1. Primary and Secondary Structure Analysis

Yuejian Mo,11510511

Introduction

Protein's primary structure is sequence of aminod acid. We can find some conservation of same protein which vary on different organs and different species. So it is possbile to find out domain from these conservative sequence. Comparing amino acid sequence and certain algorithm, we can print the phylogenetic tree.

The information of amino acid sequence is not enough to build up protein with certain biological activity, but the secondary structure and higher dimemnsion structure do. Based on the features of known experiment result between primary structure and secondary structure, we can predict the helix, strand and turn structure in unkown protein.

Methods

In the first part, we selected 10 homology sequences of human UNC5b, including both orthologues and paralogues from uniprot.org. There are

> UNC5B_HUMAN,
> K7DDTY1_PANTR,K7B3D8_PANTR,K7BVI5_PANTR,H2NAM6_PONAB,G7N281_MACMU,
> F6UDU4_MACMU,F6UDV7_MACMU,F7A1Z5_CALJA,U3F339_CALJA

Then we aligned and generated a phylogenetic tree sequence with the Clustal Omega program in uniprot.org.
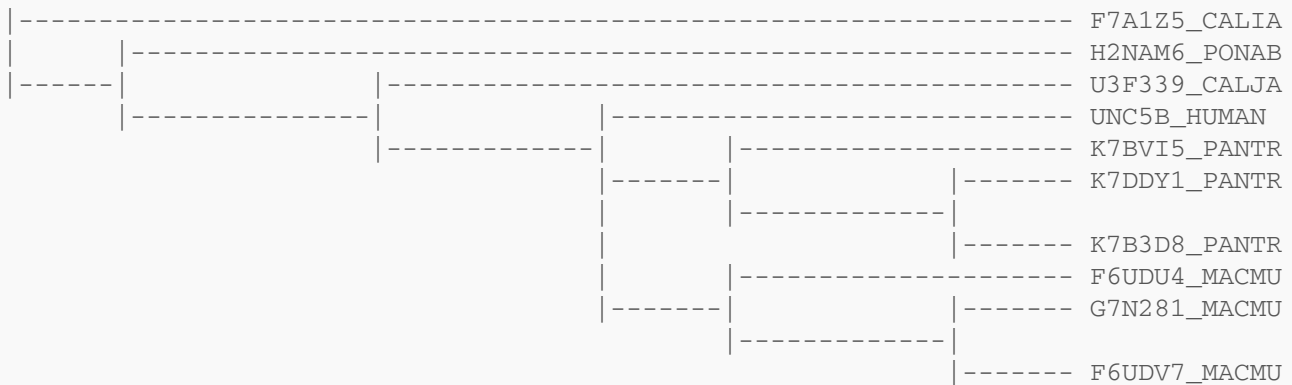
In the second part, we used JPred to predicte secondary structure of the cytoplasmic domain of human UNC5b. Then we compared the predicted result with X-ray structure(PDB:3G5B) to get error.

Last part, we design a truncation construct with minimal size, wheres retaining the overall fold. We keep all the E and H domain sequence and binding site for residue PO4A.

