# Computational Analysis and Characterization of RLM3 gene, A Disease Resistance gene of Brassica Rapa

Sumit Joshi , Shivani Banchariya

**ABSTRACT:**

Also known as stem canker , Black leg is one of the most economically important disease of Brassica family. Black leg is caused by a necrotrophic fungus Leptosphaeria Maculans that competes with plant for carbon source making various portions such as stem and leaves dead resulting in patches in various parts of plant. It has been experimentally found that there are 14 Leptosphaeria Maculans resistance genes discovered till yet. Previous studies have identified RLm3 as a TIR gene that acts as a resistance factor for a genotype that is susceptible to Black leg in Brassica. In this study protein encoded by Rlm3 gene was selected. Predicted sequence of the selected protein was obtained from NCBI (XM_009115633.2) for computational analysis. For further analysis primary, secondary and tertiary structures of this protein were obtained. Results obtained showed that query protein was acidic, hydrophilic and stable. Secondary structure of protein showed that protein contained 221 alpha helix (52.74%), 32 beta turns (7.88%) and 92 random coils (21.96%). Protein localization revealed that protein was present in cytoplasm of the cell. In I TASSER predicted 3D model model1 with confidence score -3.07 was found to be most productive. The results of I TASSER were validated by phyre. Further more tests were carried out such as determining function of RLM3 protein, Determination of Active site , homology modelling , determining localization and sub cellular localization of protein , Multiple sequence alignment , protein-protein interaction , Validation of predicted sequence of RLM3 , bond angle was also determined , Ramachandran plot was predicted for RLM3 along with plotting phylogenetic tree. Protein minimization was carried out in presence of Cl and water for 20 Nano seconds.



**INTRODUCTION**:

The Brassicaceae family includes most important plants along with model plant of plant kingdom Arabidopsis thaliana. With many varieties such as B.rapa, B.napus, B.juncea, B.campestris and B. oleracea. Brassica stands as a very important crop used for oil seed production. Brassica is grown worldwide in many parts to fulfil the oil seed demands of the world. The genus Brassica is one of the 51 genera in the tribe Brassicaceae belonging to the crucifer family and is most important gene within this tribe. Containing 37 species (Gomez-Campo 1980). Brassica family includes many

crop species that provide edible roots, leaves, stem, buds, flowers and seeds. Vegetables of Brassica are of great importance all over the globe. Over the last 10 years Brassica production has increased and it has become one of the most important source of vegetable oil after soyabean and cotton. Largest Expanse of Brassica is found in India followed by China and Canada. (Rakow G., 2004). But due to low seed yields the output production of seeds in India is much less than that of China and Canada.

Canada and Australia get their major share of economy by exporting Rape seed. While China and Japan stand to be the largest importers of oil seed. Last 3 decades have witnessed an enormous growth in cultivation and production of canola seeds (B.napus) as it is being rapidly grown all over the world becoming the second largest crop being cultivated and produced worldwide. Canola seeds have witnessed an estimated production of 67.91 million tons annually (United States Department of Agriculture 2017). Canola is one of the largest cash crop in Canada with 18.4 million tons of production in year 2016(Statistics Canada 2016). In year 2016 the contribution of canola to Canadian economy was $26.5 billion. About 90% of the canola seeds (meal and oil) are exported to approximately 55 foreign markets worldwide (Canola council of Canada 2017).

In Australia rapeseed production began in 1960s (cutting 1975) with many varieties that were introduced from Canada. Realizing the potential of new cash crop the area under cultivation greatly increased in 1970s. Thus canola Produced began a major source of economy for Australia along with Canada.

Canola or rapeseed is the fourth largest cultivated crop in China. China stands as the leading producer of canola worldwide since 1980/1981.According to USDA data. Canola production by china has reached 19.7million tones in 2014-2015 (Qiong Hu et al Rapeseed research and production in China, 2017).

India proudly stands as the third largest producer contributing 1% of the total production worldwide (Arvind et al). Area under cultivation is greatly increasing day by day thus turning green revolution to yellow revolution.

Caused by Leptosphaeria maculans Blackleg is an important disease of oilseed rape. It has caused an extensive crop loss in many parts of the world including Western Australia. (McGee DC, Petrie GA, 1979). Blackleg has estimated to cause $900 million of crop loss throughout the world every year (Howlett 2003; Fitt et al 2006). Severe Blackleg infection can result in complete loss of Brassica napus canola or oilseed rape crops (Li et al 2003; Rouxel et al.2003)

In Canada in 1980s Blackleg caused 50% of the yield loss in individual fields (Fisk et al 1997). In 1990s many black leg resistance were released that controlled Blackleg until 2005 (Kutcher, H. R., et al. 2013). Again in 2012 Blackleg became susceptible in Canada resulting in major loss to canola production (Hwang, Sheau-Fang et al 2016)

In 1972 due to Canadian cultivars Blackleg proved to be highly susceptible to Australian cultivars and by the end of 1972 there was a major Blackleg epidemic in Australia posing a major threat to industry (Murray and Stovold 1970; Bokor et al. 1975; Wightman 1982). Due to Blackleg the area under cultivation has decreased from 49,000 hectare in 1972 to 2000 hectare in 1999 (Bokor, et al 1975)

China was also not left untouched with the effect of Blackleg but the impact of this disease was very less because of strict import ban on canola import from Canada that was the major exporter of canola to China in order to prevent its crops from the effect of Blackleg. The major organism responsible for blackleg is Leptosapharia maculans that is a necrotrophic fungus. It is believed that fungus competes with plant for carbon source thus making some part of the plant dry thus residing in it and feeding it. Leptosapharia maculans is believed to spread with the help of spores and single spore can destroy the entire cultivation.

Necrotrophic term is given to such organism that kills the host, lives in it and reproduces on the energy derived from the dead cells. Until recent years gene to gene interaction for resistance in Arabidopsis thaliana has only been reported for biotrophic pathogens (Tsuda et al 2005).

Recent studies have predicted that gene to gene type of resistance in Arabidopsis thaliana is present. Whereas earlier studies described resistance for only for biotrophic fungus both in case of Arabidopsis and Brassica. (Delourme R et al. 2006; Staal et al 2006). The resistance based on gene to gene type of resistance in biotrophic response interactions led to salicyclic acid dependent response. On the other hand defense mechanism towards necrotrophs showed that the resistance was mainly due to a set of defense responses that require jasmonic acid and ethylene that are mostly regulated by complex trait (Denby et al 2004). Leptosphaeria maculas being a hemibiotrophic fungus shows necrotrophic nature but clear gene to gene relation with Brassica host is not shown (Fitt, Bruce DL, et al. 2006). Most interestingly resistance to Leptosphaeria maculans was found to require a resistant gene (dominant) that carried the same protein structure(TIR-NB-LRR)as is required for resistance against some biotrophs ( Staal et al 2006). Most probably gene to gene relationship for both Leptosphaeria maculans and Brassica share same family of resistance genes but this is yet to be proved (Saal et al 2005). But Arabidopsis on the other hand in comparison to B.napus is said to show a strong resistance gene independent for invasion response that indicates that Arabidopsis – Leptosphaeria maculans model is of a non-host interaction. (Elliott and Howlett 2008)

Leptosapharia maculans is able to attack all parts of the plant including cotyledon, stem, pores etc. (Gabrielson RL, 1983). The pathogen is believed to cause both leaf lesions and stem canker (west et al 2001). The first part effected during infection is cotyledons or the true leaves then infection goes down to stem and roots causing severe damage in form of stem canker( Huang et al 2016). The fungus is believed to survive on infected bodies and other parts of crop residues for several years and is capable of producing for sexual and asexual fruiting bodies (west et al 2001).

Leptosapharia maculans has both sexual and asexual stages on host plant and can be both monocyclic and polycyclic according to the source of inoculum (Li et al 2007, b). Both ascospores and pycnidiospores can get attached to cotyledons or young leaves, germinate and produce hyphae that penetrate through stomata or wounds (Li et al 2004). Stem canker is observed at the end of growing season even if no leaf lesions are visible at earlier stages of growth.

Infection of seedling occurs due to invasion of leaves by Leptosapharia maculans through leaves. The tissue is initially colonized as biotroph but behind the hyphal front the fungus becomes necrotrophic and asexual fruiting bodies known

as pycnidia are produced in the dead tissue (Hammond et al 1985; Hammond and lewis 1987). Pycnidiospores act as secondary inoculum and are spread to other leaves and neighboring plants by rain and air. The fungus invades and destroys the cells of stem cortex resulting in a canker that possibly completely destroys the base of stem (Hammond et al 1985)

A bank of characterized Leptosapharia maculans mediated by Agrobacterium tumefaciens was developed that analyzed the role of pathogenicity genes but the pathogenicity mechanism of Leptosapharia maculans has been largely unstudied (Howlett, Barbara J. 2004)

In 2002 using reverse genetic approach Idnurm and Howlett found that isocytrate lyase encoded by isocytrate lyase gene (*icl1*) is found to be essential for pathogenicity of Leptosapharia maculans. Studies made till date state a few related genes have been functionally studied principally including *THIOL* gene (Elliott and Howlett 2006), the *Ipa*gene (Elliott and Howlett 2008), the *Lmpma1*gene (Remyet al.2008a), the *Lmgpi15* gene (Remyet al.2008b), the *LmIFRD g*ene (van de Wouwet al.2009b), *Lmepi*gene (Remyet al.2009), and the *LmSNF1*gene (Feng et al 2014).

