFSACKG
NDKLRVEVSSPYFKVNSDGGGLVALRNITAVGKTLFVHARTPHAEDMAELVIVGGKDIQGS
LQDIFKFARTSPVPRQKRSIVVSPILIPENQRQPFPRDVGKVVSDRPERSKFRLTGKGV
DQEPKGFIRINENTGSVSVTRTLDREVIAYVQLFVETTDVNGKTEGVPVLEVIDQND
NRPIFREGPYIGHVMEGSPTGTTVMRMTAFDADDPATDNALLRYNIRQQTPDKPSNMFY
IDPEKGDIVTVVSPALLDRETLENPKYELIEAQDMAGLDVGLTGTATATIMIDDKNDHS
PKFTKKEFQATVEEGAVGVIVNLTVEDKDDPTTGAWRAAYTIINGNPGQSFEIHTNPQTN
EGMLSVMKPLDYEISAFHTLLIKVENEDPLVPDVSYPSSSTATVHITVLDVNEGPVFYFD
PMMVTRQEDLSVGSVLLTVNATDPDSLQHQTIKRSVYKDPAGWLNINPINGTVDTTAVLD
RESPVDNSVYTALFLAIDSGNPPATGTGTLITLEDVNDNAPFIYPTVAEVCDDAKNLS
VVILGASDKDLHPNTDPFKFEIHKQAVPDVWVKISKINNTHALVSLQLNKNKANYNLPIM
VTDGSKPPMTNITDLRVQVCSRSKVDCAAGALRFSVLLLSLFLACL

```

**Fig2. FASTA sequence of CAD13\_HUMAN Cadherin**

**PRABI-GERLAND**  
RHONE-ALPES BIOINFORMATIC POLE GERLAND SITE  
Institute of Biology and Protein Chemistry

Home Services Teaching Publications Links Jobs Contact

GOR IV SECONDARY STRUCTURE PREDICTION METHOD

[Abstract](#) | [NPS@ help](#) | [Original server](#)

Sequence name (optional):

Paste a protein sequence below: [help](#)

