

# Protein Secondary Structure Prediction

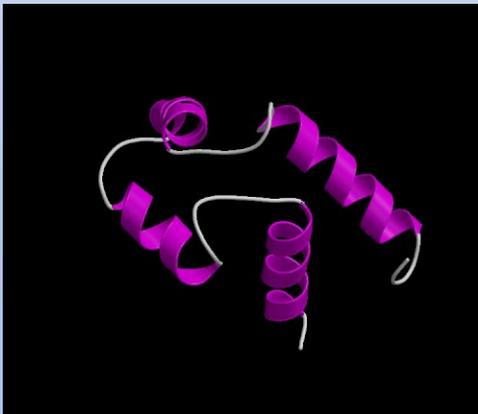
# Protein Secondary Structure

Every protein assumes a specific shape and performs a specific function.

Protein secondary structure refers to regular, recurring arrangements in space of adjacent amino acid residues in a polypeptide chain.

It is maintained by hydrogen bonds between amide hydrogens and carbonyl oxygens of the peptide backbone.

The two most common secondary structural elements are  $\alpha$ -helices and  $\beta$ -structures, though beta turns and omega loops occur as well.



*$\alpha$ -Helix*

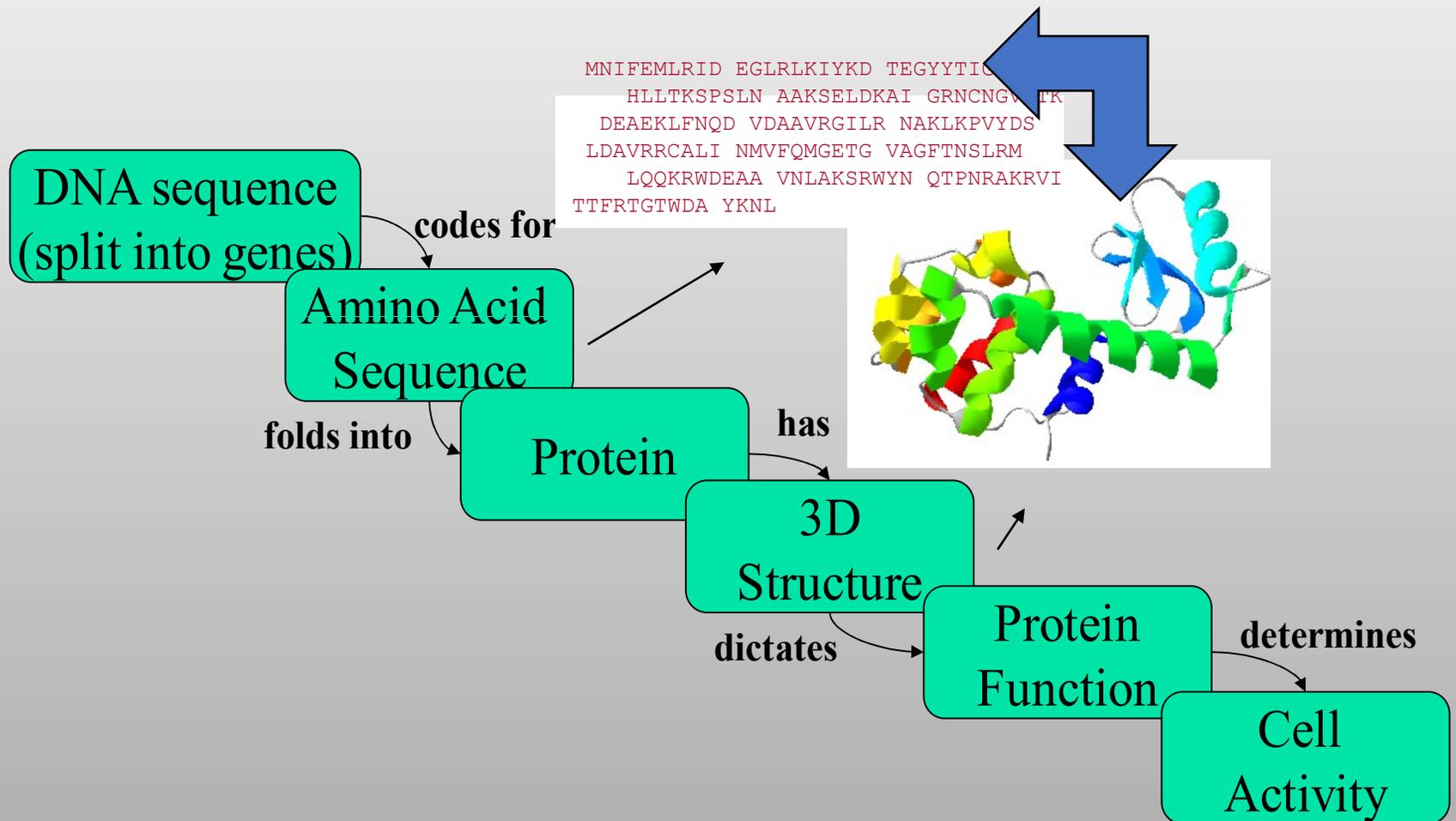


*$\beta$ -Strand*



*Coil*

# From DNA to Cell Function



# Protein Secondary Structure Prediction

Research in protein structure and function is one of the most important areas of research in proteomics.

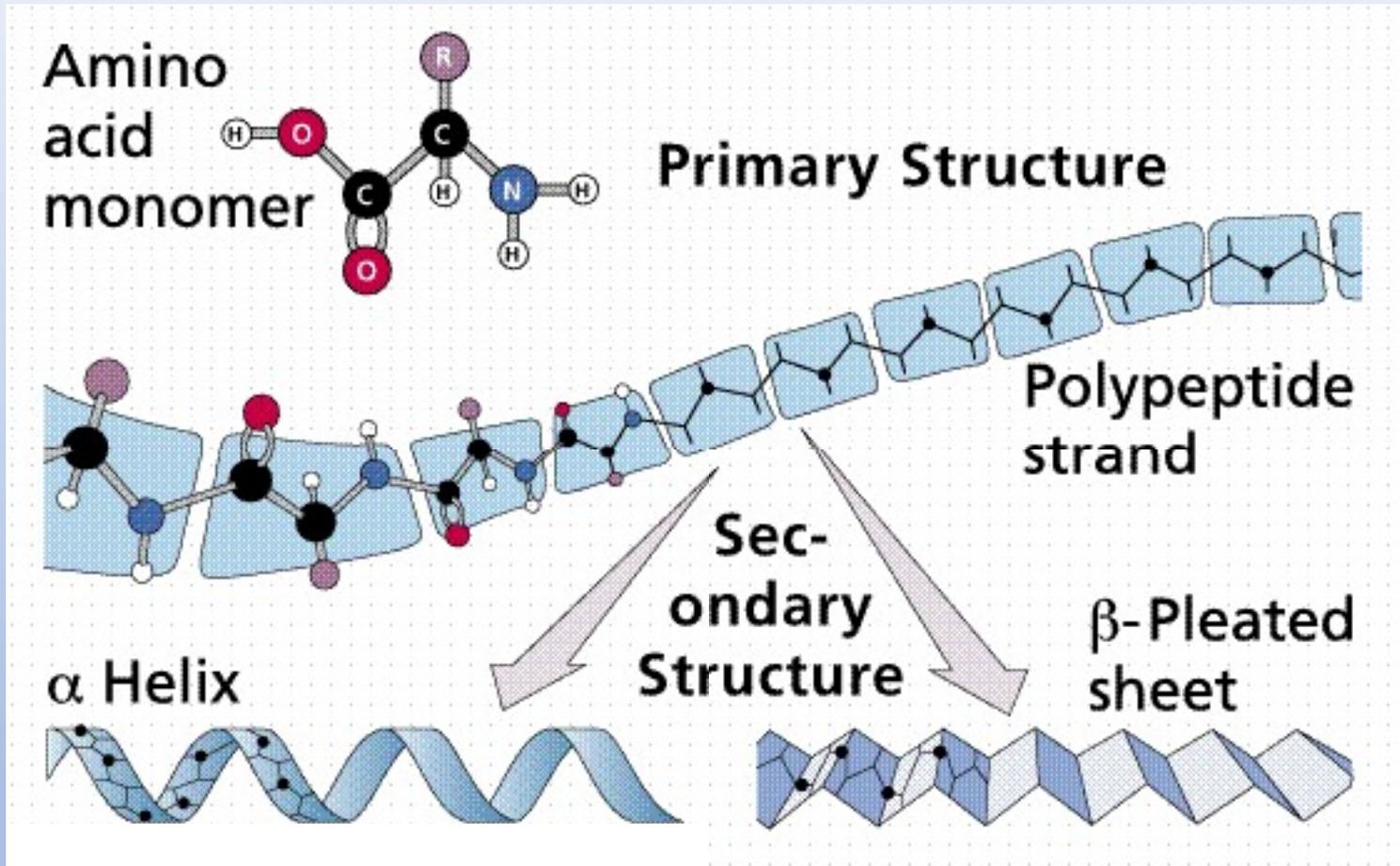
A key characteristic of the protein is the three-dimensional structure into which the polypeptide chain folds which is referred to as tertiary structure.