It is also said that Leptosapharia maculans has ability to produce phytotoxins that are essential for virulence. Sirodesmin PL is the well-studied phytotoxin said to be responsible for virulence of blackleg. (Rouxel, Thierry, et al 1988; Elliot et al 2011)

Large number of ascospores are released from infected plant that that contribute to a great increase in black leg severity (Wherret et al 2004). The integration of genetic resistance and cultural strategies such as tillage, fungicide, crop rotation etc are able to affect the concentration of ascospores to a great extent. (West JS et al 2005). Combination of appropriate crop rotation and tillage have proven to reduce amount of airborne inoculum and infection level in Canada (Guo et al 2005). Various disease resistance genes have been isolated but they are not much effective for long duration. The most important approach of controlling blackleg is through genetic breeding and by use of resistance canola varieties (Rimmer 2006, Kutcher et al 2011, and Kutcher et al 2013). But the durability of long term objectives of resistance genes gets effected by the biology of pathogen and its ability to undergo mutation and recombination of antivirulance genes. (Kutcher et al 2011, Howlett et al 2015). In order to prevent blackleg infection destruction of tillage has been recommended. Australians recommend to grow new canola seeds at least 500m away from old scribble. (Marcroft et al 2003). Fungicides have also not proven much effective to Blackleg (Huang et al 2011). Till date 18 Leptosapharia maculans resistance genes have been conferred.

| *R gene* | Originated from | Chromosome | References |
|---|---|---|---|
| *Rlm1* | *B. napus* | A7 | Ferreria et al. 1995 |

| | | | |
|---|---|---|---|
| *Rlm3* | *B. napus* | A7 | Ansan Melayah et al. 1998 |
| *Rlm4* | *B. napus* | A7 | Zhu and Rimmer, 2003 |
| *Rlm7* | *B. napus* | A7 | Rimmer 2006 |
| *Rlm9* | *B. napus* | A7 | Delourme et al. 2006 |
| *BLMR1* | *B. napus* | A10 | Long et al. 2011 |
| *BLMR2/RlmS* | *B. napus* | A10 | Van de Wouw et al. 2009; Long et al. 2011; Larkan et al |
| *LepR1* | *B.rapa ssp.sylvestris* | A2 | Yu et al. 2005 |
| *LepR2* | *B.rapa ssp.sylvestris* | A10 | Yu et al. 2007 |
| *LepR3* | *B.rapa ssp.sylvestris* | A10 | Larkan et al. 2013 |

| | | | |
|---|---|---|---|
| *Rlm2* | *B.rapa ssp.sylvestris* | A10 | Mayerhofer et al. 1997, Larkan et al. 2015 |
| *LepR4* | *B.rapa ssp.sylvestris* | A6 | Yu et al. 2008 |
| *Rlm8* | *B. rapa* | ----- | Balesdent et al. 2002 |
| *Rlm11* | *B. rapa* | ----- | Balesdent et al. 2013 |
| *Rlm5* | *B. juncea* | ----- | Chèvre et al. 1997 |
| *Rlm6* | *B. juncea* | B8 | Balesdent et al. 2002 |
| *Rlm10* | *B.nigra* | B4 | Chèvre et al. 1996; Eber et al. 2011 |



Fig.1: Genes conferring resistance to blackleg in Brassica

Studies have conferred that during 1990s RLm 3 gene was most present in cultivars. Recent studies have identified RLm3 as a protein that is resistant to Leptosphaeria maculans and some other necrotrophic fungi.

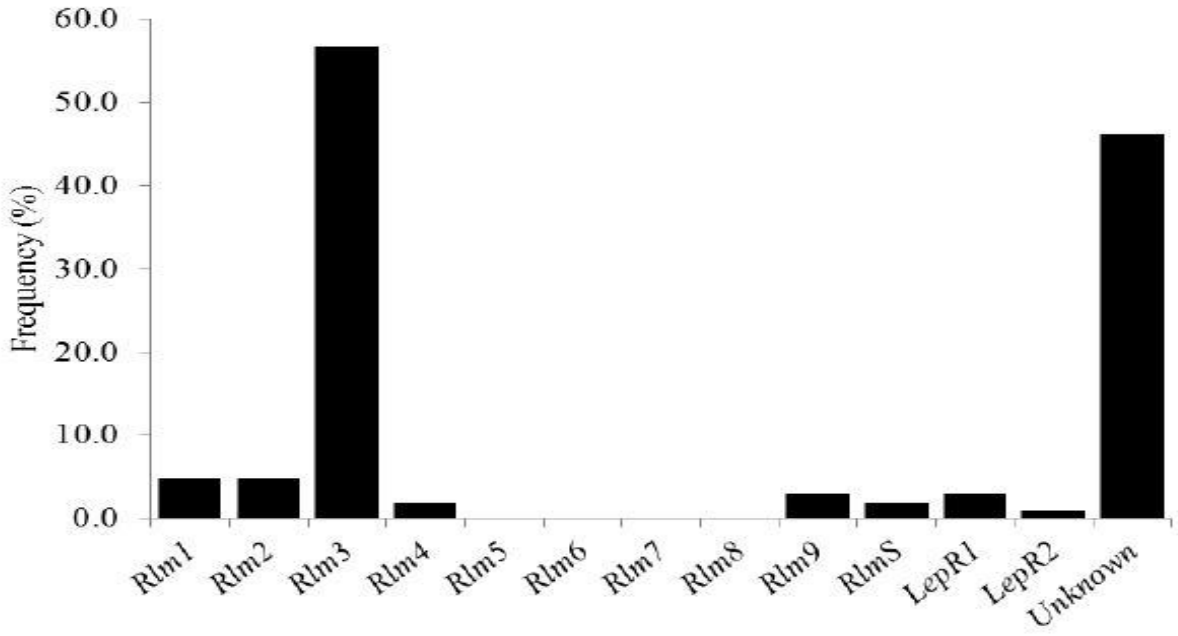



Fig 2: Efficiency (in %) of various genes offering resistance to blackleg

**MATERIAL AND METHOD:**

**Retrieval of sequence:** we obtained the sequence of 1526 bp locus with accession number XM-00915633.2.from NCBI(https://www.ncbi.nlm.nih.gov ) .We used megablast in order to compare the query protein with the available database and identified the sequence that resembles our query protein above a certain threshold. We used the tool Protparam (https://web.expasy.org/protparam/) available in expassy server in order to obtain the primary structure of protein. Protparam provided us with complete primary sequence of query protein. . We used SOPMA (Self-Optimized Prediction Method with Alignment) tool (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html ) available in the server of Institute of biology and plant chemistry. Another tool used by us in order to obtain the secondary structure of protein was PSIPRED (PSI-blast based secondary structure prediction) tool that is available in the server of University of California. The major benefit of using PSIPRED is that it assembles various structure prediction methods under a single platform. We used I-TASSER (Iterative Threading ASSEmbly Refinement) tool (https://zhanglab.ccmb.med.umich.edu/I-TASSER ) that is available in the server of university of Michigan. Biological function of protein is defined by its 3D structure. High quality 3D model of the query protein is generated by ITASSER. We used Verify 3D **(**http://servicesn.mbi.ucla.edu/Verify3d) tool available in the server of university of California to verify our 3D structure of query protein. The Composability of an atomic 3D model with its own amino acids is determined by verify 3D. We used another tool that is known as ProSA (protein structure analysis) tool (https://prosa.services.came.sbg.ac.at/prosa.php ) available in the server of Structural bioinformatics group, to validate our 3D structure. It's the most widely used tool that checks the 3D structure of protein for potential errors.Prediction of active site was done by using active site prediction tool (http://scfbio-iitd.res.in/dock/ActiveSite.jsp ) available in the

server of supercomputing facility of IIT Delhi. Binding site refers to the residues that form temporary bond with the substrates and reaction takes place. To find the available binding site residues we used COACH tool available in the server (https://zhanglab.ccmb.med.umich.edu/COACH ) of university of Michigan to predict the available binding sites and the available ligands that bind with our desired protein.