```

PMMVTRQEDLSVGSVLLTVNATDPDSLQHQTIKRSVYKDPAGWLNIN
PINGTVDTTAVLD
RESPVDNSVYTALFLAIDSGNPPATGTGTLITLEDVNDNAPFIYPTVA
EVCDDAKNLS
VVILGASDKDLHPNTDPFKFEIHKQAVPDVWVKISKINNTHALVSLQ
NLNKNANYNLPIM
VTDGSKPPMTNITDLRVQVCSRSKVDCAAGALRFSVLLLSLFLACL

```

Output width:

User: public@45.115.185.190. Last modification time: Thu Jan 14 21:53:29 2016. Current time: Thu Feb 17 18:45:20 2022

**Fig3. Search bar with FASTA sequence of cadherin**



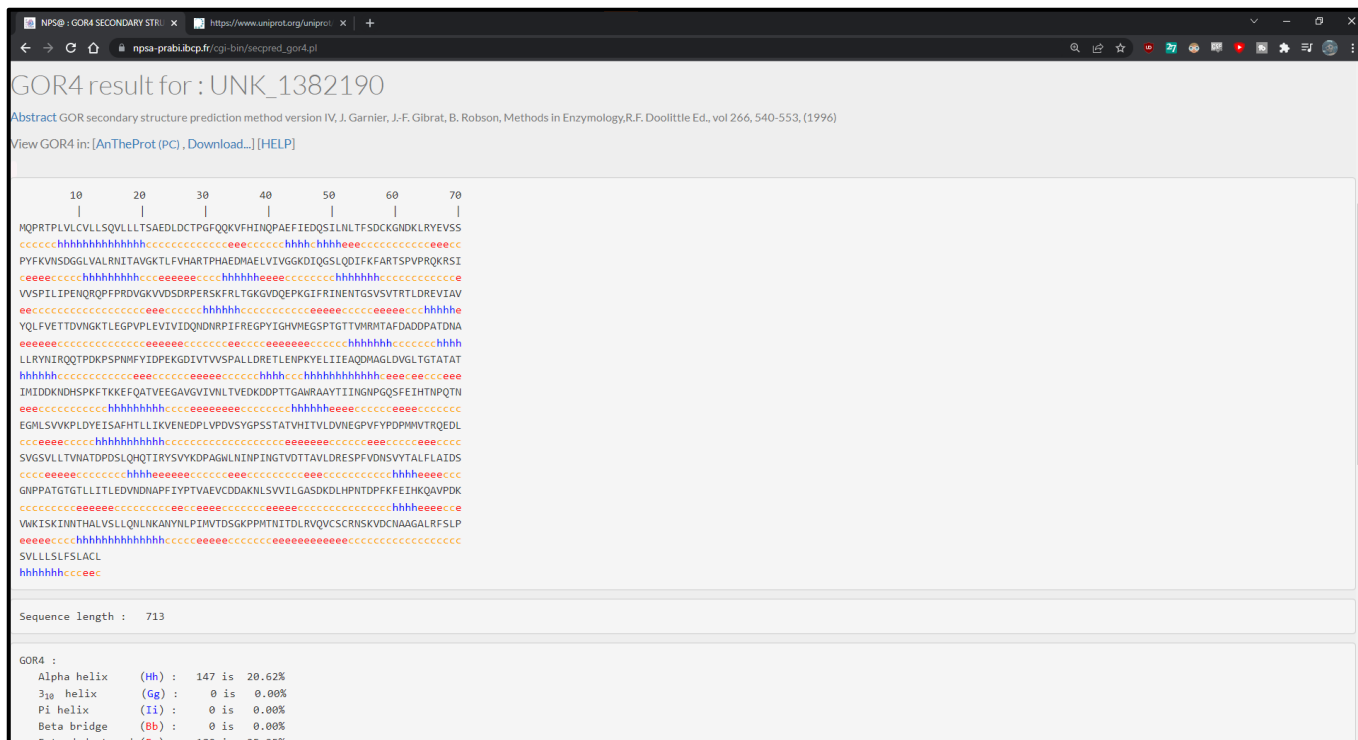


Fig4. Result page of GOR IV for Cadherin

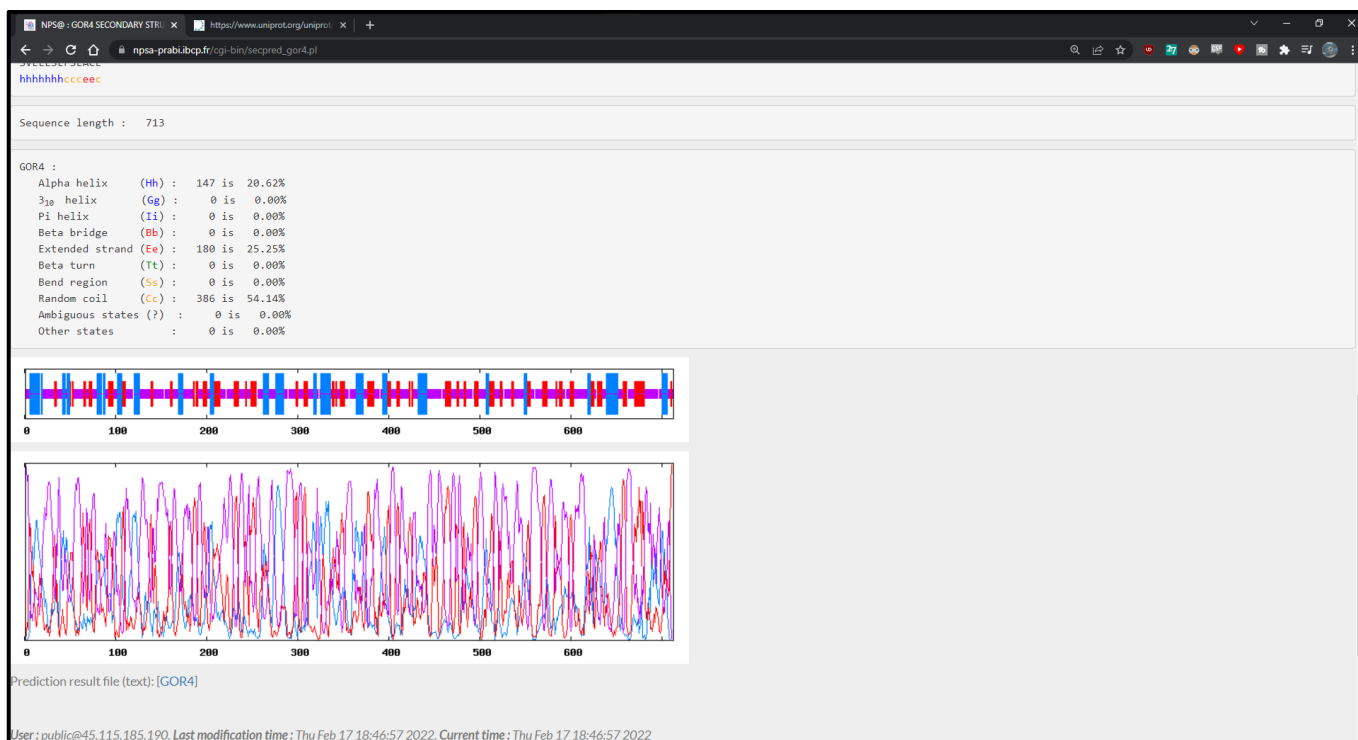


Fig5. Frequency graph for Cadherin



## → Third Generation: PSIPRED

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)

**Fig1. Homepage for PSIPRED**

```
>sp|P55290|CAD13_HUMAN Cadherin-13 OS=Homo sapiens OX=9606 GN=CDH13 PE=1 SV=1
MQPRTPLVLCVLLSQVLLLTSAEDLDCTPGFQQKVFHINQPAEFIEDQSILNLTFSACKG
NDKLRYEVSSPYFKVNSDGGGLVALRNITAVGKTLFVHARTPHAEDMAELVIVGGKDIQGS
LQDIFKFARTSPVPRQKRSIVSPILIPENQRQPFPRDVGKVVSDRPERSKFRLTGKGV
DQEPKGFIRINENTGSVSVTRTLDREVIAYVQLFVETTDVNGKTEGVPVLEVIVIDQND