The information necessary for protein folding resides completely within the primary structure.

The development of rapid methods of DNA sequencing coupled with the straightforward translation of the genetic code into protein sequences has amplified the urgent need for automated methods of interpreting these one-dimensional(1D), linear amino acid sequence.

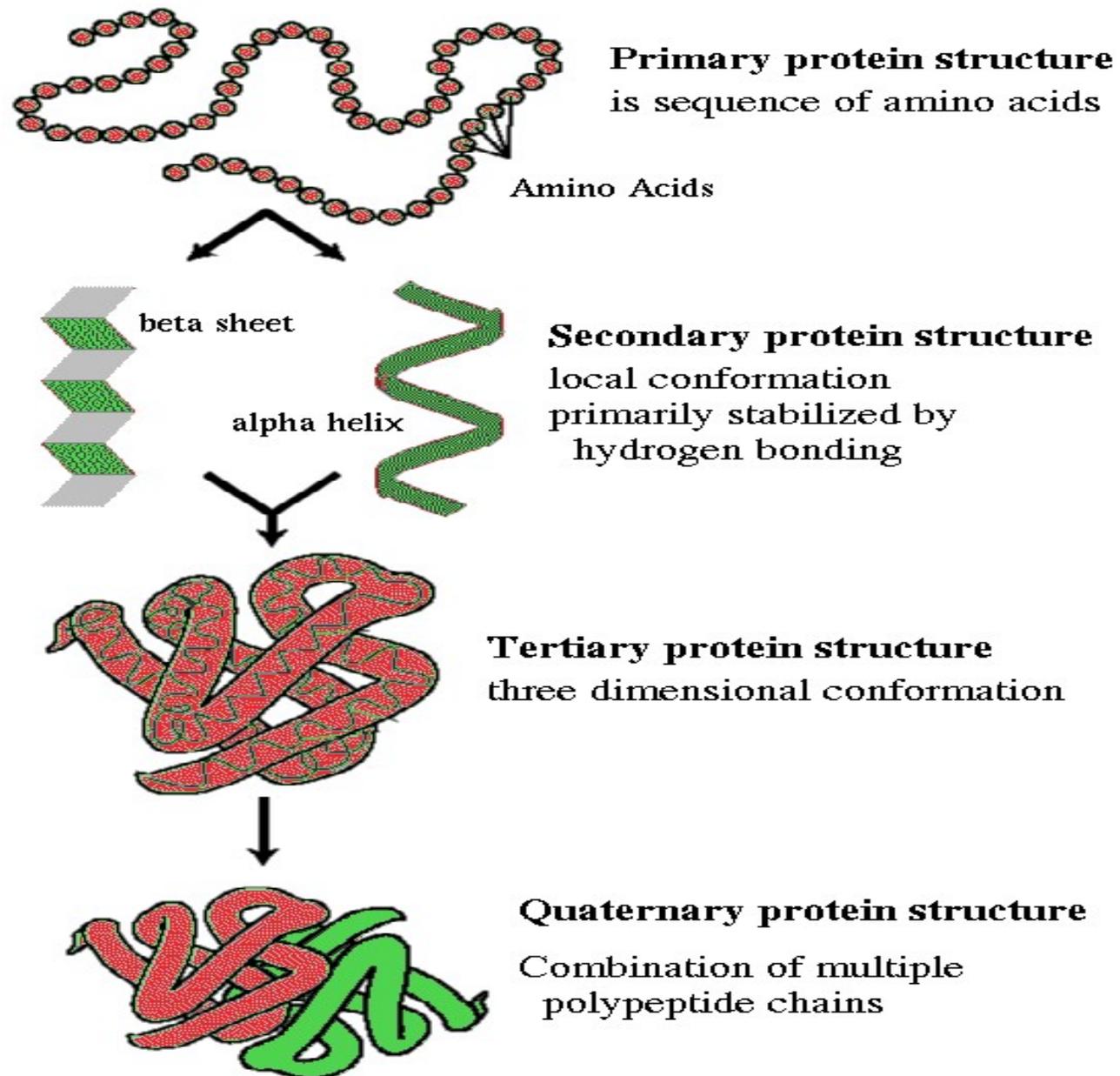
Classifiers are used to efficiently to predict the protein secondary structure which are used to classify the three types of  $\alpha$ -helix,  $\beta$ -sheet and C-coil in terms of two-dimensional(2D) structure.