Results

Target Sequence and Aligment results

```
Q8IZJ1 UNC5B_HUMAN    359  TLSDPNSHLLEASGDAALYAGLVVAIFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   418
K7DDY1 K7DDYI_PANTR   359  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   418
K7B3D8 K7B3D8_PANTR   359  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   418
K7BVI5 K7BVI5_PANTR   359  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAILMAVGVVVYRHNCRDFDTDITDSSAALT   418
H2NAM6 H2NAM6_PONAB    73  TLSDPNSHLLEASGDAALYAGLVVAIFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   132
G7N281 G7N281_MACMU   359  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   418
F6UDU4 F6UDU4_MACMU   356  --------MLEASGDAALYAGLVVAVFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   407
F6UDV7 F6UDV7_MACMU   359  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAILMAVGVVVYRRNCRDFDTDITDSSAALT   418
F7A1Z5 F7A1Z5_CALJA   335  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAVLMAVGVVVYRRNCHDFDTDITDSSAALT   394
U3F339 U3F339_CALJA   359  TLSDPNSHLLEASGDAALYAGLVVAVFVVVAVLMAVGVVVYRRNCHDFDTDITDSSAALT   418
                                :.**************:.*****.**********:.**:*************

Q8IZJ1 UNC5B_HUMAN    419  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   478
K7DDY1 K7DDYI_PANTR   419  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   478
K7B3D8 K7B3D8_PANTR   419  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   478
K7BVI5 K7BVI5_PANTR   419  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   478
H2NAM6 H2NAM6_PONAB   133  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   192
G7N281 G7N281_MACMU   419  GGFHPVNFKTARPNNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   478
F6UDU4 F6UDU4_MACMU   408  GGFHPVNFKTARPNNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   467
F6UDV7 F6UDV7_MACMU   419  GGFHPVNFKTARPNNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP   478
F7A1Z5 F7A1Z5_CALJA   395  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPMCALQDSTDKILMTNSPLLDPLP   454
U3F339 U3F339_CALJA   419  GGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPMCALQDSTDKILMTNSPLLDPLP   478
                           *************.*.***************************:    *********  ***********

Q8IZJ1 UNC5B_HUMAN    479  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHFLHLRSASLGSQQLLGLP   538
K7DDY1 K7DDYI_PANTR   479  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHFLHLRSASLGSQQLLGLP   538
K7B3D8 K7B3D8_PANTR   479  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHFLHLRSASLGSQQLLGLP   538
```

```
K7BVI5  K7BVI5_PANTR    479  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHFLHLRSASLGSQQLLGLP  538
H2NAM6  H2NAM6_PONAB    193  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFAQDTHFLHLRSASLGSQQLLGLP  252
G7N281  G7N281_MACMU    479  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHLLHLRSASLGSQQLLGLP  538
F6UDU4  F6UDU4_MACMU    468  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFTRDTHLLHLRSASLGSQQLLGLP  527
F6UDV7  F6UDV7_MACMU    479  SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFTRDTHLLHLRSASLGSQQLLGLP  538
F7A1Z5  F7A1Z5_CALJA    455  SLKVKVYSSSTTGSGQGLTDGADLLGVLPPGTYPSDFTRDSHFLHLRSASLGSQQLLGLP  514
U3F339  U3F339_CALJA    479  SLKVKVYSSSTTGSGQGLTDGADLLGVLPPGTYPSDFTRDSYFLHLRSASLGSQQLLGLP  538
                             *********** **.:****************.::*:::*****************

Q8IZJ1  UNC5B_HUMAN     539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
K7DDY1  K7DDY1_PANTR    539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
K7B3D8  K7B3D8_PANTR    539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
K7BVI5  K7BVI5_PANTR    539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
H2NAM6  H2NAM6_PONAB    253  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  312
G7N281  G7N281_MACMU    539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
F6UDU4  F6UDU4_MACMU    528  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  587
F6UDV7  F6UDV7_MACMU    539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
F7A1Z5  F7A1Z5_CALJA    515  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPSGAIPQGKFYEMYLLINKAESTLPLSEGTQ  574
U3F339  U3F339_CALJA    539  RDPGSSVSGTFGCLGGRLSIPGTGVSLLVPSGAIPQGKFYEMYLLINKAESTLPLSEGTQ  598
                             ******************************.********************

Q8IZJ1  UNC5B_HUMAN     599  TVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLKTQAHQGHWEEVVTLDEETLN  658
K7DDY1  K7DDY1_PANTR    599  TVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLKTQAHQGHWEEVVTLDEETLN  658
K7B3D8  K7B3D8_PANTR    599  TVLSPSVTCGPTGLLLCHPVILTMPHCAEVSARDWIFQLKTQAHQGHWEEVVTLDEETLN  658
K7BVI5  K7BVI5_PANTR    599  TVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLKTQAHQGHWEEVVTLDEETLN  658
H2NAM6  H2NAM6_PONAB    313  TVLSPSVTCGPTGLLLCRPVILTLPHCAEVSARDWIFQLKTQAHQGHWEEVVTLDEETLN  372
G7N281  G7N281_MACMU    599  TVLSPSVTCGPTGLLLCRPVILTVPHCAKVSVGDWIFQLKTQAHQGHWEEVVTLDEETLN  658
F6UDU4  F6UDU4_MACMU    588  TVLSPSVTCGPTGLLLCRPVILTVPHCAKVSVGDWIFQLKTQAHQGHWEEVVTLDEETLN  647
F6UDV7  F6UDV7_MACMU    599  TVLSPSVTCGPTGLLLCRPVILTVPHCAKVSVGDWIFQLKTQAHQGHWEEVVTLDEETLN  658
F7A1Z5  F7A1Z5_CALJA    575  TVLSPSVTCGPAGLLLCRPVILTVPHCAEVSAGDWIFQLKTQAHQGHWEEVVTLDEETLN  634
U3F339  U3F339_CALJA    599  TVLSPSVTCGPAGLLLCRPVILTVPHCAEVSAGDWIFQLKTQAHQGHWEEVVTLDEETLN  658
                             ***********:*****.*****:****.**. ***************************

Q8IZJ1  UNC5B_HUMAN     659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  718
K7DDY1  K7DDY1_PANTR    659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  718
K7B3D8  K7B3D8_PANTR    659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  718
K7BVI5  K7BVI5_PANTR    659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  718
H2NAM6  H2NAM6_PONAB    373  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  432
G7N281  G7N281_MACMU    659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  718
F6UDU4  F6UDU4_MACMU    648  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  707
F6UDV7  F6UDV7_MACMU    659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE  718
F7A1Z5  F7A1Z5_CALJA    635  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPSLCTSLEYSLRVYCLE  694
U3F339  U3F339_CALJA    659  TPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPSLCTSLEYSLRVYCLE  718
                             ********************************************:***************

Q8IZJ1  UNC5B_HUMAN     719  DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
K7DDY1  K7DDY1_PANTR    719  DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
K7B3D8  K7B3D8_PANTR    719  DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
K7BVI5  K7BVI5_PANTR    719  DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
H2NAM6  H2NAM6_PONAB    433  DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWR-KLLAKYQEIP  491
G7N281  G7N281_MACMU    719  DTPVALKEVLDLERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
F6UDU4  F6UDU4_MACMU    708  DTPVALKEVLDLERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  767
F6UDV7  F6UDV7_MACMU    719  DTPVALKEVLDLERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
F7A1Z5  F7A1Z5_CALJA    695  DTPVALKEVLELERTLGGYLVEEPKPLVFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  754
U3F339  U3F339_CALJA    719  DTPVALKEVLELERTLGGYLVEEPKPLVFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIP  778
                             **********:*****************:***************************** **********

Q8IZJ1  UNC5B_HUMAN     779  FYHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  833
K7DDY1  K7DDY1_PANTR    779  FCHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  833
K7B3D8  K7B3D8_PANTR    779  FCHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  833
K7BVI5  K7BVI5_PANTR    779  FCHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  833
H2NAM6  H2NAM6_PONAB    492  FYHIWSGSQKALHCTFTLERHSLASTELICKICVRQVEGEGQIFQLHT-----TLAETPA  546
G7N281  G7N281_MACMU    779  FYHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  833
F6UDU4  F6UDU4_MACMU    768  FYHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHTTLLHTTLAETPA  827
F6UDV7  F6UDV7_MACMU    779  FYHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TL--TPA  831
F7A1Z5  F7A1Z5_CALJA    755  FYHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  809
U3F339  U3F339_CALJA    779  FYHIWSGSQKALHCTFTLERHSLASTELTCKICVRQVEGEGQIFQLHT-----TLAETPA  833
                             * *************************************************   **  ***

K7BVI5  K7BVI5_PANTR    834  GSLDTLCSAPGSTVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  893
H2NAM6  H2NAM6_PONAB    547  GSLDTLCSAPGSTITTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  606
G7N281  G7N281_MACMU    834  GSLDTLCSAPGSAVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  893
F6UDU4  F6UDU4_MACMU    828  GSLDTLCSAPGSAVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  887
F6UDV7  F6UDV7_MACMU    832  GSLDTLCSAPGSAVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  891
F7A1Z5  F7A1Z5_CALJA    810  GSLDTLCSAPGGAITTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  869
U3F339  U3F339_CALJA    834  GSLDTLCSAPGGAITTQLGPYAFKIPLSIRQKICNSLDAPNSRGNDWRMLAQKLSMDRYL  893
                             ***********.::.:***********************************************

Q8IZJ1  UNC5B_HUMAN     894  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  945
K7DDY1  K7DDY1_PANTR    894  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  945
K7B3D8  K7B3D8_PANTR    894  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  945
K7BVI5  K7BVI5_PANTR    894  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  945
H2NAM6  H2NAM6_PONAB    607  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  658
G7N281  G7N281_MACMU    894  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  945
F6UDU4  F6UDU4_MACMU    888  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  939
F6UDV7  F6UDV7_MACMU    892  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  943
F7A1Z5  F7A1Z5_CALJA    870  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  921
U3F339  U3F339_CALJA    894  NYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC  945
                             ***************************************************
```