We carried out the analysis of 3D structure using PROVE (http://servicesn.mbi.ucla.edu/PROVE ) that is available in the server of University of California. We drew Ramachandran plot with the help of PROCHECK (http://servicesn.mbi.ucla.edu/PROCHECK) available in the server of University of California. (Molecular biology institute). We used INTERPRO (http://www.ebi.ac.uk/interpro ) available in the server of European Bioinformatics Institute in order to predict the function of query protein. . We carried out prediction of subcellular localization with the help of CELLO tool (http://cello.life.nctu.edu.tw ) available National chiao Tung university. . MEMSAT SVM tool (http://bioinf.cs.ucl.ac.uk/psipred/?memsatsvm ) present in the server of University of California was used to predict Membrane Structure and Topology. SWISS MODEL tool (https://swissmodel.expasy.org ) available in EXPASY server was used in order to compute Homology Modelling. CLUSTAL OMEGA (https://www.ebi.ac.uk/Tools/msa/clustalo) tool available in the server of European Bioinformatics Institute. Was used to carry out multiple sequence alignment. We used STRING tool (https://string-db.org) that has been developed by a consortium of academic institutions including CPR, EMBL, KU, SIB, TUD and UZH in order to find protein protein interaction of our query protein. We constructed phylogenetic tree using MEGA software version 7 that was downloaded from (https://www.megasoftware.net)



**RESULTS AND DISCUSSIONS:**

**Retrieval of sequence:** After going through the database of NCBI (https://www.ncbi.nlm.nih.gov ) we obtained a sequence of RLm3. The predicted sequence thus obtained by us is as follows (https://www.ncbi.nlm.nih.gov/nuccore/XM_009115633.2?report=fasta ).

**PREDICTED: Brassica rapa disease resistance protein RLM3 (LOC103839142), mRNA**

NCBI Reference Sequence: XM_009115633.2

GenBank Graphics

>XM_009115633.2 PREDICTED: Brassica rapa disease resistance protein RLM3 (LOC103839142), mRNA

GACCATTCTGCTTATGGACGTTTTAAAGCAGAATATTAATTATGCTTTTTCATAATAATATATAAAGTTT
GACAAAAAAAAATCAAGTTGTACATAATGTGAGATACCCTTTCTCCAATTTTCATATGGTTGACTTTATTT
AACTTTTGAAATGACGTCTTATTGAAGAAGTGGTTCCAATGACAGTGTCCGGTGAAGAAGTCCCACCGCA
GCACCAAGTTTTCATCAACTATAGAGGAGACGAGCTGCGGAAAAGCTTCCTCGGGTTCGTAGTGAAAGCC
ATGCGAGAAGCAAAGATCAACGTCTTCACAGACGAAATAGAAGTCAAAGGTAGAGATCTACAGAATCTTT
TCAGGAGGATCGAAGAATCCAGAGTCGCTGTTGCGATCTTGTCTGAAAGATACACTGAATCTAGTTGGTG
CTTGGATGAACTAGTGAAGATGAAAGAGCAGATGGACCAAGACAAGCTCGTTGTGATTCCAATTTTTTAC
AGATTGGATGCTACCAACTGTAAGAGGCTCGTTGGAACTTTCGGTGACAATTTCAGGAATCTGGAGAGAG
AGTATCGGAGCGAGCCTGAGAGAGTCAAGAAATGGAAAGAAGCTTTGATCTCTATTCCCCAGAAGGTTGG
CTTGACTTTGGAAGGACACAGGGATGAGTCTGAGTTGGTCGATTCAATCGTTAAGGAGGTGAAGAAAGTT
CTAATCGATGTTTCAAACAAGAAAAGAGGTCAGTCCACAAATCACCGAACCCAAACTGACATCATATCAT
CATTTGAGCCTCTTGTGGCAGAGCTGGACCGACTACCATCACGTCTCCCTCAAGTGTTTGTCAGTTTCTG
CAAAGAGGAGCTTGGCGACAACTTTGTCAGACATCTCGCGTGGGCTTTGAGAGAATTAGGGATCAATGTA
CTCATGGACAGCTACAACCGTAGAGGAAAAGAGGACGAACAACAACAAGTATTCACGAATATAGAGAAGT
CGAATATCGTTCTAGCCATCTTCTCAAAAAGGTACTCGGAATCAGATAGGTGCTTGAACGAGCTTGTGAA
GATGGAAGAGCTTACAAAGGAAGGAAAACTCGTGGTCATCCCCGTCTTCTACAATGTGAAAACAAACGAA
GTGAGAAGACTCCAAGGAGAGTTTGGAATCCATTTTGCAGATTCAGTGAAGAGGTTTTCTATGGAGCCAA
TGATGGTCCAGAGCTGGGAGGAAGCTTTGAATTTTATTATCAAAAGGAGTACTGGCTTAAGCTTGGAAAG
ACACAGAAACGAATTTGCTCTTGTTGCTGCTATTGTTAAAGAAGTTACAAGATTGCTACCAAGCTCTNCG
AAGAATGAGAAGAAGAAGGAGATTAGGGATTGGAGAAGTTTTTATCAGGGCTTTAACAGCGACTTTTCTA
TTTGCTCTCTTCATATCCCCTATACGTCGTACCCCTGATGTGAACTTTCTCAAAATTGGTAACTTGCTAG
TTGGGTTTCCGTTTTTGGTTGTTGTCTTGCACAAAATTCTATGTTAACAAAAACAA

Fig 3: figure showing the predicted nucleotide sequence of RLm3 obtained from NCBI database



**BLAST:** The results of Megablast explained that on total there are 13 sequences that match our query protein. Three sequences with accession number NM_123868.4 and AK221580.1 have 86% cover with query protein , sequences with accession number CP002688.1 and AB010693.1 have 79% cover with query protein , sequence with accession number XM_010443500.2 has 47% cover with query protein , sequences with accession number XM_019232620.1 and XM_019242103.1 have 35% cover with query protein, sequence with accession number XM_013754668.1 has 33% cover with query protein, sequence with accession number LT669793.1 has 21% cover with query protein, sequence with accession number NM_123869.4 has 12% cover with query protein, sequence with accession number XM_013802756.2 has 6% cover with query protein, sequences with accession number XM_015304013.1 and XM_015304012.1 have 2% cover with query protein.

**Sequences producing significant alignments:**

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|---|---|---|---|---|---|---|
| PREDICTED: Brassica rapa disease resistance protein RLM3 (LOC103839142), mRNA | 2815 | 2815 | 100% | 0.0 | 100% | XM_009115633.2 |
| Arabidopsis thaliana Disease resistance protein (TIR-NBS-LRR class) family mRNA | 1223 | 1223 | 86% | 0.0 | 84% | NM_123868.4 |
| Arabidopsis thaliana mRNA for putative protein, complete cds, clone: RAFL07-83-N01 | 1223 | 1223 | 86% | 0.0 | 84% | AK221580.1 |
| PREDICTED: Brassica oleracea var. oleracea vesicle-associated protein 1-4-like (LOC106316797), mRNA | 861 | 861 | 33% | 0.0 | 97% | XM_013754668.1 |
| PREDICTED: Camelina sativa vesicle-associated protein 1-4-like (LOC104724929), transcript variant X1, mRNA | 699 | 699 | 47% | 0.0 | 84% | XM_010443500.2 |
| PREDICTED: Camelina sativa vesicle-associated protein 1-4-like (LOC104724929), transcript variant X2, mRNA | 590 | 590 | 35% | 6e-164 | 86% | XM_019232620.1 |
| PREDICTED: Camelina sativa disease resistance-like protein CSA1 (LOC104771696), mRNA | 566 | 566 | 35% | 1e-156 | 86% | XM_019242103.1 |
| Arabidopsis thaliana chromosome 5 sequence | 497 | 1187 | 79% | 4e-136 | 86% | CP002688.1 |
| Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K21C13 | 497 | 1187 | 79% | 4e-136 | 86% | AB010693.1 |
| Arabis alpina genome assembly, chromosome: 6 | 185 | 290 | 21% | 3e-42 | 80% | LT669793.1 |
| PREDICTED: Brassica napus putative F-box/FBD/LRR-repeat protein At5g44950 (LOC106362933), mRNA | 178 | 178 | 6% | 5e-40 | 98% | XM_013802756.2 |
| Arabidopsis thaliana DSS1 homolog on chromosome V (DSS1(V)), mRNA | 171 | 171 | 12% | 9e-38 | 83% | NM_123869.4 |
| PREDICTED: Solanum tuberosum TMV resistance protein N-like (LOC107058378), mRNA | 56.5 | 56.5 | 2% | 0.003 | 97% | XM_015304013.1 |
| PREDICTED: Solanum tuberosum uncharacterized LOC102590743 (LOC102590743), mRNA | 54.7 | 54.7 | 2% | 0.009 | 97% | XM_015304012.1 |



Fig 4: figure showing the results of BLAST where 13 sequences match our query sequence.