# Protein Secondary Structure



**Secondary structure = spatial arrangement of amino-acid residues that are adjacent in the primary structure**

# Levels of Protein Structure



# Protein Structure Prediction Tools

- **Experimental Techniques**

- X-ray Crystallography
- NMR

- **Computational tools**

- **I-TASSER Online**
- **SOPMA**
- **PHYRE2**
- **MODELLER**
- **HHPred**

# I-TASSER

I-TASSER is an online bioinformatics platform for predicting protein structure and function. It has been developed by Zhang Lab (University of Michigan)

Home Research COVID-19 Services Publications People Teaching Job Opening News Forum Lab Only

Online Services

- I-TASSER
- QUARK
- LOMETS
- COACH
- COFACTOR
- MetaGO
- MUSTER
- CEthreader
- SEGMER
- FG-MD
- ModRefiner
- REMO
- DEMO
- SPRING
- COTH
- BSpred
- ANGLOR

**I-TASSER**  
Protein Structure & Function Predictions

(The server completed predictions for [582565 proteins](#) submitted by [139928 users](#) from [151 countries or regions](#))  
(The template library was updated on [2020/11/22](#))

I-TASSER (Iterative Threading ASSEMBly Refinement) is a hierarchical approach to protein structure prediction and structure-based function annotation. It first identifies structural templates from the PDB by multiple threading approach [LOMETS](#), with full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the target are then derived by re-threading the 3D models through protein function database [BioLiP](#). I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide [CASP7](#), [CASP8](#), [CASP9](#), [CASP10](#), [CASP11](#), [CASP12](#), and [CASP13](#) experiments. It was also ranked the best for function prediction in [CASP9](#). The server is in active development with the goal to provide the most accurate protein structure and function predictions using state-of-the-art algorithms. Please report problems and questions at [I-TASSER message board](#) and our developers will study and answer the questions accordingly. ([>> More about the server ...](#))

[Structure models for the 2019-nCov Coronavirus genome by C-I-TASSER](#) **NEW**

[\[Queue\]](#) [\[Forum\]](#) [\[Download\]](#) [\[Search\]](#) [\[Registration\]](#) [\[Statistics\]](#) [\[Remove\]](#) [\[Potential\]](#) [\[Decoys\]](#) [\[News\]](#) [\[Annotation\]](#) [\[About\]](#) [\[FAQ\]](#)

