Investigate of Zea Mays protein: GRMZM2G105004

1. Summary[100-200 words]:

This protein is found in Zea mays. Due to the lack of research, we are not sure the function of this protein right now.

By the following analysis, we can see this protein contains two domains: PB1 and RWP-RK. PB1 can works in protein binding and RWP-RK is involved in nitrogen-controlled development like dimerisation and DNA binding.

Both of these domain are highly related to dimerisation. It is a reasonable assumption that this protein is related to cell dimerisation activity.

## Basic genomic information: Chromosome, gene IDs, links to NCBI etc.

Chr. 10

Gene ID: 100191188

Link: <https://www.ncbi.nlm.nih.gov/protein/1284804524>

## Basic physic-chemical properties

Number of amino acids: 945

Molecular weight: 102830.44

Theoretical pI: 6.06

Amino acid composition:

Ala (A) 70 7.4%

Arg (R) 50 5.3%

Asn (N) 41 4.3%

Asp (D) 59 6.2%

Cys (C) 20 2.1%

Gln (Q) 38 4.0%

Glu (E) 49 5.2%

Gly (G) 67 7.1%

His (H) 25 2.6%

Ile (I) 36 3.8%

Leu (L) 81 8.6%

Lys (K) 45 4.8%

Met (M) 25 2.6%

Phe (F) 32 3.4%

Pro (P) 54 5.7%

Ser (S) 117 12.4%

Thr (T) 49 5.2%

Trp (W) 8 0.8%

Tyr (Y) 21 2.2%

Val (V) 58 6.1%

Pyl (O) 0 0.0%

Sec (U) 0 0.0%

(B) 0 0.0%

(Z) 0 0.0%

(X) 0 0.0%

Total number of negatively charged residues (Asp + Glu): 108

Total number of positively charged residues (Arg + Lys): 95

Atomic composition:

Carbon C 4458

Hydrogen H 7053

Nitrogen N 1277

Oxygen O 1428

Sulfur S 45

Formula: C4458H7053N1277O1428S45

Total number of atoms: 14261

Extinction coefficients:

Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.

Ext. coefficient 76540

Abs 0.1% (=1 g/l) 0.744, assuming all pairs of Cys residues form cystines

Ext. coefficient 75290

Abs 0.1% (=1 g/l) 0.732, 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:

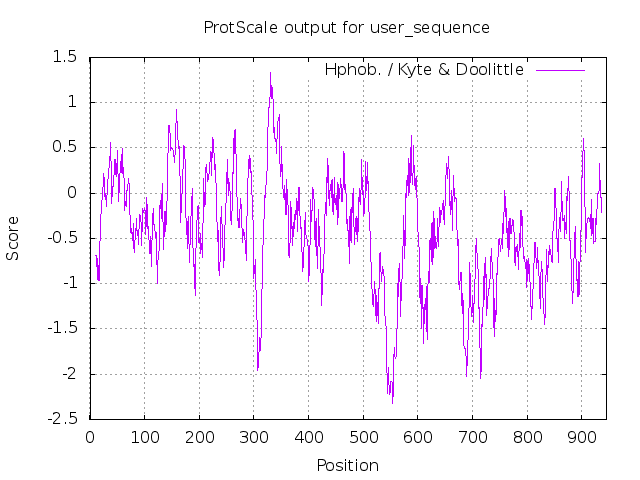
The instability index (II) is computed to be 52.81

This classifies the protein as unstable.