Phylogenetic Tree

```
    |------------------------------------------------------------ F7A1Z5_CALIA
    |       |---------------------------------------------------- H2NAM6_PONAB
    |------|                  |--------------------------------- U3F339_CALJA
           |--------------|    |--------------------------- UNC5B_HUMAN
                          |-----------|    |------------------- K7BVI5_PANTR
                                 |-------|          |------- K7DDY1_PANTR
                                 |        |-----------|
                                 |                   |------ K7B3D8_PANTR
                                 |       |------------------- F6UDU4_MACMU
                                 |------|          |------ G7N281_MACMU
                                        |-----------|
                                                    |------ F6UDV7_MACMU
```

```
Seq :
YRRNCRDFDTDITDSSAALTGGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP
Pre :-----------EE-----------------------------------------------------------------
Exp :
Err :            12


SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHFLHLRSASLGSQQLLGLPRDPGSSVSGTFGCLGGRLSI
-------------------------------------------------------------EEEEEEE---EEEEE
                                                              EEEEE     EEEEE
                                                              3


PGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQTVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLK
E--EEEEEE---------EEEEEEE-------------EEEEEEEEE-------EEEEEEE--------EEEEE
    EEEEE          EEEEEEE              EEE   EEE          EEEEE          EEEEE
4                  5                   6  78910  11         12    13         14


TQAHQGHWEEVVTLDEETLNTPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE
E-------EEE-----------EEEEE---EEEEEE-----HHHHH-------EEEEEEEE---------EEEEEE--
EE      EEEEE          EEEEE   EEEEE   EEEEEEEE        EEEEEEEEEE       EEEEEEEEEE
15      16  17                        1819    20       21      2223      2425    2627


DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIPFYHIWSGSQKALHCTFTLER
----HHHHHHHHHH----EEE----EEEEE----EEEEEEE------E------EEEEE---HH--------EEEEEE-
   HHHHHHHHHHHHHHH EEE    EEEEE    EEEEEE       EE         EEE HHHHH       EEEEEEEE
2829           3031 32 33  34    35                36      37  383940       41      42


HSLASTELTCKICVRQVEGEGQIFQLHTTLAETPAGSLDTLCSAPGSTVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGN
--------EEEEEEE-------EEEE-------------------------------HHHHHHHHHH---------
        EEEEEEEEEE   EEEEEEEEE                           EEE       HHHHHHHHHH
        43      4445   464748 4950                        515253             54


DWRMLAQKLSMDRYLNYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC
HHHHHHHH------HHHHH----HHHHHHHHHHH-----HHHHHHHHHH---HHHHHHHHH----
 HHHHHHH       HHHH   HHHHHHHHHHHHH     HHHHHHHHH    HHHHHHH
55            56                57     58          59        60


--------------------------------------------------------------------------
```