I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence within [10, 1500] residues in [FASTA format](#). [Click here for a sample input](#):

```
>protein
MAKSSFKISNPLEARMSESSRIREKYPDRIPVIVEKAGQSDVDPIDKKKYLVPADLTVGQ
FVYVVRKRIKLGAEKAI FVFKNTLPPTAALMSAIYEEHKDEDGFLYMTYSGENTFGSLT
VA
```

# I-TASSER

zhanglab.ccmb.med.umich.edu/I-TASSER/

- DEMO
- SPRING
- COTH
- BSPred
- ANGLOR
- EDock
- BSP-SLIM
- SAXSTER
- FUpred
- ThreaDom
- ThreaDomEx
- EvoDesign
- GPCR-I-TASSER
- MAGELLAN
- BindProf
- BindProfX
- SSIPe
- ResQ
- IonCom
- STRUM
- DAMPred
- TM-score
- TM-align
- MM-align

## I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence within [10, 1500] residues in [FASTA format](#). [Click here for a sample input](#):

```
>protein
MAKSSFKISNPLEARMSESSRIREKYPDRIPVIVEKAGQSDVPDIDKKKYLVPADLTVGQ
FVYVVRKRIKLGAEKAI FVFKNTLPPTAALMSAIYEEHKDEDGFLYMTYSGENTFGSLT
VA
```

Or upload the sequence from your local computer:

No file chosen

Email: (mandatory, where results will be sent to)

Password: (mandatory, please click [here](#) if you do not have a password)

ID: (optional, your given name of the protein)

- Option I:** [Assign additional restraints & templates to guide I-TASSER modeling.](#)
- Option II:** [Exclude some templates from I-TASSER template library.](#)
- Option III:** [Specify secondary structure for specific residues.](#)

Keep my results public (uncheck this box if you want to keep your job private, and a key will be assigned for you to access the results. We received numerous requests from users for their key to access result. To save your time, please keep results public, or ensure you remember the key if you choose to keep job private)

(Please submit a new job only after your old job is completed)

# I-TASSER

→ [npsa-prabi.ibcp.fr/tmp/5782121d4108.simpa96](https://npsa-prabi.ibcp.fr/tmp/5782121d4108.simpa96)

06 secondary structure prediction program

elix b = strand . = other

indicates prediction strength: 0-weak, 2-medium, 5-strong 9-very strong

ary structure prediction without homologues

nce 1

53160

```
           10      20      30      40      50      60
nce  MKKITYDLAELSGVSASAVSAILNGNWKKRRISAKLAEKVTRIAEEQGYAINRQASHLR
tion  .. bbbHHHHHHH... HHHHHHHHH... HHHHHHHHHHHHHHHHHHHHH... HHHHHHHHH
      14212210221123320123222124421122234554445444331243000023443
```

```
           70      80      90      100     110     120
nce  SKKSHVIGMIIPKYDNRYFGSIAERFEEMARERGLLPITCTRRRPELEIEAVKAMLSWQ
tion  H.HHHHHbbb..... HHHHHHHHHHHHHHH... bbbbbb..... HHHHHHHHHHHHHH.
      001010001202345310002455566654422640244212223111233344443201
```

```
           130     140     150     160     170     180
nce  VDwVVATGATNPDKISALCQQAGVPTVNLDLPGSLSPSVISDNYGGAALTKILANSAR
tion  .. bbbb..... HHHHHHHHHH... bbbb..... bb..... HHHHHHHHHHHHHH.
      010230255673024433331476402101467545411114445410223433431110
```

```
           190     200     210     220     230     240
nce  RRGELAPLTFIGGRRATITPASVYAASTMRIASWGLACRRRIFWLP AIRKATLRTACRS
tion  ..... bbbb..... bb... bbbb..... HHHHHH... HHHHHHHHHH..
      013310034304542110343200122101100133210000001211100122331023
```

```
           250     260     270
nce  LAARRRCCRGYLLTRRYPWKGLCAGCRRW
```