**Primary sequence analysis:** Primary sequence analysis reveals the primary structure of protein. Data obtained from Protparam reveals that our query protein contains 1526 amino acids with total molecular weight of 124640.81. The chemical formula for our query protein is $C_{4635}H_{7765}N_{1527}O_{1943}S_{271.}$ Atomic composition reveals that our query protein contains 4635 atoms of carbon, 7765 atoms of hydrogen, 1527 atoms of oxygen, and 271 atoms of sulphur. Amino acid composition reveals that our query protein contains 480 molecules of Alanine (31.5%), 415 molecules of Threonine (27.2%), 359 molecules of glycine (23.5%), 271 molecules of cysteine (17.8%) and 1 molecule of Asparagine (.1%). We also came to know that at pH 5.03 our query protein has no net charge. Instability index 37.10 reveals that our query protein is stable. The aliphatic index of 37.10 reveals that query protein is thermostable. Hydrophobic nature of query protein is revealed by the positive value of hydropathy (**GRAVY**). The hydropathy (**GRAVY**) value of our query protein is.723.

Number of amino acids: 1526

Molecular weight: 124640.81

Theoretical pI: 5.03

Amino acid composition: CSV format

| | | |
|---|---|---|
| Ala (A) | 480 | 31.5% |
| Arg (R) | 0 | 0.0% |
| Asn (N) | 1 | 0.1% |
| Asp (D) | 0 | 0.0% |
| Cys (C) | 271 | 17.8% |
| Gln (Q) | 0 | 0.0% |
| Glu (E) | 0 | 0.0% |
| Gly (G) | 359 | 23.5% |
| His (H) | 0 | 0.0% |
| Ile (I) | 0 | 0.0% |
| Leu (L) | 0 | 0.0% |
| Lys (K) | 0 | 0.0% |
| Met (M) | 0 | 0.0% |
| Phe (F) | 0 | 0.0% |
| Pro (P) | 0 | 0.0% |
| Ser (S) | 0 | 0.0% |
| Thr (T) | 415 | 27.2% |
| Trp (W) | 0 | 0.0% |
| Tyr (Y) | 0 | 0.0% |
| Val (V) | 0 | 0.0% |
| 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): 0
Total number of positively charged residues (Arg + Lys): 0

Aliphatic index: 31.45

Grand average of hydropathicity (GRAVY): 0.723

**Atomic composition:**

| | | |
|---|---|---|
| Carbon | C | 4635 |
| Hydrogen | H | 7745 |
| Nitrogen | N | 1527 |
| Oxygen | O | 1943 |
| Sulfur | S | 271 |

**Formula:** $C_{4635}H_{7745}N_{1527}O_{1943}S_{271}$
**Total number of atoms:** 16121

**Extinction coefficients:**

This protein does not contain any Trp residues. Experience shows that
this could result in more than 10% error in the computed extinction coefficient.

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

| | |
|---|---|
| Ext. coefficient | 16875 |
| Abs 0.1% (=1 g/l) | 0.135, assuming all pairs of Cys residues form cystines |

| | |
|---|---|
| Ext. coefficient | 0 |
| Abs 0.1% (=1 g/l) | 0.000, assuming all Cys residues are reduced |

**Estimated half-life:**

The N-terminal of the sequence considered is G (Gly).

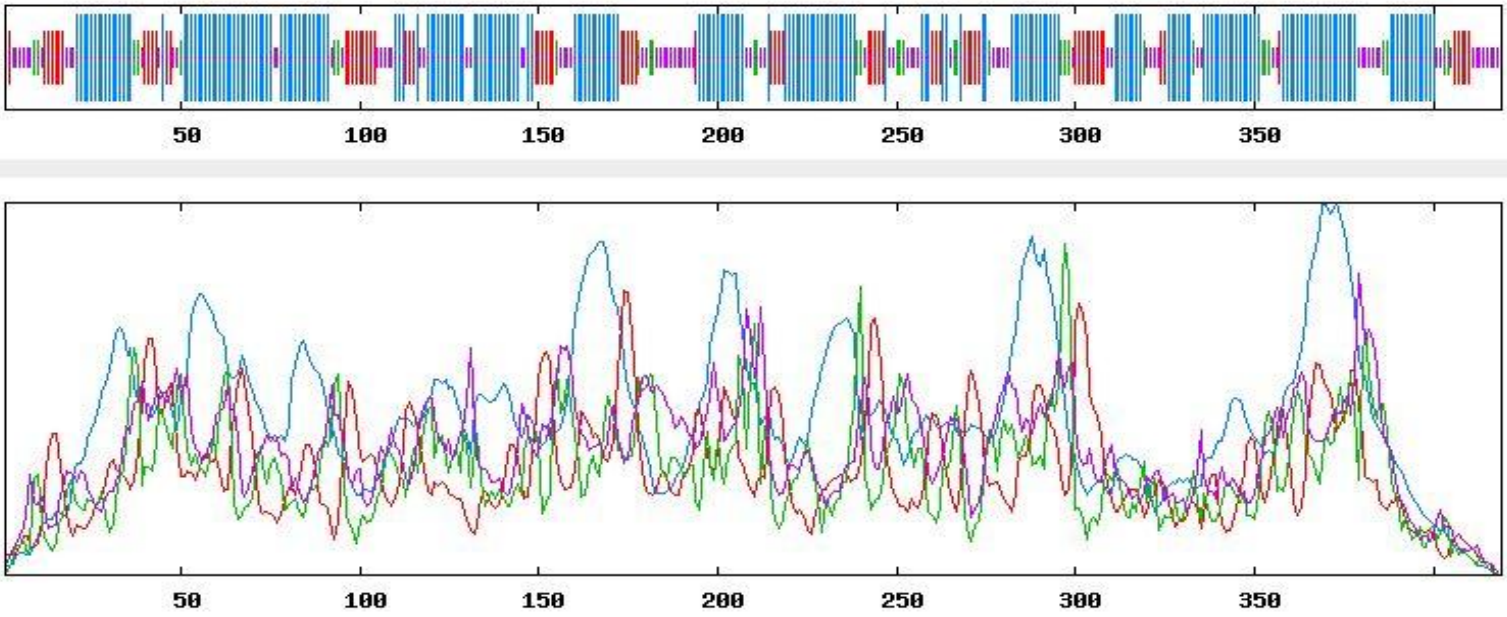
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

**Instability index:**

The instability index (II) is computed to be 37.10
This classifies the protein as stable.



Fig 5: Molecular characterization (Primary structure) of Query protein




```
Parameters :
   Window width          :  17
   Similarity threshold  :   8
   Number of states      :   4
```

**SECONDARY STRUCTURE ANALYSIS :** Secondary Structure analysis using SOPMA reveals that our query protein contains 221 alpha helixes (Hh) 52.74%, 92 random coils (Cc) 21.96%, 73 extended strands (Ee) 17.42%, 33 beta turns(Tt)7.88% and does not contain any $3_{10}$ helix(Gg), pi helix(Ii), beta bridge(Bb),bend region(Ss), ambiguous states and other states .



Fig 6: Secondary structure of Query protein

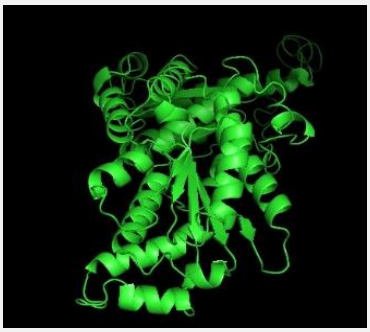
**Tertiary Structure Prediction**: Total 10 tertiary structures were generated through ITASSER. Tertiary structure or 3D model generated through ITASSER predicted that model 1 with C score -3.07, Estimated TM-score = 0.37±0.12 and Estimated RMSD = 14.5±3.7Å was the most efficient structure obtained. We carried out this complete study taking this structure.

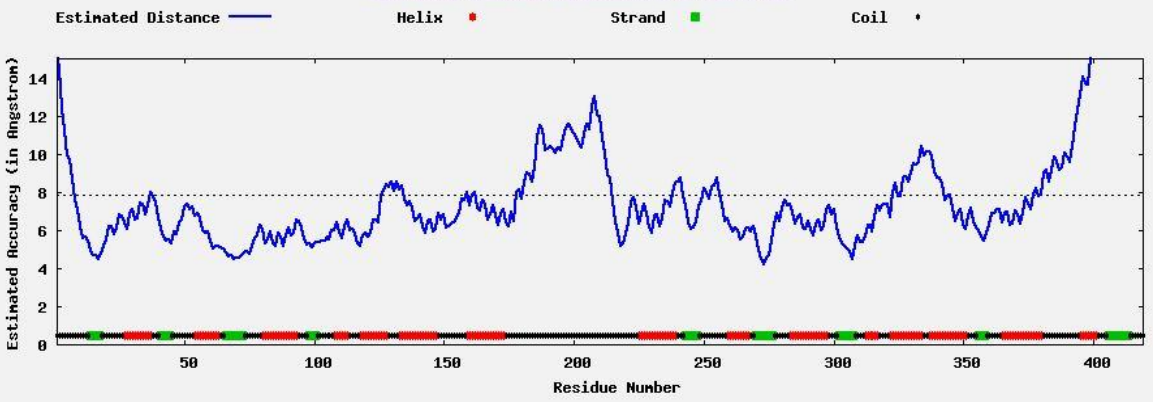


```
SOPMA :
   Alpha helix     (Hh) :   221 is  52.74%
```