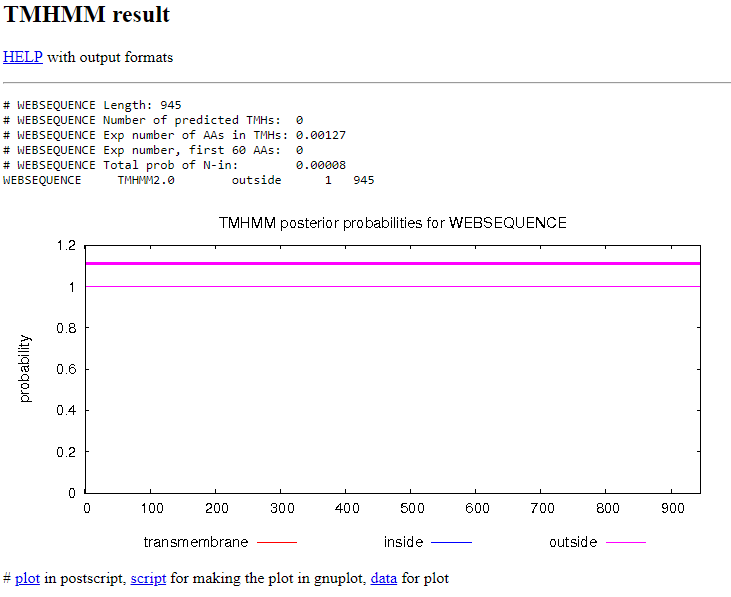
Aliphatic index: 73.49

Grand average of hydropathicity (GRAVY): -0.406

## Transmembrane domain prediction



Based on ProtScale, this protein has no transmembrane domain



TMHMM result confirmed our result.

## PROSITE patterns

Using Gene Fasta, I found the following pattern.

FT MYHIT 157 165 pat:2FE2S\_FER\_1 [!]

FT MYHIT 638 646 pat:2FE2S\_FER\_1 [!]

FT MYHIT 1185 1193 pat:2FE2S\_FER\_1 [!]

FT MYHIT 1454 1462 pat:2FE2S\_FER\_1 [!]

FT MYHIT 1538 1546 pat:2FE2S\_FER\_1 [!]

FT MYHIT 2078 2086 pat:2FE2S\_FER\_1 [!]

FT MYHIT 2129 2137 pat:2FE2S\_FER\_1 [!]

FT MYHIT 3418 3426 pat:2FE2S\_FER\_1 [!]

FT MYHIT 223 234 pat:4FE4S\_FER\_1 [!]

FT MYHIT 59 93 pat:ANAPHYLATOXIN\_1 [!]

FT MYHIT 143 177 pat:ANAPHYLATOXIN\_1 [!]

FT MYHIT 204 239 pat:ANAPHYLATOXIN\_1 [!]

FT MYHIT 2115 2150 pat:ANAPHYLATOXIN\_1 [!]

FT MYHIT 2943 2972 pat:ANAPHYLATOXIN\_1 [!]

FT MYHIT 2830 2867 pat:CTCK\_1 [!]

FT MYHIT 93 122 pat:DEFENSIN [!]

FT MYHIT 193 221 pat:DEFENSIN [!]

FT MYHIT 3409 3438 pat:DEFENSIN [!]

FT MYHIT 175 186 pat:EGF\_1 [!]

FT MYHIT 193 204 pat:EGF\_1 [!]

FT MYHIT 1693 1704 pat:EGF\_1 [!]

FT MYHIT 2460 2471 pat:EGF\_1 [!]

FT MYHIT 2474 2485 pat:EGF\_1 [!]

FT MYHIT 2504 2515 pat:EGF\_1 [!]

FT MYHIT 3126 3137 pat:EGF\_1 [!]

FT MYHIT 3263 3274 pat:EGF\_1 [!]

FT MYHIT 3396 3407 pat:EGF\_1 [!]

FT MYHIT 3610 3621 pat:EGF\_1 [!]

FT MYHIT 317 331 pat:INTEGRIN\_BETA [!]

FT MYHIT 3055 3068 pat:THIOLASE\_3 [!]

FT MYHIT 3246 3259 pat:THIOLASE\_3 [!]

FT MYHIT 3565 3578 pat:THIOLASE\_3 [!]

FT MYHIT 56 102 pat:VWFC\_1 [!]

FT MYHIT 152 203 pat:VWFC\_1 [!]

FT MYHIT 920 967 pat:VWFC\_1 [!]

FT MYHIT 1193 1238 pat:VWFC\_1 [!]

FT MYHIT 2858 2906 pat:VWFC\_1 [!]

FT MYHIT 3113 3152 pat:VWFC\_1 [!]

FT MYHIT 3329 3383 pat:VWFC\_1 [!]

FT MYHIT 3427 3474 pat:VWFC\_1 [!]

