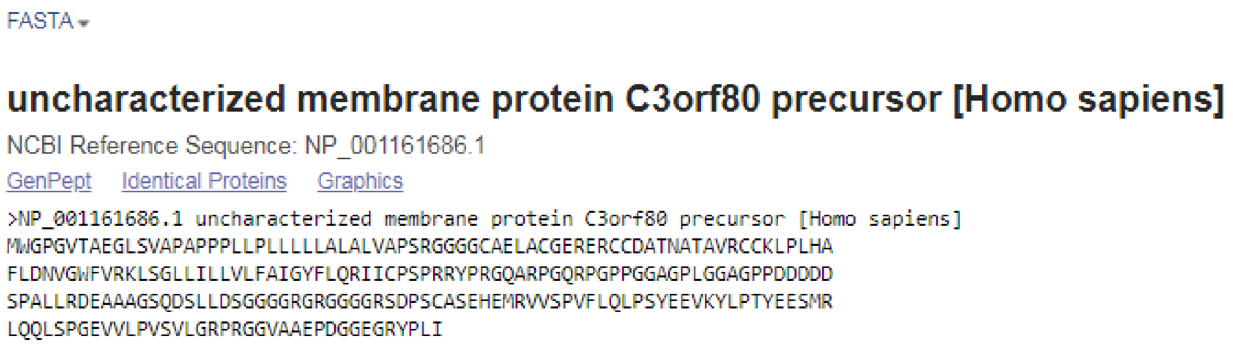
DEPARTMENT OF BIOINFORMATICS & COMPUTATIONAL BIOLOGY
VIRTUAL UNIVERSITY OF PAKISTAN

Course: BIF401- Bioinformatics – I	Term: FALL 2019
Instructor: Syed Hassan Abbas	Assignment #2
Assigned Date: 04-11-2019	**Due Date: 15-01-2020**
Total Marks: 15	

Zuraiz Butt

Take a candidate human gene amino acid sequence, preferably whose structure is not yet predicted.



uncharacterized membrane protein C3orf80 precursor [Homo sapiens]

NCBI Reference Sequence: NP_001161686.1

```
>NP_001161686.1 uncharacterized membrane protein C3orf80 precursor [Homo sapiens]
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDATNATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQRPGPPGGAGPLGGAGPPDDDDD
SPALLRDEAAAGSQDSLLDSGGGGRGRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESMR
LQQLSPGEVVLPVSVLGRPRGGVAAEPDGGEGRYPLI
```

```
>NP_001161686.1 uncharacterized membrane protein C3orf80 precursor [Homo
sapiens]
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDATNATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQRPGPPGGAGPLGGAGPPDDDDD
SPALLRDEAAAGSQDSLLDSGGGGRGRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESMR
LQQLSPGEVVLPVSVLGRPRGGVAAEPDGGEGRYPLI
```

Take the sequence of seven ortholog for same protein. (Like Mouse, rat, Chimpanzee etc.)

```
>XP_004037972.1 uncharacterized membrane protein C3orf80 homolog [Gorilla
gorilla gorilla]
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDATNATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQRPGPPGGAGPLGGAGPPDDEDD
SPALLRDEAAAGSQDSLLDSGGGGRGRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESMR
LQQLSPGEVVLPVSVLGRPRGGVAAEPDGGEGRHPLI
```

```
>XP_028723059.1 uncharacterized membrane protein C3orf80 homolog [Peromyscus
leucopus]
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDSANATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQARPGPAGGSGPPGTAGQPDDDD
DSPALLRDEVAAGSLDSLLDSGGGGRGRGGGGRLPPTCVSEHELRVVSPVFLQLPSYEEVKYLPTYEESM
RLQQLSPAEVVLPVSVLGHPRGGSAGDPDGGQGRFPLI
```

```
>XP_004598394.1 PREDICTED: uncharacterized membrane protein C3orf80 homolog
[Ochotona princeps]
MWGPGVTTEGLSAAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDAANATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARAGQARPGTPGAAGPPGGAGPPDDDD
DSPALLRDEVAAGSQDSLLDSSGGGGRGRGGGSRPAPSCASEHELRVVSPVFLQLPSYEEVKYLPTYEES
MRLQQPSPGEVTLPVSVLGRPRGCGAGESDGGEGRFPLI
```