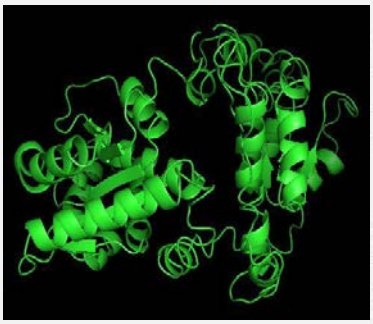
**Generated 3D models**

Download the estimated local accuracy of models



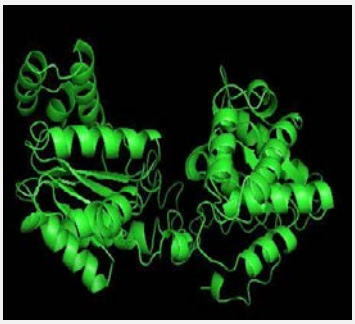


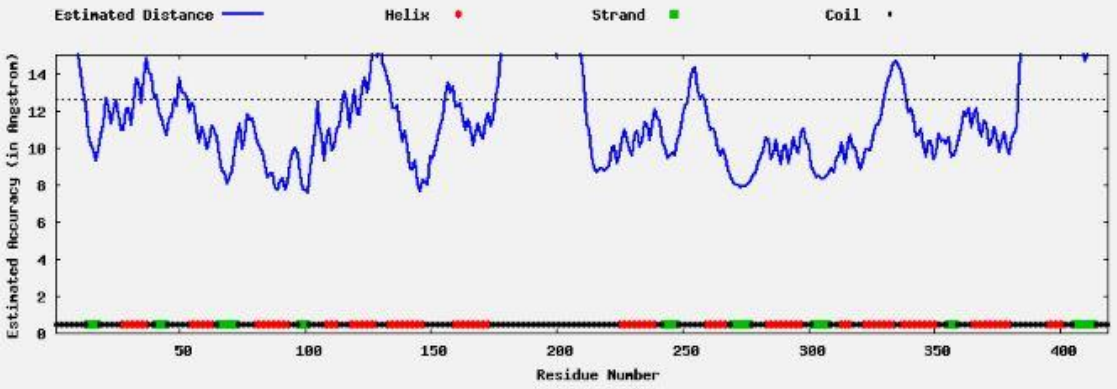

- Download Model 1
- C-score=-3.07 (Read more about C-score)
- Estimated TM-score = 0.37±0.12
- Estimated RMSD = 14.5±3.7Å






- Download Model 2
- C-score = -3.53






- Download Model 3
- C-score = -3.53

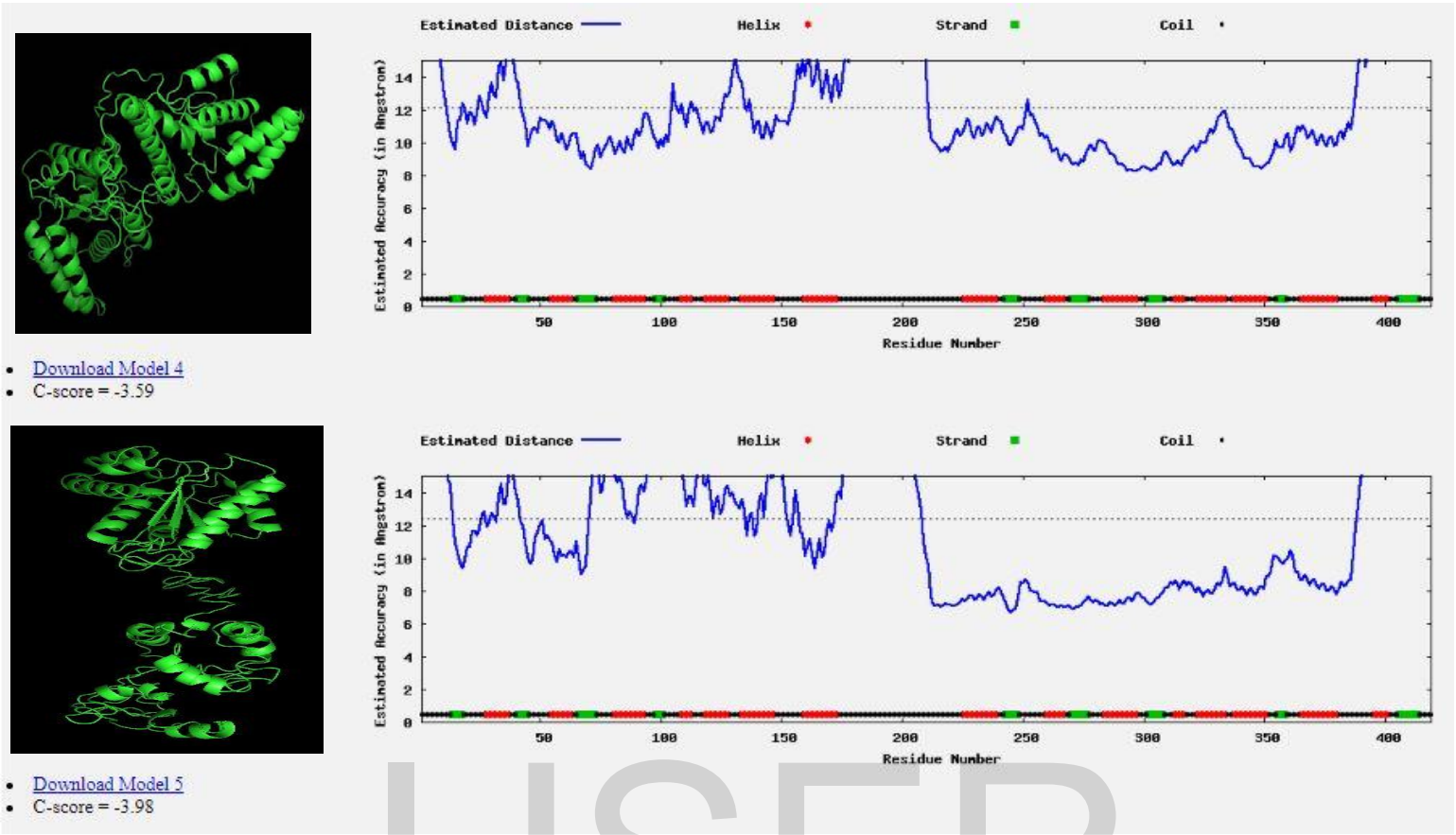



Fig 7: Tertiary structure (3D structure ) of query protein

**Validation of 3D structure:** The results obtained by ProSA predicted that Z score of our protein is -8.76. That means that the occurrence of model is equivalent to a model generated by X-ray diffraction with a scoring value of -8.76. Local model quality id depicted by considering the score below the threshold line calculated quantitatively for all residues of modeled proteins.