## Protein domains

[PS51519](https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS51519)   **RWP\_RK**   *RWP-RK domain profile* :

|  |  |
| --- | --- |
| **556 - 637:** | **score =** 17.006 |

MRVADKKRSASEK---NFSLDVLRKYFSGSLRDAAMSLGVCPTTLKRICRQHGISRWPSR

KINKVNRSLKKIEKVIKSVHGVDRS

**Predicted feature:**

|  |  |  |  |
| --- | --- | --- | --- |
| [DOMAIN](https://www.uniprot.org/help/domain) | 556 | 637 | RWP-RK |

[PS51745](https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS51745)   **PB1**   *PB1 domain profile* :

|  |  |
| --- | --- |
| **842 - 924:** | **score =** 23.473 |

ALTVKATYNGDT-VRFKFVPARGWYHLLEEIAKRFKL--TAGaFQLKYKDDEDEWVILAN

DADLQECMDVLDSIssRNMKLQVRDL

**Predicted feature:**

|  |  |  |  |
| --- | --- | --- | --- |
| [DOMAIN](https://www.uniprot.org/help/domain) | 842 | 924 | PB1 |

## Subcellular localization

SeqID: NM\_001358370.1 Zea mays uncharacterized LOC100191188 (LOC100191188), mRNA

Analysis Report:

CMSVM+ Unknown [No details]

CWSVM+ Unknown [No details]

CytoSVM+ Unknown [No details]

ECSVM+ Extracellular [No details]

ModHMM+ Unknown [No internal helices found]

Motif+ Unknown [No motifs found]

Profile+ Unknown [No matches to profiles found]

SCL-BLAST+ Unknown [No matches against database]

SCL-BLASTe+ Unknown [No matches against database]

Signal+ Unknown [No signal peptide detected]

Localization Scores:

Cytoplasmic 0.24

CytoplasmicMembrane 0.05

Cellwall 0.80

Extracellular 8.91

Final Prediction:

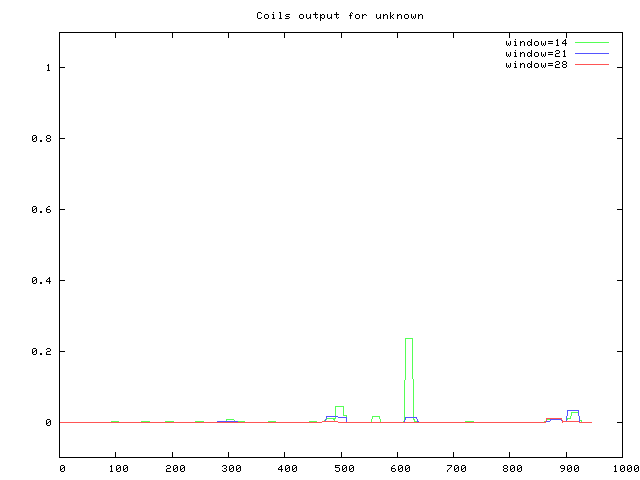
Extracellular 8.91

## Secondary structure prediction using algorithms [here](https://www.expasy.org/resources/search/keywords:secondary%20structure%20prediction)

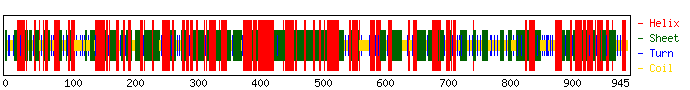
### APSSP:

Does not work

### Coils:



CFSSP:



**Secondary Structure:**

\* \* \* \* \* \*

Query 1 MDVPTPPNRAGCNSNIGSPMQSFDDPFGVAAMTSFDGYSELCSPSVADHIFSLLNDPSAAQQMVAMWSSL 70

Helix 1 HHHHHHHHHHHHHHH HHHHHHHH HHHHHHHHHHH 70

Sheet 1 EEEE EEEEEEE EEEEEEEEE EEEE EEEEEEEE EEEEEEEEE 70

Turns 1 TT T T T T TT 70

Struc 1 EEEECCTTCCCCCTCEETEHHHHHTHEHHHHEEHECCEEEECCTCHHEEEEEHCCTTCHHEHHHHHEEEC 70

\* \* \* \* \* \*

Query 71 GSSPRASAVREDMSFDTFPGPVDGTSSLAQRFNSAAASSPTGVDRGLKESDGLVLPSNGSQQGSSIIPRS 140

Helix 71 HHHHHHHHHH HHHHHHHHH HHH 140

Sheet 71 EEEEE EEEEEEE EE EEEEEE 140

Turns 71 T T T T T T T T T T T T T T 140

Struc 71 CTCTCHHHHHHHHEEEECCTCCCCETEHHEEHHHHHCCTCCCCCTCCCTCTCEECCTCCCEEEEEECHTH 140

\* \* \* \* \* \*

Query 141 VGNALADRMLMALSLFRKSLSDGVLAQVWMPVEHNGRVVLSTSEQPYLLDQDLAGYREVSRNFLFSVKEE 210

Helix 141 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH HHHHHHHHHHHHHH HHHHHHHHHHH 210

Sheet 141 EEEEEEEEEEEEEEEEEEEEEEEEEEEEE EEEEEEEEEEEEEEEE EEEEEEEE 210

Turns 141 T T T T T TT T T T 210

Struc 141 HHHHHHHHHHHHEEHHHTHHEEEEEEEEEHHHHCCTCEEEHHHTEEEHHHHHECCCCCEEEEEEEHHHTH 210

\* \* \* \* \* \*

Query 211 PGLHLGLPGRVFISGVPEWTSSVIYYSKPEYLRMEHALLHEIQGSLAMPVYDPSKGSCCAVLELVTNKEK 280

Helix 211 HH HHHHHH HHHHHHHHHHHHHHHHHHHH HHHHHHHHHHH 280

Sheet 211 EEEEEEEEEEEEEEEEEEEEEEE EEEEEEEEEEEEEEEEEEEE EEEEEEEE 280

Turns 211 T T T TT T T T T 280

Struc 211 THEEEEEEEEEEEEHHEEEEEEEEECCTTHEHHHHHHHHHHEETHHHEEECCTCCTCEEHHHHEEHHHHH 280

\* \* \* \* \* \*

Query 281 PNFDAEMDIVCDALQAVNLQTTTDRSNQKVYSENQKSAFTEILDVLRAICHAHMLPLALTWVPTSNGIAN 350

Helix 281 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 350

Sheet 281 EEEEEEEEEEEEEEEE EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE EEE 350