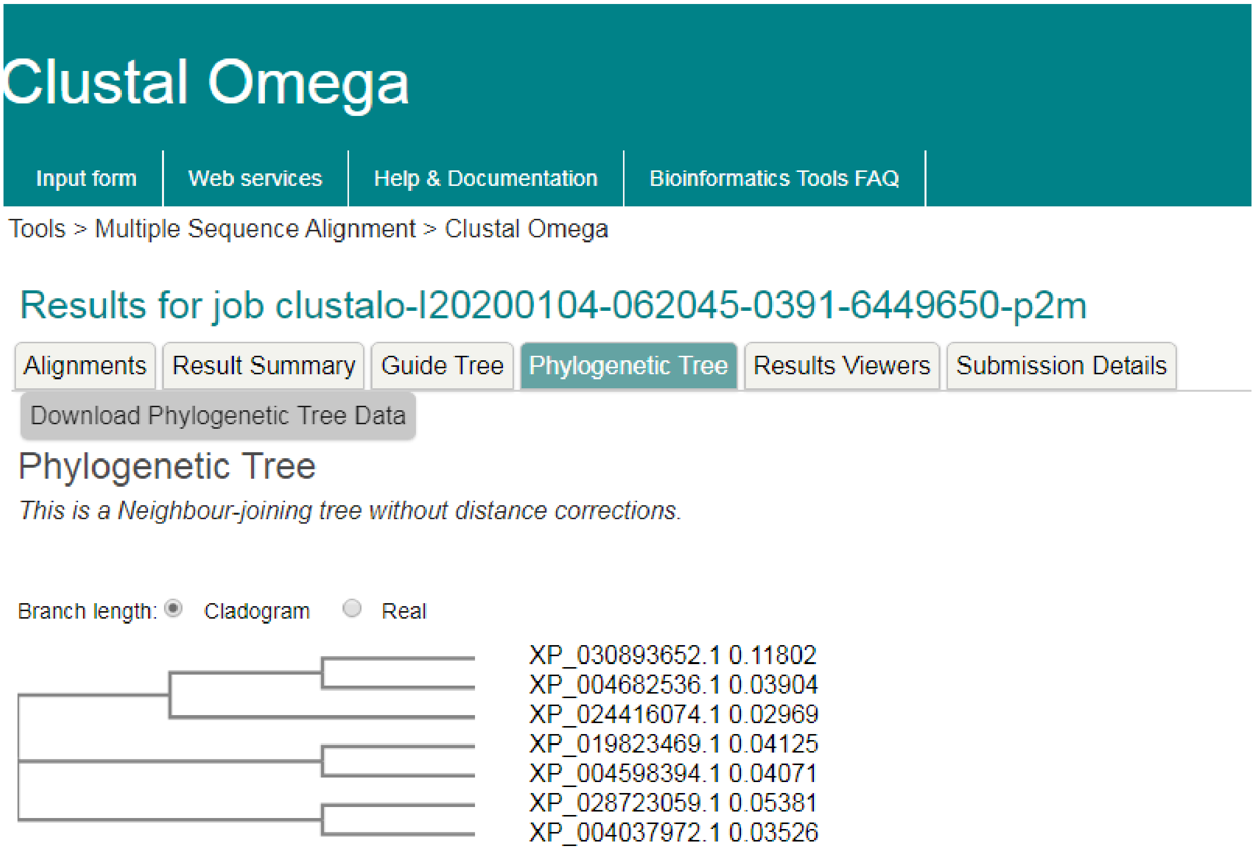
```
>XP_024416074.1 LOW QUALITY PROTEIN: uncharacterized membrane protein C3orf80
homolog [Desmodus rotundus]
MWGTGVTAEGLSAAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDAANATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPXDDDDDSPALLRDEAAAGSQDSLLDSGSGGRGRGGG
GFSVASEHELRLVSPAFLQLPSYEEVKYLPTYEESMRLQQLSPGEVVLPVSVLGRPRGGGSGEPDGGEGR
FPLI
```

```
>XP_030893652.1 uncharacterized membrane protein C3orf80 homolog
[Leptonychotes weddellii]
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDAANATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPHGQARPGQPRPXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXGGGRGRGGGGRSAPSCASEHELRVVSPVFLQLPSYEEVKYLPTYEESMRL
QQLSPGEVVLPVSVLGRPRGGCAGESDGGEGRFPLI
```

```
>XP_004682536.1 PREDICTED: uncharacterized membrane protein C3orf80 homolog
[Condylura cristata]
MWGPGVTAEGLPVAPALLLALVALVAPSRGGGSCAELACGERERCCDAANATAVRCCKLPLHAFLDNVGW
FVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRAAGSQDSLLDSGGRGRGGARAVPASALEHELRVVSP
VFLQLPSYEEVKYLPTYEESMRLQQLSPGEVVLPVSVLGVPEPDGEGRFPLI
```

```
>XP_019823469.1 PREDICTED: uncharacterized membrane protein C3orf80 homolog
[Bos indicus]
MWGPGITAEGVSVAPAPPXXLPLLLLLALALVAPSRGGGGCAELACGERERCCDAANATAVRCCKLPLHA
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGXXRPGPPGGAGQPGAAGPPDDEDDDSPE
LLRDEVAAGSQDSLLDSGGGGGRGRAGGGRSAPSCASEHELRVVSPVFLQLPSYEEVKYLPTYEESMRLQ
QLSPGEVVLPVSVLGRPRGGGAGESDGGGEGRFPLI
```