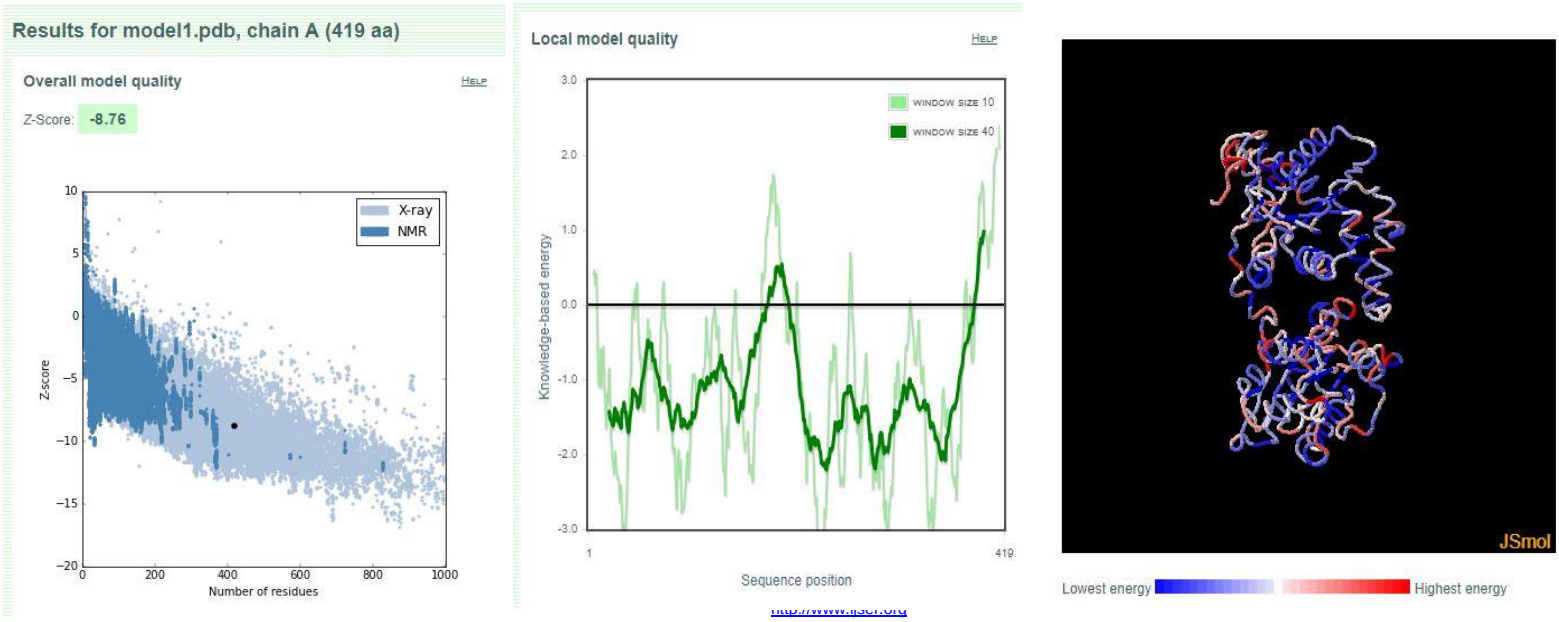
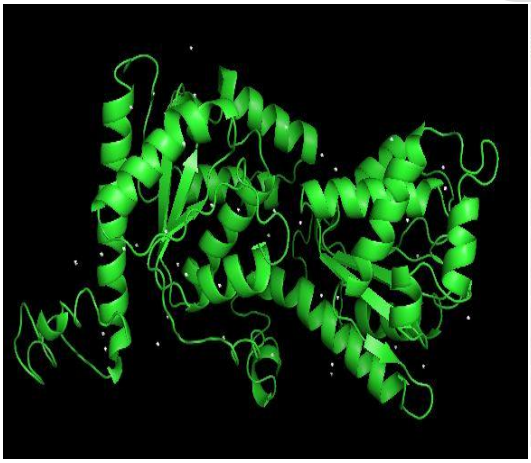


Fig 8: (a) overall quality of model (b) local quality of model (c) energy model

**Active site and binding site:** By the results of active site prediction we came to know that there are total 66 cavities present in our protein that act as active site to which our ligand can bind.

| Cavities | |
|---|---|
| cavity_1_TRQDAKEVFPISLGMN | cavity_2_DRQLPTAEKVNFSIYGMCW |
| cavity_3_RQKEDTVPIGNLFSYMC | cavity_4_QKEDIGRPSNYLVMWF |
| cavity_5_KEFLWVDPRSYQGN | cavity_6_VKLFREQNDITSHMGWP |
| cavity_7_KSLEIPNRVCGWTF | cavity_8_KELNSCPGVWRTFI |
| cavity_9_EQLFKVIDRYMSNCAWPH | cavity_10_KSREMVDYLNQFPHIA |
| cavity_11_ILGEKNDRMTVSHFQWP | cavity_12_KREYMNDSCLQVFWHIGA |
| cavity_13_LGKRDEVATMINQFHPS | cavity_14_YSNPTFIHCKLQGREWVA |
| cavity_15_HRTEDQAKLSVPFGI | cavity_16_TRDQNVAKEFPIGLSM |
| cavity_17_PRYTNLQKGVIMEDWASFH | cavity_18_DRLEASVPFKITQNH |
| cavity_19_PRYTDKFSLNICQEVG | cavity_20_FSERKLGDVTWINQP |
| cavity_21_GKQERDVMFYSNCLHI | cavity_22_PICHNKYSLERFTDWVA |
| cavity_23_ERDSLFKITVQAHM | cavity_24_SVRLTEAKFGIHNQ |
| cavity_25_RLQDEKSMFWPHNVIAG | cavity_26_YNDTQKLREVIGMWAFSC |
| cavity_27_RPDQYLSNTEKVIGMFA | cavity_28_YLTIPKSAVNWREGH |
| cavity_29_QVKEDFRTLWNPSGM | cavity_30_TRVNGKQLIDFSEWP |
| cavity_31_NHERYKVTSGIQ | cavity_32_FKAGELRVTMISNQHP |
| cavity_33_TDENRFSLYWPVKG | cavity_34_EKRVNSGMFQHWPLA |
| cavity_35_MWRASNYVDHGFKCE | cavity_36_EADLPSKFVIGTQNR |
| cavity_37_EVMKFLSWHNQRDIGPA | cavity_38_SKRELNFVHDIGAQ |
| cavity_39_PRYTLSNQKGVIEMDFA | cavity_40_EALTSRGNHKIVQ |
| cavity_41_PRYTSHNLCQGVIMEAWKF | cavity_42_FELRWPDYSVKGQ |
| cavity_43_KDTIRAEHVNMSQ | cavity_44_HTRDEGNVAKQLFPSYI |
| cavity_45_TNYRQGVIMELWAPKFH | cavity_46_DVGNERLTFY |
| cavity_47_LAVWFHRKENTGS | cavity_48_TRFEYNSLKDVW |
| cavity_49_FPYSTRKINHCLQE | cavity_50_RLMVDWEAYSNFHGK |
| cavity_51_FKISLCERNPWYV | cavity_52_DESYFKLAGRVT |
| cavity_53_LKENSCVIGFMTRA | cavity_54_REWAMLVFDHGNY |
| cavity_55_DQTLGNERMFKIVYSCA | cavity_56_HREGDTLFSY |
| cavity_57_VGLTRFDENKSWPY | cavity_58_RVLDIGNFKSEWYP |
| cavity_59_RELDSVWTKGF | cavity_60_RKIQLFPSEMWDV |
| cavity_61_WELAVFHRKNTSG | cavity_62_TDEVRGNAKLFPIY |
| cavity_63_YPSTNLCQGVIREK | cavity_64_DESYFKLAGRVT |
| cavity_65_MASTRLGFKIV | cavity_66_LGPVIED |





(a)

| Rank | C-score | Cluster size | PDB Hit | Lig Name | Download Complex | Consensus Binding Residues |
|---|---|---|---|---|---|---|
| 1 | 0.09 | 5 | 1hloA | Nuc.Acid | Rep, Mult | 375,378 |
| 2 | 0.05 | 3 | 2wpnA | SBY | Rep, Mult | 213,216,271,273,373,377 |
| 3 | 0.05 | 3 | 1e1yA | PO3 | Rep, Mult | 64,65,66,92 |
| 4 | 0.04 | 2 | 1p29A | GLC | Rep, Mult | 89,90 |
| 5 | 0.04 | 2 | 3qf1A | PZE | N/A | 74,77 |
| 6 | 0.02 | 1 | 2yl5A | MG | N/A | 140,150 |
| 7 | 0.02 | 1 | 3cejA | AVF | Rep, Mult | 310,311,315,318 |
| 8 | 0.02 | 1 | N/A | N/A | N/A | 93,95,99,149,249,252,254,255,257,258,261,262,284,286,287,288,290,291,333,339,343 |
| 9 | 0.02 | 1 | 1s0vC | Nuc.Acid | N/A | 316,374 |
| 10 | 0.02 | 1 | 3bb1H | MG | N/A | 161,247 |

(b)

(c) (d)

Fig 8 : (a & b ) cavities present in our query protein (c)binding sites in 3D structure (c) Binding residues and ligands in our query protein

Structure Analysis: The results of prove showed the difference between average Z-score and root mean square Z-score. Results showed that there is a variation in value of average and RMS zscore with increase in resolution. It has been found that with increase in resolution average Z-score remains same but in case of RMS Z-score with increase in resolution Z-score increases.