# I-TASSER

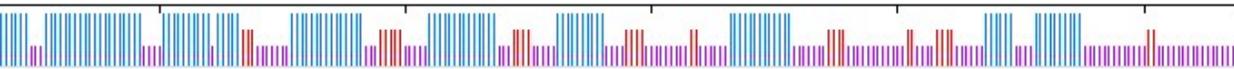
npsa-prabi.ibcp.fr/cgi-bin/secpred\_simpa96.pl

PA96 in: [AnTheProt (PC), Download...] [HELP]

```
10      20      30      40      50      60      70
|      |      |      |      |      |      |
DLAELSGVSASAVSAILNGMWKKRRISAKLAEKVTRIAEEQGYAINRQASMLRSKSKSHVIGMI
hhhhhCCc hhhhHhhhhcCChhhhhHHHHHHHHHHHHcCCc hhhhHHHHhchhhhheee
YFGSIAERFEEMARERGLLPITCTRRRPELEIEAVKAMLSWQVDWVATGATNPKISALCQ
hhhhHHHHHHHHHHcCCeeEEeccccChhhHHHHHHHHhhccceeeEecCCCCHhhHHHHH
WLDLPGSLSPSVISDNYGGAKALTKILANSARRRGELAPLTFIGRRATITPASVYAASTMR
eeCCCCCCCCceeeCCCCChhhHHHHHHhhcccCCceEEeCCCCccccCCCCeeccccce
RRRIFWLPAIRKATLRTACRSLAARRRCRGYLLTRYPWKGLCAGCRRWV
hhhhhcccc hhhhHHhhcCccccccCCceeeCCCccCCcCCCCccC
```

length : 270

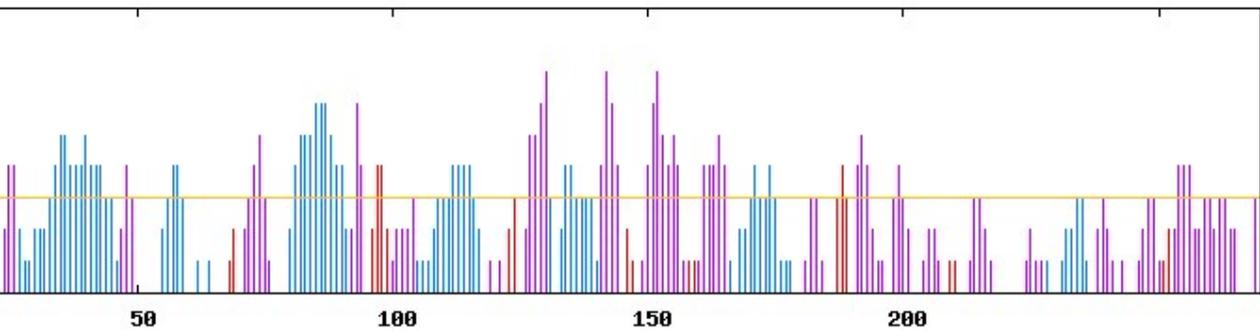
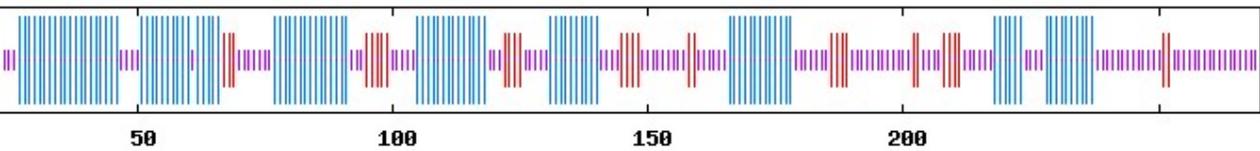
```
:
alpha helix (Hh) : 118 is 43.54%
beta helix (Gg) : 0 is 0.00%
gamma helix (Ii) : 0 is 0.00%
beta bridge (Bb) : 0 is 0.00%
beta sheet (Ee) : 33 is 12.18%
turn (Tt) : 0 is 0.00%
loop region (Ss) : 0 is 0.00%
random coil (Cc) : 119 is 43.91%
ambiguous states (?) : 0 is 0.00%
other states : 1 is 0.37%
```