Make a phylogenetic tree after doing MSA (Multiple Sequence Alignment).



Results for job clustalo-I20200104-062045-0391-6449650-p2m

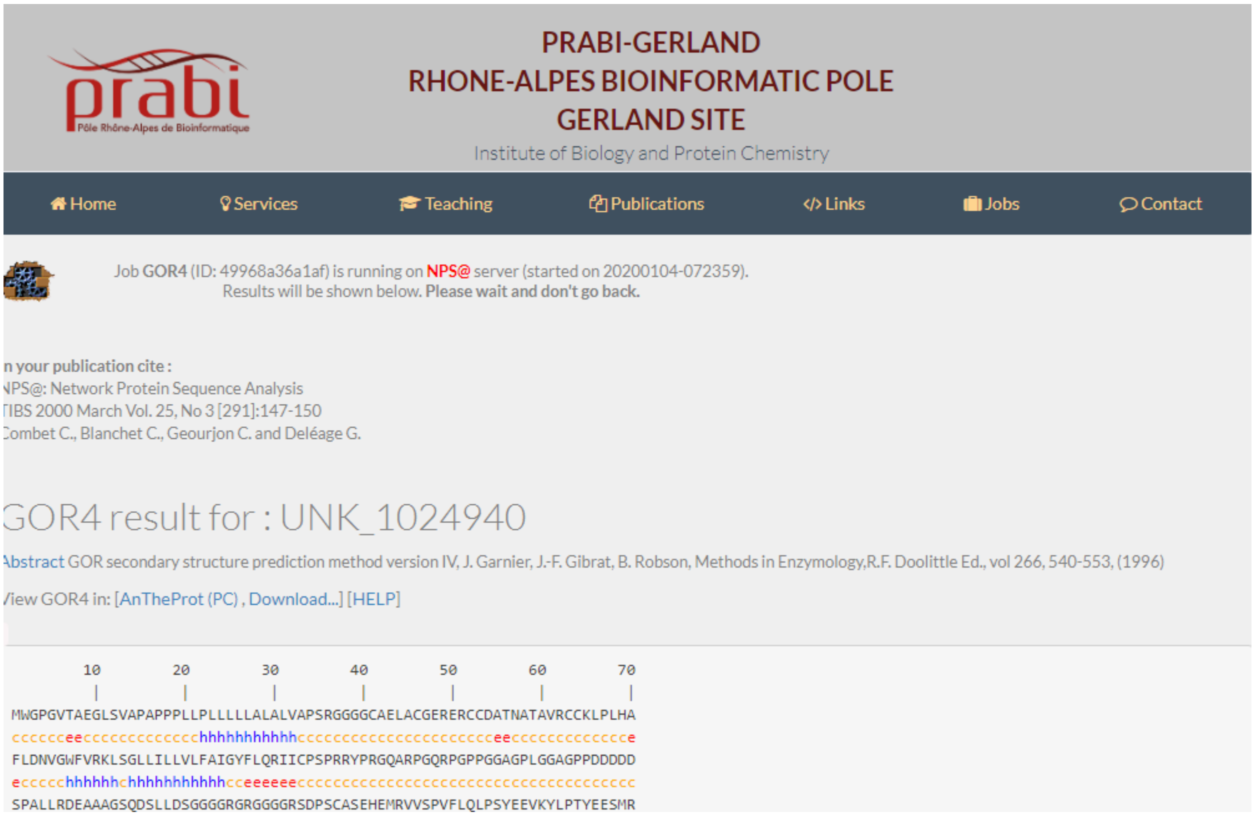
Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ● Cladogram ○ Real