NRPIFREGPYIGHVMEGSPTGTTVMRMTAFDADDPATDNALLRYNIRQQTPDKPSPNMFY
IDPEKGDIVTVVSPALLDRETLENPKYELIEAQDMAGLDVGLTGTATATIMIDKNDHS
PKFTKKEFQATVEEGAVGVIVNLTVEDKDDPTTGAWRAAYTIINGNPGQSFEIHTNPQTN
EGMLSVMKPLDYEISAFHTLLIKVENEDPLVPDVSYGPSSTATVHITVLDVNEGPVFYPD
PMMVTRQEDLSVGSVLLTVNATDPDSLQHQITIRYSVYKDPAGWLNINPINGTVDTTAVLD
RESPFDNSVYTALFLAIDSGNPPATGTGTLITLEDVNDNAPFIYPTVAEVCDDAKNLS
VVILGASDKDLHPNTDPFKFEIHKQAVPDKVWKISKINNTALVSLQLNKNKANYNLPIM
VTDSGKPPMTNITDLRVQVCSCRNSKVCNAAGALRFSLPVLLLSLFSLACL
```

**Fig2. FASTA sequence of CAD13\_HUMAN Cadherin**

Domain Prediction

☐ DomPred (Protein Domain Prediction)

Function Prediction

☐ FFPred 3 (Eukaryotic Function Prediction)

[Help...](#)

**Submission details**

**Protein Sequence**

```
MQPRTPLVLCVLLSQVLLLSAEDLDCTPGFQQKVFHINQPAEFIEDQSILNLTFSCKG
NDKLRVEVSSPYFKVNSDGLVALRNITAVGKTLFVHARTPHAEDMAELVVGKDIQGS
LQDIFKFARTSPVPROKRSIVVSPILPENQRQPFPRDVGKVVSDRPERSKFRLTGKGV
```

[Help...](#)  
If you wish to test these services follow this link to retrieve a [test fasta sequence](#).