Sequence total length: 339 Error ratio= (mismatch of predition compare eith correct)/(total sequence length)=60/339=17%

Minisize sequence:

The minisize sequence are marked as *

```
Seq     :
YRRNCRDFDTDITDSSAALTGGFHPVNFKTARPSNPQLLHPSVPPDLTASAGIYRGPVYALQDSTDKIPMTNSPLLDPLP
Known   :
Design  :


SLKVKVYSSSTTGSGPGLADGADLLGVLPPGTYPSDFARDTHFLHLRSASLGSQQLLGLPRDPGSSVSGTFGCLGGRLSI
                                                            EEEEEE      EEEEE
                                                            * * * * * *   * * * * *


PGTGVSLLVPNGAIPQGKFYEMYLLINKAESTLPLSEGTQTVLSPSVTCGPTGLLLCRPVILTMPHCAEVSARDWIFQLK
      EEEEEE          EEEEEEE              EEE      EEE           EEEEE         EEEEE
      * * * * * *        * * * * * * *        * * *     * * *         * * * * *       * * * * *


TQAHQGHWEEVVTLDEETLNTPCYCQLEPRACHILLDQLGTYVFTGESYSRSAVKRLQLAVFAPALCTSLEYSLRVYCLE
EE      EEEEE          EEEEE   EEEEE   EEEEEEEE     EEEEEEEEEE      EEEEEEEEEE
* *       * * * * *        * * * * *   * * * * * *   * * * * * * * *   * * * * * * * * * *   * * * * * * * * * *


DTPVALKEVLELERTLGGYLVEEPKPLMFKDSYHNLRLSLHDLPHAHWRSKLLAKYQEIPFYHIWSGSQKALHCTFTLER
   HHHHHHHHHHHHHH EEE     EEEEE        EEEEE       EE       EEE HHHHH          EEEEEEE
   * * * * * * * * * * * * * * * * *   * * * * *        * * * * * *         * *       * * *   * * * * *         * * * * * * *


HSLASTELTCKICVRQVEGEGQIFQLHTTLAETPAGSLDTLCSAPGSTVTTQLGPYAFKIPLSIRQKICNSLDAPNSRGN
       EEEEEEEEEE     EEEEEEEEE                               EEE          HHHHHHHHHHH
       * * * * * * * * * *     * * * * * * * * *                             * * *          * * * * * * * * * * *
DWRMLAQKLSMDRYLNYFATKASPTGVILDLWEALQQDDGDLNSLASALEEMGKSEMLVAVATDGDC
  HHHHHHH           HHHH     HHHHHHHHHHHHH       HHHHHHHHH       HHHHHHH
  * * * * * * *          * * * *     * * * * * * * * * * * * *       * * * * * * * * *       * * * * * * *
```

Conclusions

Here are conservation of protein in different organs and speices, which suggested the evolution of protein. Using cetain technique, secondary structure can be predicted in acceptable range. But we think that the different of our choose target is small and not enough to show the essential domain. And we risks to keep beta strand and alpha helix.