XP_030893652.1 0.11802
XP_004682536.1 0.03904
XP_024416074.1 0.02969
XP_019823469.1 0.04125
XP_004598394.1 0.04071
XP_028723059.1 0.05381
XP_004037972.1 0.03526

Predict the Secondary structure of the candidate gene amino acid sequence. (alpha helix, beta sheets, strands)



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

Job **GOR4** (ID: 49968a36a1af) is running on **NPS@** server (started on 20200104-072359).
Results will be shown below. **Please wait and don't go back.**

n your publication cite :
NPS@: Network Protein Sequence Analysis
TIBS 2000 March Vol. 25, No 3 [291]:147-150
Combet C., Blanchet C., Geourjon C. and Deléage G.

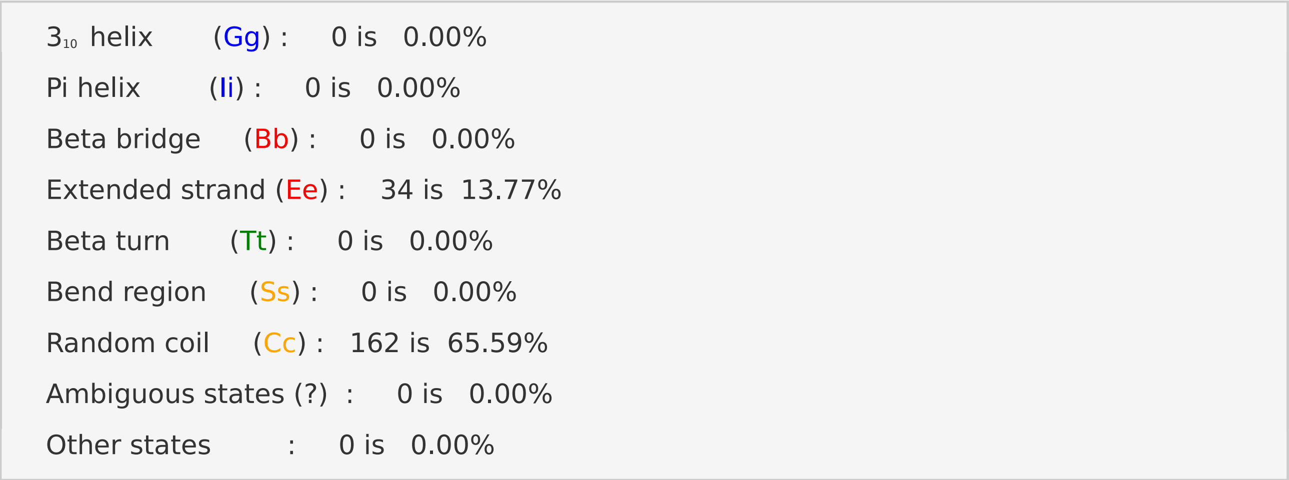
GOR4 result for : UNK_1024940

Abstract GOR secondary structure prediction method version IV, J. Garnier, J.-F. Gibrat, B. Robson, Methods in Enzymology,R.F. Doolittle Ed., vol 266, 540-553, (1996)

View GOR4 in: [AnTheProt (PC) , Download...] [HELP]

```
       10        20        30        40        50        60        70
        |         |         |         |         |         |         |
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDATNATAVRCCKLPLHA
cccccceeccccccccccccchhhhhhhhhhhcccccccccccccccccccccceecccccccccccccce
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQRPGPPGGAGPLGGAGPPDDDDD
ecccccchhhhhhchhhhhhhhhhhhcceeeeeecccccccccccccccccccccccccccccccccccccc
SPALLRDEAAAGSQDSLLDSGGGGRGRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESMR
```

```
      10        20        30        40        50        60        70
       |         |         |         |         |         |         |
MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDATNATAVRCCKLPLHA
cccccceeccccccccccccchhhhhhhhhhhcccccccccccccccccccccceecccccccccccccce
FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQRPGPPGGAGPLGGAGPPDDDDD
ecccccchhhhhhchhhhhhhhhhhhcceeeeeecccccccccccccccccccccccccccccccccccccc
SPALLRDEAAAGSQDSLLDSGGGGRGRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESM
R
chhhhhhhhhhcchhhhhcccccccceeccccccccccccccccceeeeceeecccccccccccccchhhhhh
LQQLSPGEVVLPVSVLGRPRGGVAAEPDGGEGRYPLI
hhcccccceeeeeeeccccccceeeccccccccceeec
Sequence length :   247
GOR4 :
   Alpha helix     (Hh) :    51 is  20.65%
```