# I-TASSER

npsa-prabi.ibcp.fr/cgi-bin/secpred\_simpa96.pl

```
ridge (Bb) : 0 is 0.00%
ed strand (Ee) : 33 is 12.18%
urn (Tt) : 0 is 0.00%
egion (Ss) : 0 is 0.00%
coil (Cc) : 119 is 43.91%
ous states (?) : 0 is 0.00%
states : 1 is 0.37%
```



with a scale value of 3 and over (uppercase letters) are predicted at better than 80%

result file (text): [SIMPA96]

157.41.100.27. Last modification time : Sun Nov 29 18:41:10 2020. Current time : Sun Nov 29 18:41:10 2020

# SOPMA

→ ↻ 🔒 npsa-prabi.ibcp.fr/cgi-bin/npsa\_automat.pl?page=npsa\_sopma.html



## PRABI-GERLAND RHONE-ALPES BIOINFORMATIC POLE GERLAND SITE

Institute of Biology and Protein Chemistry

🏠 Home

📍 Services

🎓 Teaching

📄 Publications

🔗 Links

📁 Jobs

🗨️

## SOPMA SECONDARY STRUCTURE PREDICTION METHOD

[Contact](#) [[NPS@ help](#)] [[Original server](#)]

Protein name (optional):

Protein sequence below: [help](#)

```
MTYDLAELSGVSASAVSAILNGNWKKRRISAKLAEKVTRIAEEQ  
RQASMLR  
IVIGMIIPKYDNRVFGSIAERFEEMARERGLLPIITCTRRRPELEI  
AMLSWQ  
VATGATNPKISALCQQAGVPTVNLDPGSLSPSVISDNYGG  
CHKILANSAR  
LAPLTFIGGRRATITPASVYAASTMRIASWGLACRRRIFWLP  
AIR  
TACRSG
```

Window width:

### Parameters

Number of conformational states:

Similarity threshold:

# SOPMA

npsa-prabi.ibcp.fr/cgi-bin/secpred\_simpa96.pl

PA96 result : UNK\_1545150

Levin JM, Garnier J (1988) Improvements in a secondary structure prediction method based on a search for local sequence homologies and its use as a building tool. Biochim. Biophys. Acta 955, 283-295  
PA96 in: [AnTheProt (PC), Download...] [HELP]

```
10      20      30      40      50      60      70
|       |       |       |       |       |       |
DLAELSGVSAVASAILNGNWKKRRISAKLAEKVTRIAEEQGYAINRQASMLRSKKS HVIGMI
hhhhhccChhhhhhhhhcChhhhhHHHHHHHHHHHHHHcCCc hhhhhHHHHh c hhhheee
RYFGSIAERFEEMARERGLLPITCTRRRPELEIEAVKAMLSWQVDWV VATGATNPKISALCQ
hhhhHHHHHHHHHHcCCeeEEeccccChhhHHHHHHHHhhccceeEecCCCCHhHHHHHH
WMLDLPGLSPSVISDN YGGAKALTHKILANSARRRGELAPLTFIGRRATITPASVYAASTMR
eeccccccccceccccChhhHHHHHHHHhhcccCCceEEeCCCCccccCCCceccccce
CRRRIFWLP AIRKATLRTACRSGLAARRRCCRGYLLTRRYPWKGLCAGCRRWV
c hhhhhccchhhhhHHhhcCcccccccccCceecCCCccCCcCCccccccC
```