**Job name**

sp|P55290|CAD13\_HUMAN

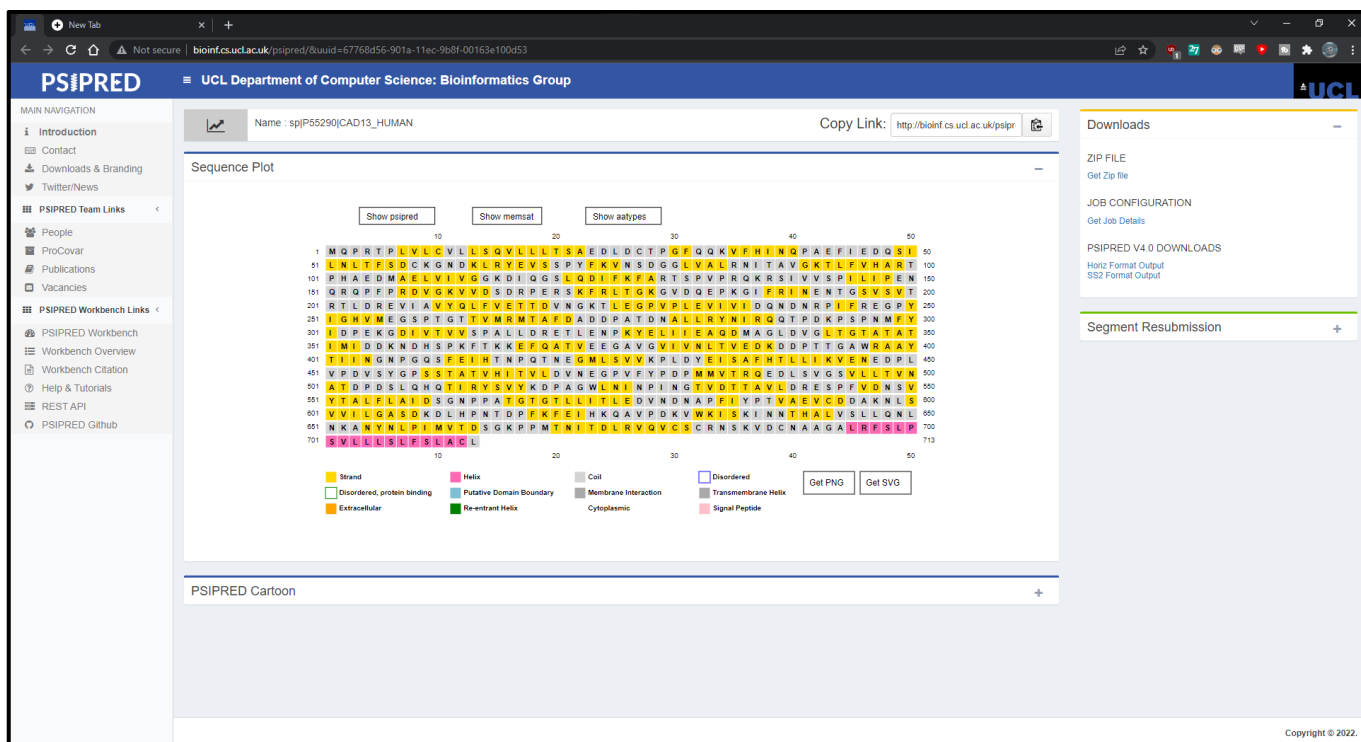
**Email (optional)**

Email (optional)

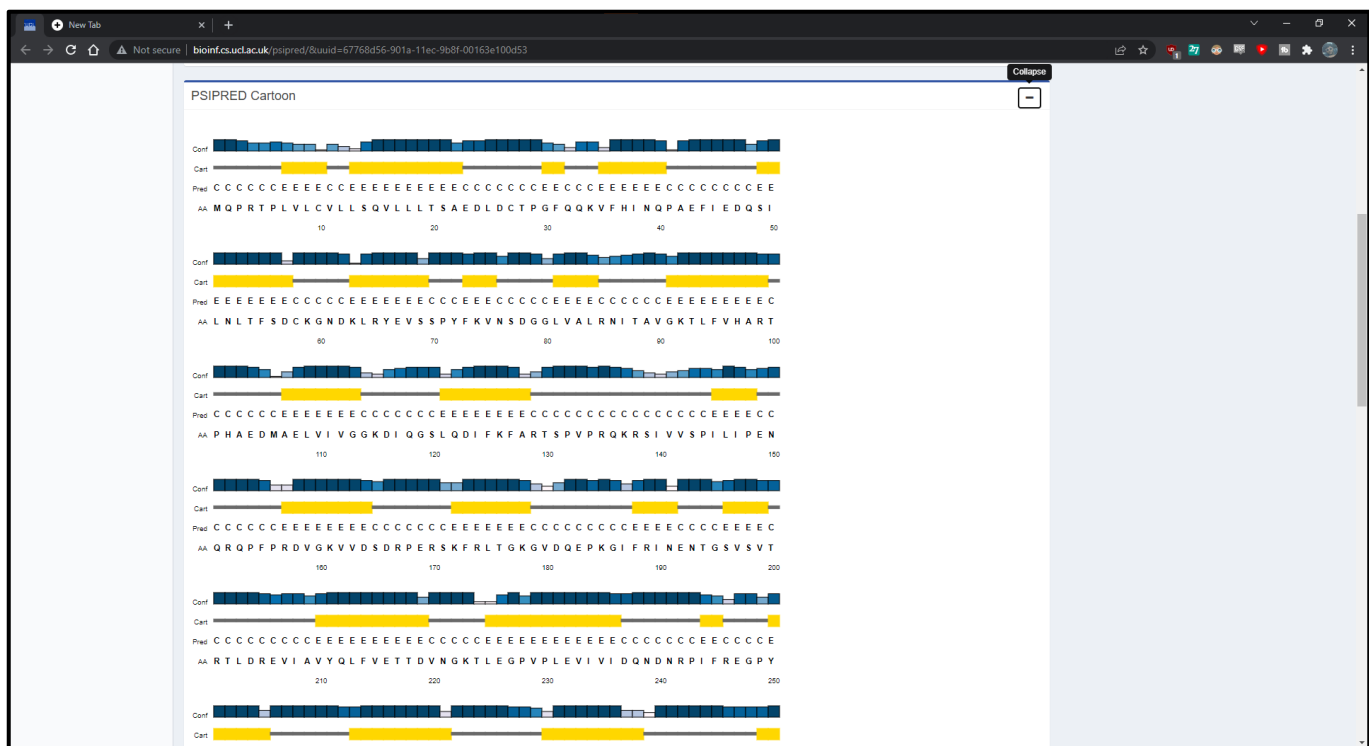
[Reset](#) [Submit](#)

Copyright © 2022.

**Fig3. Search bar of PSIPRED with FASTA sequence of Cadherin**



**Fig4. Sequence plot of CAD13\_HUMAN Cadherin in PSIPRED**



### Results:

### ➔ First Generation: Chou and Fasman

- In secondary structure prediction using Chou and Fasman tools. The FASTA Sequence length of query Cadherin is 713 and their total residues are H: 455; E: 470; T:95; and the percentage is H: 63.8; E: 65.9; T:13.3;

## ➔ Second Generation: GOR IV

- In secondary structure prediction using GOR IV tools. The FASTA Sequence length of query Cadherin is 713 and total residues are Alpha helix: 147 (20.62%); Extended Strand: 180 (25.25%); Random coil: 386 (54.14%).

### ➔ Third Generation: PSIPRED

- In secondary structure prediction using PSIPRED tools. The FASTA sequence length of query Cadherin is 713. Here the maximum helix residues are predicted and we can find legend of confidence of prediction, 3-state assignment cartoon and target sequence.

## Conclusion:

- ➔ The secondary structure of the protein is Alpha helix and Beta sheet. Protein structure plays a key role in its function. Secondary structure prediction is important from the structural and functional point of view. Secondary structure is local interactions between stretches of a polypeptide chain and include alpha helix and beta pleated sheet structures.

## References:

- ➔ Ranscht, B. (2010). Cadherin Regulation of Adhesive Interactions. Handbook of Cell Signaling, 1975–1988. <https://doi.org/10.1016/b978-0-12-374145-5.00242-4>
- ➔ CHOU-FASMAN. (2022, February 12). CHOU-FASMAN. Retrieved February 14, 2022, from <http://www.biogem.org/tool/chou-fasman/index.php>
- ➔ Ashok Kumar, T. (2013). CFSSP: Chou and Fasman Secondary Structure Prediction server. WIDE SPECTRUM: Research Journal. 1(9):15-19.
- ➔ GOR IV. (2022, February 12). GOR IV. Retrieved February 14, 2022, from [https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_gor4.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html)
- ➔ PSIPRED. (2016, September 12). Http://Bioinf.Cs.Ucl.Ac.Uk/Psipred. Retrieved February 14, 2022, from <http://bioinf.cs.ucl.ac.uk/psipred>