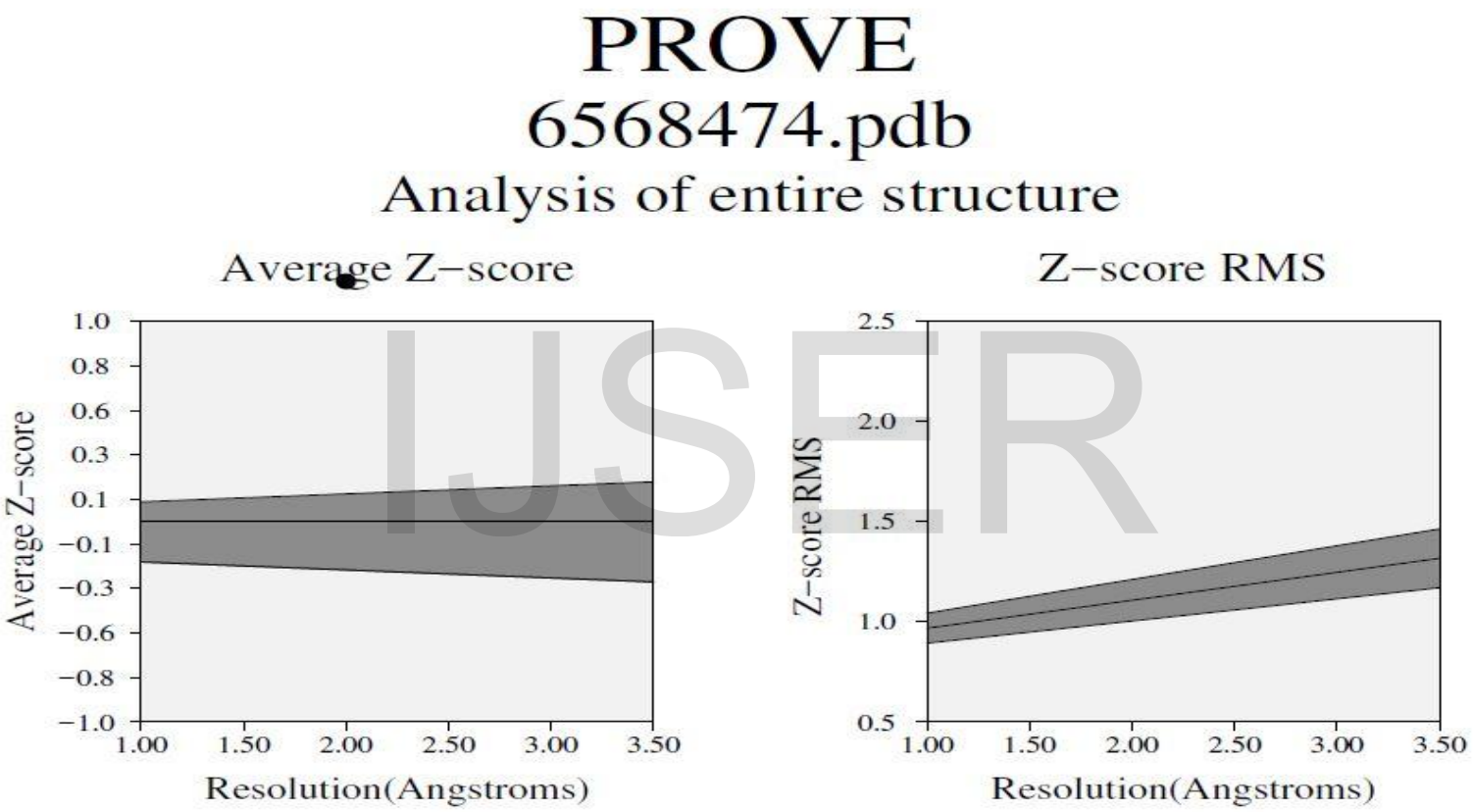
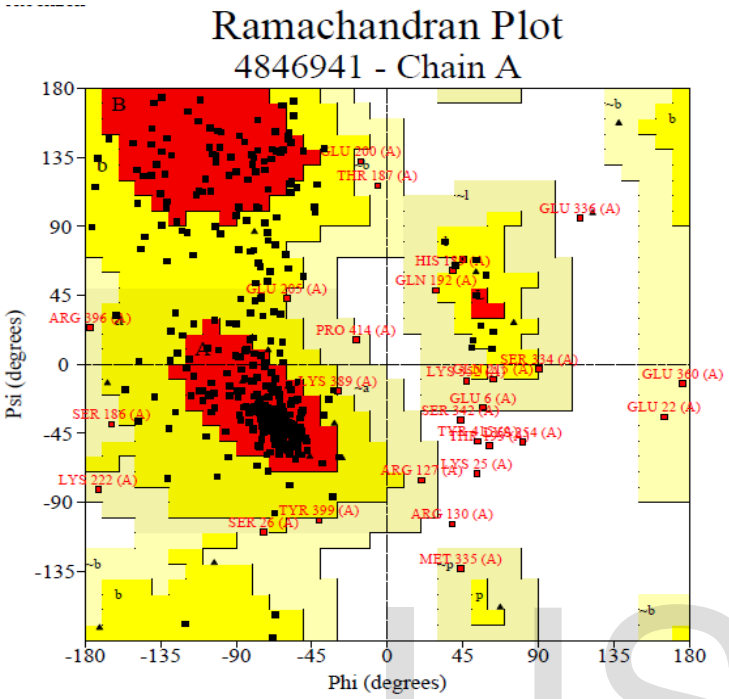



Fig 9: Average and RMS Z-score of our query protein

**amachandran plot:** Results of Ramachandran plot depicted that our protein contains 269 residues in most favored regions, 93 residues additional allowed regions, 19 residues in generously allowed regions, 7 residues in disallowed regions, 1 residue in end regions, 17 glycine residues and 13 proline residues.



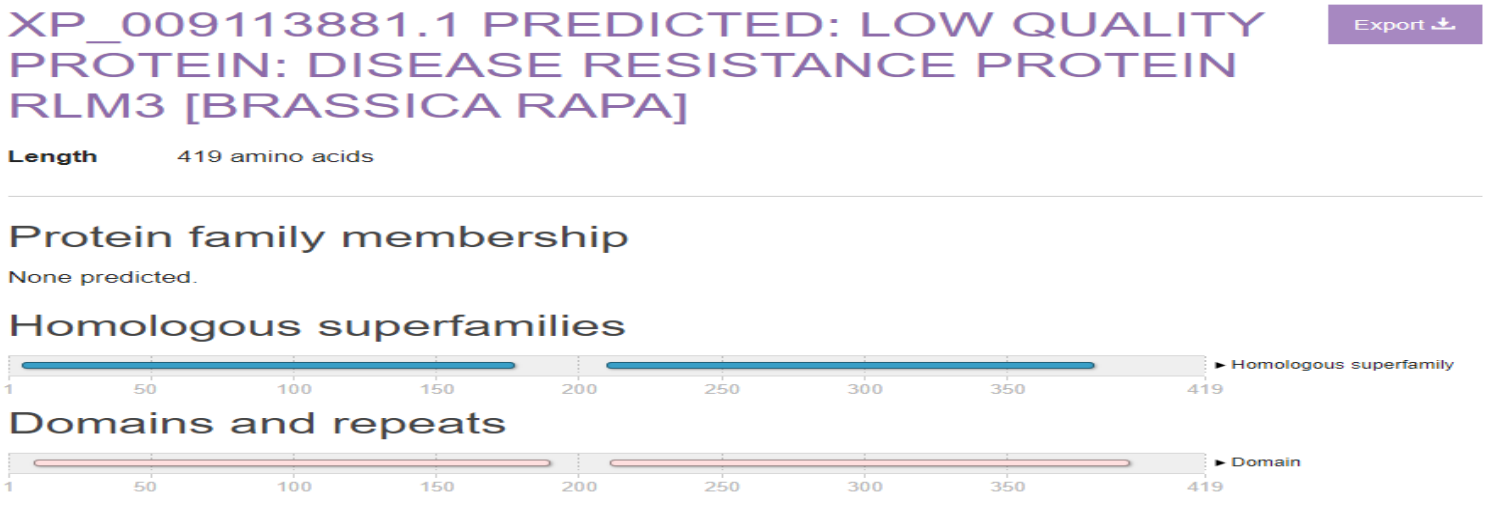

Plot statistics

| | | |
|---|---|---|
| Residues in most favoured regions [A,B,L] | 269 | 69.3% |
| Residues in additional allowed regions [a,b,l,p] | 93 | 24.0% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 19 | 4.9% |
| Residues in disallowed regions | 7 | 1.8% |
| Number of non-glycine and non-proline residues | 388 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 1 | |
| Number of glycine residues (shown as triangles) | 17 | |
| Number of proline residues | 13 | |
| Total number of residues | 419 | |

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig 10 : Ramachandaran plot and residues of our query protein

**otein Function Prediction** : The results of protein prediction through interpro demonstrated that our query protein that was predicted has the same properties as the properties of RLm3 protein in Arabidopsis Thaliana this shows that the properties of RLm3 in Arabidopsis thaliana and Brassica napus contain same properties.