Turns 281 T T T T T TT T T 350

Struc 281 THHHHHEEEEHHHEEEEEEEEEHHTHHEEEHHHHHHEHHHHEEEEHEEEHHHHHHHHEEEEEECCTCEET 350

\* \* \* \* \* \*

Query 351 GFCVGKNIGIDPQPGKAALRIHESACYVNDAKMQGFLHACADRHLEKGQGIAGRALKSNLPFFSPDIREY 420

Helix 351 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 420

Sheet 351 EEEEEEEEE EE EEEEEEE EEEEEEEEEEE 420

Turns 351 T T T T T T T TT T T 420

Struc 351 EEEEETEEECCTHHHHHHHHHHHHHEEHHHHHEEEHHHHHHHHHHHTHHHHHHHHHHTHEEEEETEEEEE 420

\* \* \* \* \* \*

Query 421 SIEDYPLAHHARKFGLHAAVAIRLRSTYTSYDDYILEFFLPVNCKGCGEQQMLLNNLSSTMQRICKSLRT 490

Helix 421 HHHHHHHHHHHHHHHHHHHHHH HHHHHHHHHH HHHHHHHHHHHHHHHHHHHHHH 490

Sheet 421 EEEEE EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 490

Turns 421 T T TT T 490

Struc 421 EEHEEHHHHHHTHHHHHHHEHEHEEEEEEEEEEEHHHEEEEEEEEEEEHHHEEEEEEHHHEEEEEHEEEH 490

\* \* \* \* \* \*

Query 491 VSEAEIENVSATAAMFEKTSGTCLPTGNSESSSHDDQPITESALQDLSLGDKQGDREPDKPQRSSMRVAD 560

Helix 491 HHHHHHHHHHHHHHHHH HHHHHHHHHHHHHHHH HHHHH 560

Sheet 491 EEEEEEEEEEEEEE EEEEEEEEEEEE 560

Turns 491 T T T TT TT T T 560

Struc 491 HHHHHHHHHHHHHHHHHTEEEEEEECCTCCCTHHHTEHHHHHHHHHEHCCTCCCCCCCCCCCCTCHHHHH 560

\* \* \* \* \* \*

Query 561 KKRSASEKNFSLDVLRKYFSGSLRDAAMSLGVCPTTLKRICRQHGISRWPSRKINKVNRSLKKIEKVIKS 630

Helix 561 HHHHHHHHHHHHHHHHHH HHHHHHHHH HH HHHHHHHHHHH 630

Sheet 561 EEEEEEEEE EEEEEEEEEEEEEEEE EEEEEEE 630

Turns 561 T T TT T T TT T 630

Struc 561 THTHHHHTHHEEEEEEEECCTCHHHHHHEEEEEEEEEEEEEEECCCCCCCTTHHCCCCTHHHHHHEEEEE 630

\* \* \* \* \* \*

Query 631 VHGVDRSLQYDPATGSLVPATSLPEKMPFSACDTLPTSSLGRAVEETCSPKSEQDFSSPDGWQRETSQFH 700

Helix 631 H HHHHHHHHHHHHHH HHHHHHHHHHHHHHHH 700

Sheet 631 EEEEEEEEE EEEEEEE EEEEEEEEEE EEEE EEEEEEEEE 700

Turns 631 T T T T T T T T T T T T T 700

Struc 631 EEEEEEEEECCTCEEEEEHHHHHHHHHHEHHEEEEECTCHHHHHHTEHHTHHHTHCCTCTCEETEEEEEE 700

\* \* \* \* \* \*

Query 701 VSGIPKRGGDEVRTLANNNKGRRNYVSGIANITQHSNSEGTHGPSYPNPIGAVNSLHTGETGNIDSLTSL 770

Helix 701 HH HH 770

Sheet 701 E EEEE EEEEEEEEE EEEEE EEEEEE 770

Turns 701 T T T T T T T T T 770

Struc 701 ECCCCTCCCTHEEEECCTCCCCTCEEEEEEEEECCCTCTCCCCTCCCTCCCCEEEEECCCCTCEEEEEEC 770

\* \* \* \* \* \*

Query 771 HPSMDGIEGQTTVRNSPFVQQADVTMVDGHDTKEQTLPSTSGMTDSSSGSASSEPTFKGNPGRALKDRSS 840

Helix 771 HHHHHHHHHHHHHHHHHHH HHHHH 840

Sheet 771 EEEEEEEEEEEEEEEEEEEE EEEEEEEEEEEE 840

Turns 771 T T TT TT T T T TTT T T T T 840

Struc 771 CTCCCCEETEEEEETEEEEHHHEEEHHHHHHHHHEEEETEEEECCTCCTCCCTTTCCCTCCTCCCHTHHT 840

\* \* \* \* \* \*

Query 841 PALTVKATYNGDTVRFKFVPARGWYHLLEEIAKRFKLTAGAFQLKYKDDEDEWVILANDADLQECMDVLD 910

Helix 841 HHHHHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 910

Sheet 841 EEEEEEEEEEEEEEEEEEEEEEEEEEEEE EEEEEEEEEEE EEE EEEEEEEEE 910

Turns 841 T T T T T T 910

Struc 841 HHEHHEEEEETEEEEEEEHHHTEEEHHHHHHHHHEHEHHHHHEEHHHTHHHHEEEHHTHHHHHHEHHEEE 910

\* \* \*

Query 911 SISSRNMKLQVRDLSCHVGSSGSSSSCLQVEAHSS 945

Helix 911 HHHHHHHHHHHHHHH HHHHHH 945

Sheet 911 EE EEEEEEEEEEEE EEE 945

Turns 911 TT T T 945

Struc 911 EEHTHHHEEEEHEEEEECCTCCCCTCEHHHHHCCC 945

Total Residues: H: 538 E: 528 T: 132

Percent: H: 56.9 E: 55.9 T: 14.0

## 3D structure prediction (optional)

## Genomic location

Map the protein sequence of your adopted protein in Assignment 9 to the Zea Maize genome using tblastn. Write down the chromosome and coordinate.

Chr: 10

Coordinate: 128604452-128609487

I authorize Dr. Xijin Ge to post this document on the internet to help researchers study this protein.

If you do not want to be involved, uncheck this box above.

Please submit a Word document.