3_{10} helix (Gg) : 0 is 0.00%

Pi helix (Ii) : 0 is 0.00%

Beta bridge (Bb) : 0 is 0.00%

Extended strand (Ee) : 34 is 13.77%

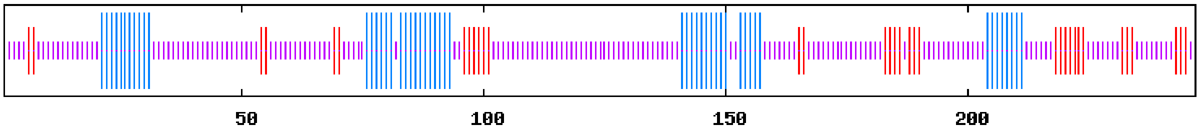
Beta turn (Tt) : 0 is 0.00%

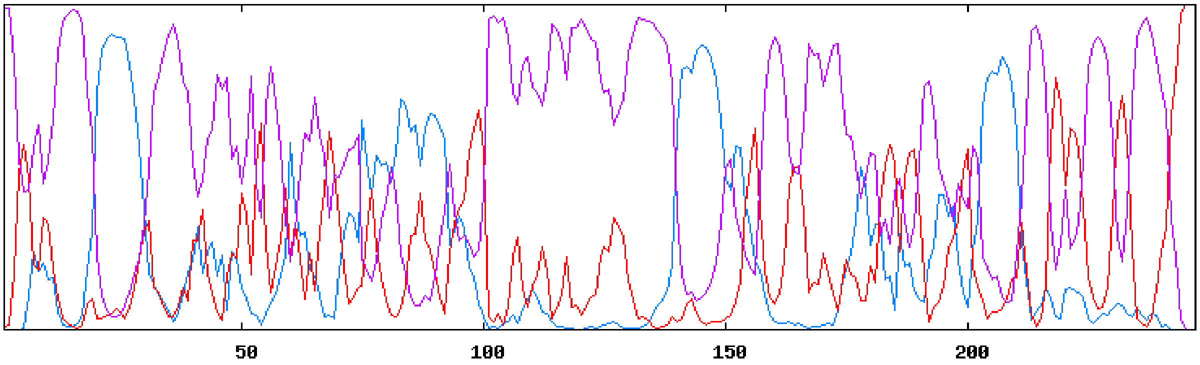
Bend region (Ss) : 0 is 0.00%

Random coil (Cc) : 162 is 65.59%

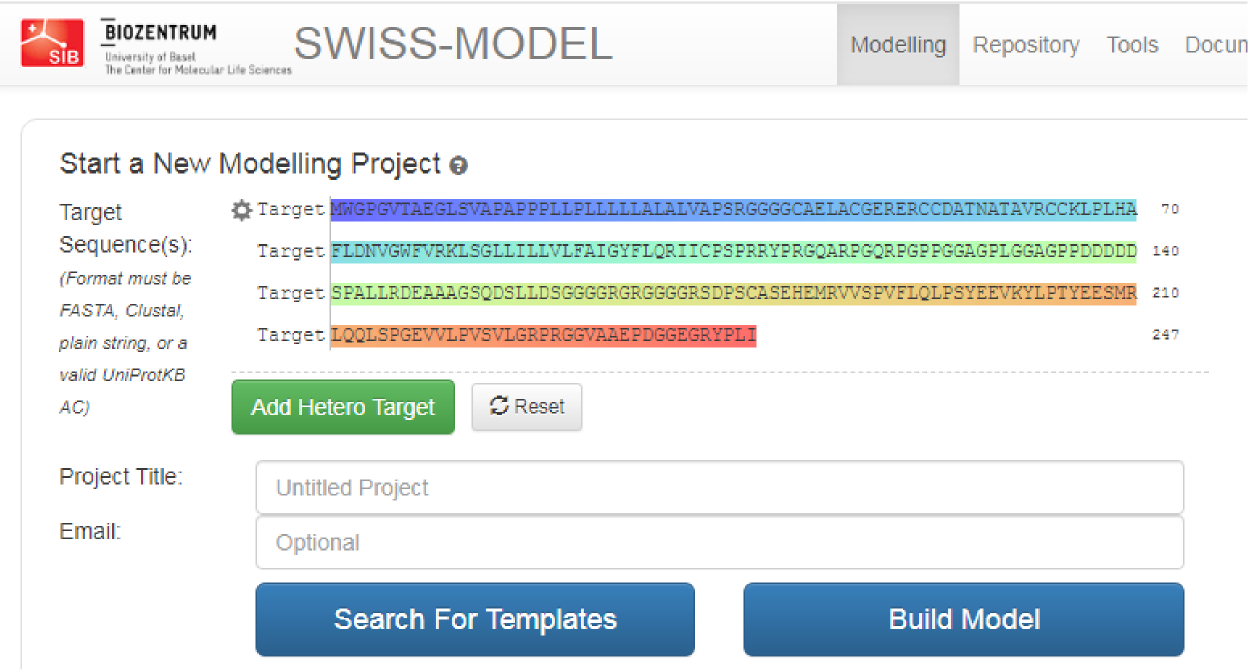
Ambiguous states (?) : 0 is 0.00%

Other states : 0 is 0.00%





Predict the tertiary structure of the sequence.