Sequence length : 270

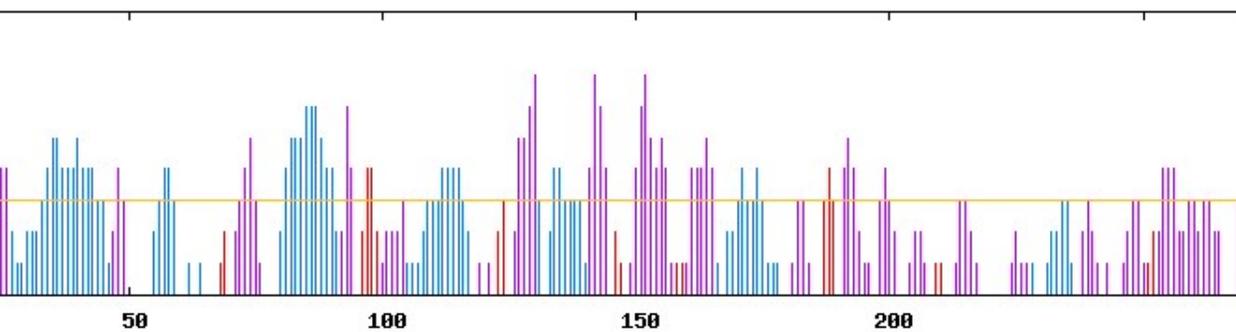
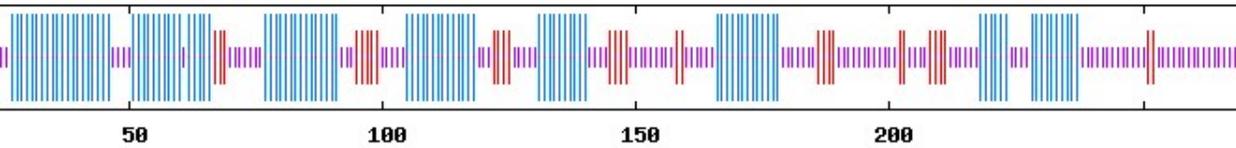
Statistics :

alpha helix (Hh)	: 118 is 43.54%
beta helix (Gg)	: 0 is 0.00%
gamma helix (Ii)	: 0 is 0.00%
alpha bridge (Bb)	: 0 is 0.00%
extended strand (Ee)	: 33 is 12.18%
beta turn (Tt)	: 0 is 0.00%
loop region (Ss)	: 0 is 0.00%
random coil (Cc)	: 119 is 43.91%
ambiguous states (?)	: 0 is 0.00%

# SOPMA

npsa-prabi.ibcp.fr/cgi-bin/secpred\_simpa96.pl

```
elix (Hh) : 118 is 43.54%
ix (Gg) : 0 is 0.00%
x (Ii) : 0 is 0.00%
idge (Bb) : 0 is 0.00%
d strand (Ee) : 33 is 12.18%
rn (Tt) : 0 is 0.00%
gion (Ss) : 0 is 0.00%
coil (Cc) : 119 is 43.91%
us states (?) : 0 is 0.00%
tates : 1 is 0.37%
```



with a scale value of 3 and over (uppercase letters) are predicted at better than 80%

result file (text): [\[SIMPA96\]](#)

# Measures of prediction accuracy

## Qindex

**Qindex:** (Qhelix, Qstrand, Qcoil, Q3)

- percentage of residues correctly predicted as  $\alpha$ -helix,  $\beta$ -strand, coil, or for all 3 conformations.

$$Q_3 = \frac{\text{correctly predicted residues}}{\text{number of residues}}$$

$$Q_3 = \frac{N_{predicted}}{N_{observed}} \cdot 100$$

# Measures of prediction accuracy

relation coefficient

$$C_{\alpha} = \frac{p_{\alpha} \cdot n_{\alpha} - u_{\alpha} \cdot o_{\alpha}}{\sqrt{([n_{\alpha} + u_{\alpha}][n_{\alpha} + o_{\alpha}][p_{\alpha} + u_{\alpha}][p_{\alpha} + o_{\alpha}] )}}$$

$$C_{\alpha} = 1 \text{ (=100\%)}$$

True positive $p_a$	False positive (overpredicted) $o_a$
True negative $n_a$	False negative (underpredicted) $u_a$