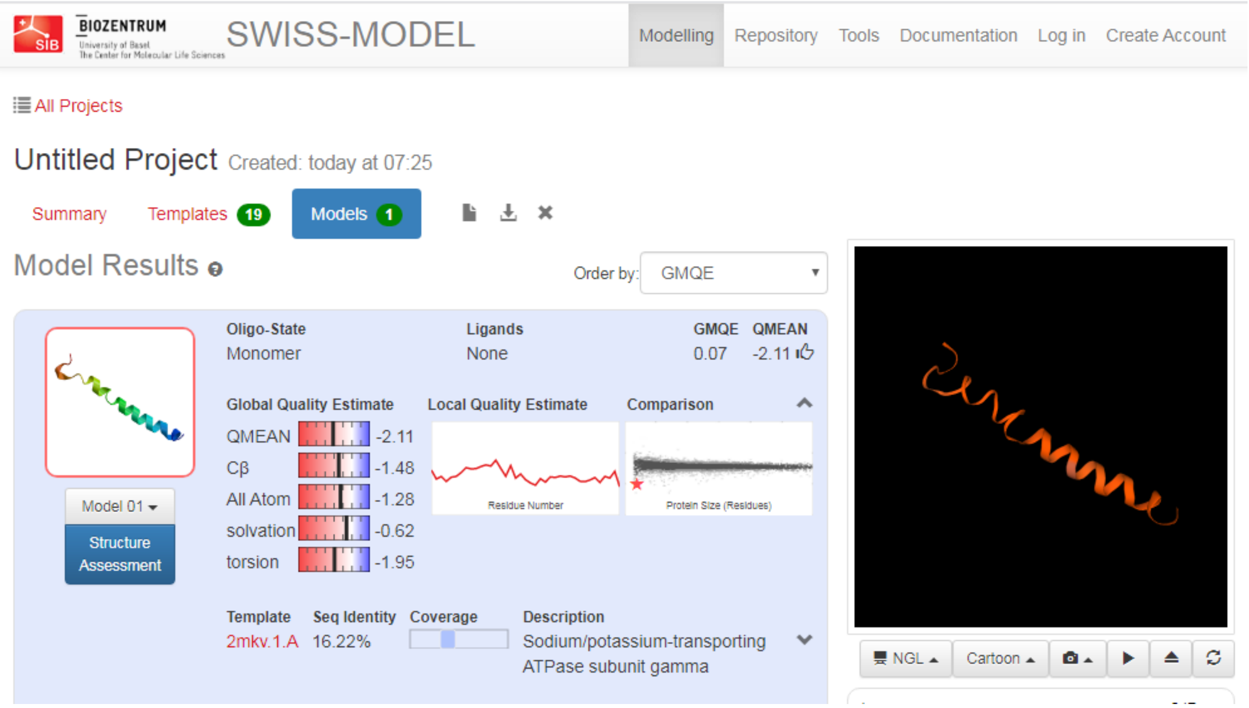
Start a New Modelling Project

Target Sequence(s): *(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)*

```
Target MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDATNATAVRCCKLPLHA  70
Target FLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYPRGQARPGQRPGPPGGAGPLGGAGPPDDDDD 140
Target SPALLRDEAAAGSQDSLLDSGGGGRGRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESMR 210
Target LQQLSPGEVVLPVSVLGRPRGGVAAEPDGGEGRYPLI                                 247
```

Project Title: Untitled Project

Email: Optional



Untitled Project Created: today at 07:25

Model Results

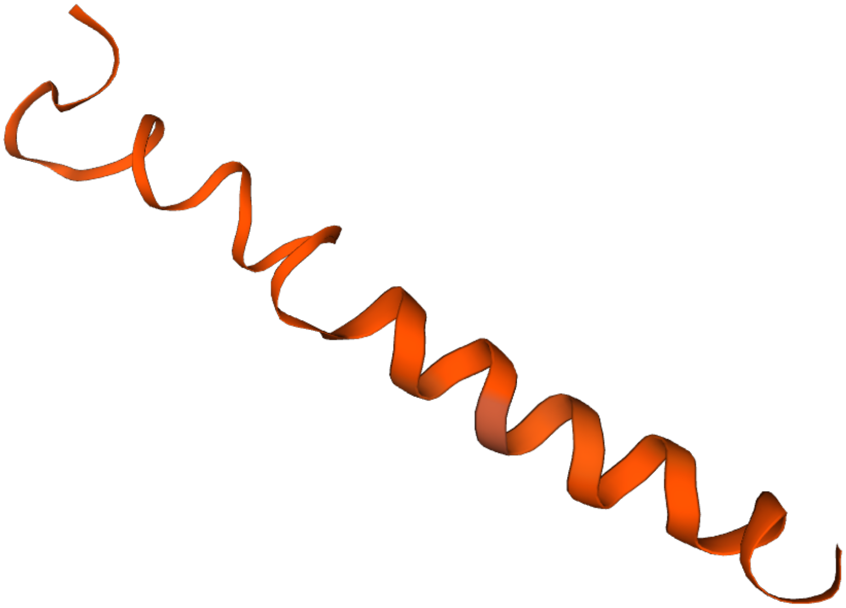
Order by: GMQE

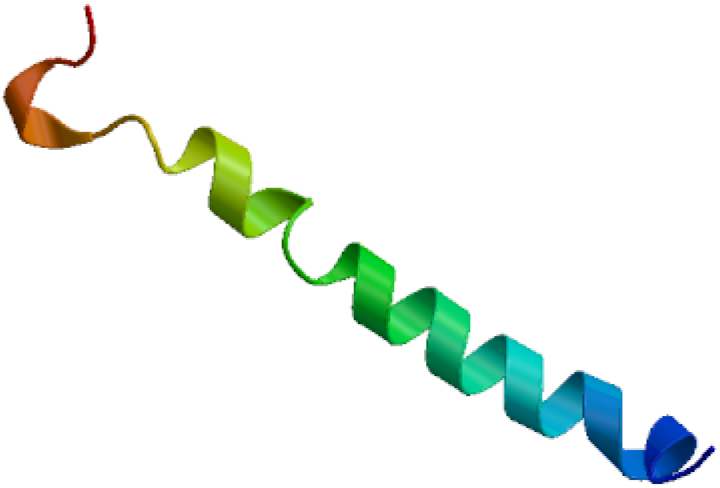
Oligo-State	Ligands	GMQE	QMEAN
Monomer	None	0.07	-2.11

Global Quality Estimate

QMEAN	-2.11
Cβ	-1.48
All Atom	-1.28
solvation	-0.62
torsion	-1.95

Template	Seq Identity	Coverage	Description
2mkv.1.A	16.22%		Sodium/potassium-transporting ATPase subunit gamma





Model 01

Oligo-State	Ligands	GMQE	QMEAN
Monomer	None	0.07	-2.11

Global Quality Estimate		Local Quality Estimate	Comparison
QMEAN	-2.11	Residue Number	
Cβ	-1.48		
All Atom	-1.28		
solvation	-0.62	★Protein Size (Residues)	
torsion	-1.95		

Template	Seq Identity	Coverage	Description
2mkv.1.A	16.22%		Sodium/potassium-transporting ATPase subunit gamma

Model-Template Alignment

```
Model_01 MWGPGVTAEGLSVAPAPPPLLPLLLLLALALVAPSRGGGGCAELACGERERCCDA  55
2mkv.1.A -------------------------------------------------------

Model_01 TNATAVRCCKLPLHAFLDNVGWFVRKLSGLLILLVLFAIGYFLQRIICPSPRRYP 110
2mkv.1.A ---------------------YETVRNGGLIFAGLAFIVGLLI--LLSRRFRSGG  52

Model_01 RGQARPGQRPGPPGGAGPLGGAGPPDDDDDSPALLRDEAAAGSQDSLLDSGGGGR 165
2mkv.1.A NKKRR--------------------------------------------------  57

Model_01 GRGGGGRSDPSCASEHEMRVVSPVFLQLPSYEEVKYLPTYEESMRLQQLSPGEVV 220
2mkv.1.A -------------------------------------------------------

Model_01 LPVSVLGRPRGGVAAEPDGGEGRYPLI                             247
2mkv.1.A ---------------------------
```