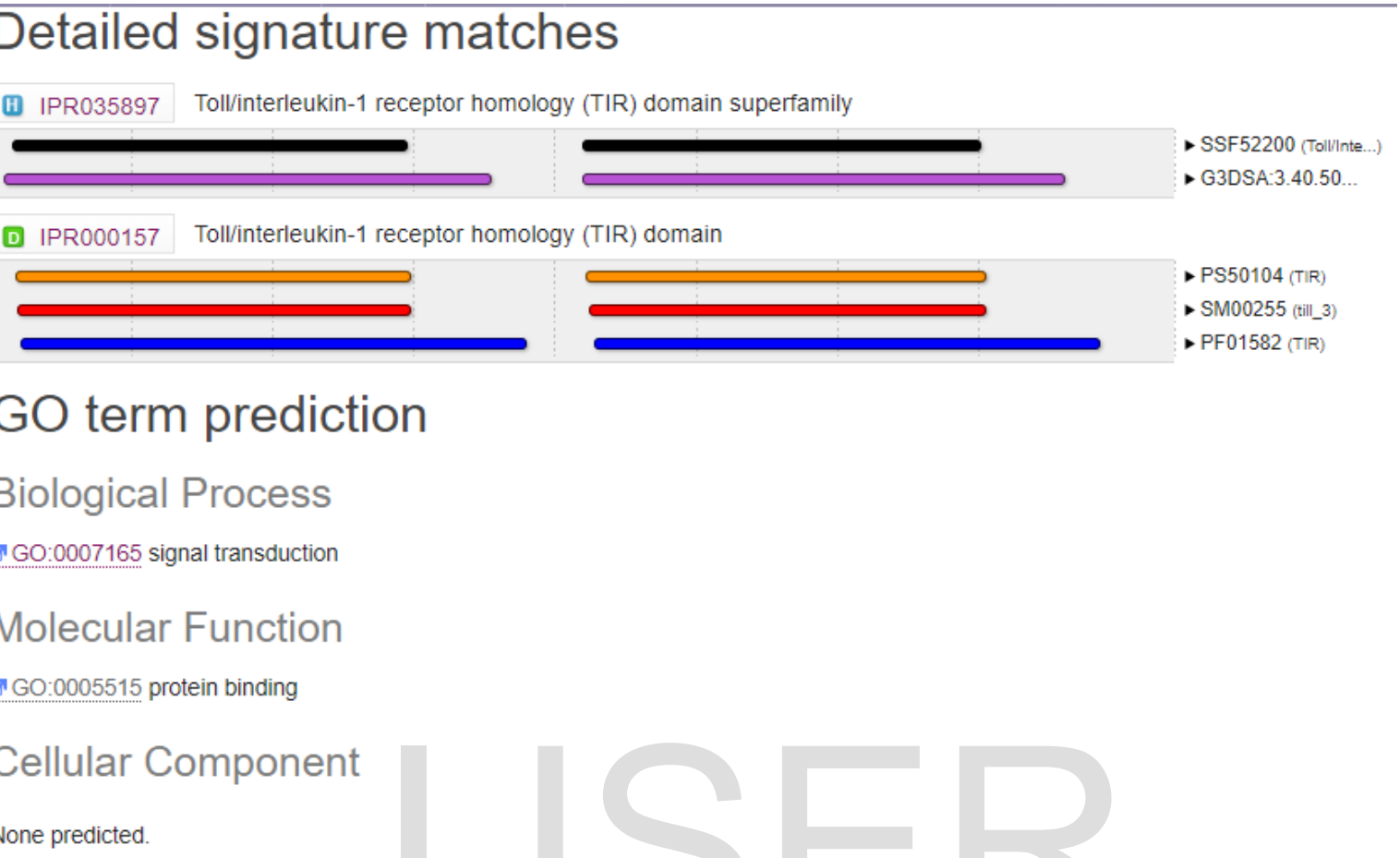





Fig 11: Function prediction showing that our query protein is a TIR protein

**b cellular Prediction:** The results of CELLO and Plant-mPloc both reveal that our query protein is present in nucleus of cell.

# CELLO RESULTS

SeqID: XP_009113881.1 PREDICTED: LOW QUALITY PROTEIN: disease resistance protein RLM3 [Brassica rapa]

Analysis Report:

| SVM | LOCALIZATION | RELIABILITY |
|---|---|---|
| Amino Acid Comp. | Nuclear | 0.419 |
| N-peptide Comp. | ER | 0.227 |
| Partitioned seq. Comp. | Nuclear | 0.669 |
| Physico-chemical Comp. | Cytoplasmic | 0.514 |
| Neighboring seq. Comp. | Cytoplasmic | 0.463 |

CELLO Prediction:

| | | |
|---|---|---|
| | Nuclear | 1.786 * |
| | Cytoplasmic | 1.689 * |
| | Mitochondrial | 0.301 |
| | ER | 0.285 |
| | Chloroplast | 0.266 |
| | PlasmaMembrane | 0.252 |
| | Golgi | 0.155 |
| | Extracellular | 0.132 |
| | Peroxisomal | 0.050 |
| | Cytoskeletal | 0.040 |
| | Lysosomal | 0.027 |
| | Vacuole | 0.017 |

**********

[ Home | Documentation]



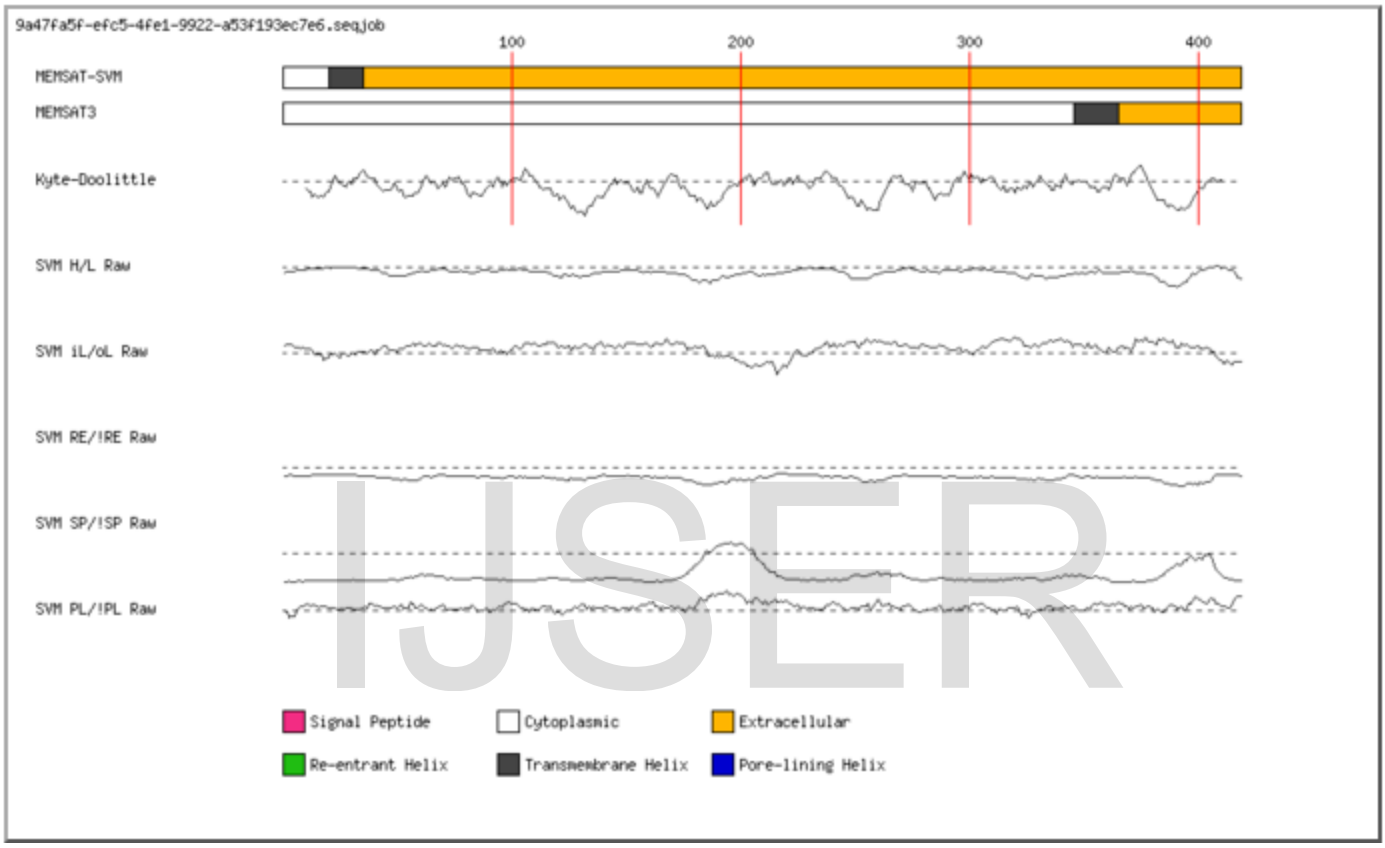
Fig11: sub cellular localization of our query protein

**Membrane Structure and Topology Prediction:** The results of membrane structure and topology prediction says that no signal peptide is present in in our query protein. No Re-entrant helices and No Pore-lining helices were present. One Trans membrane helix with score -0.129308 in region 20-35 was present.

# Sequence analysis results for job: memsat
# ID: 84a1d4de-77b5-11e8-97ee-00163e110593

Summary | MEMSAT-SVM | Downloads

## MEMSAT-SVM Schematic



## MEMSAT-SVM Prediction

**Summary of MEMSAT-SVM Topology Analysis**

| | |
|---|---|
| Signal peptide | **Not detected.** |
| Signal score | **0** |
| Topology | **20-35** |
| Re-entrant helices | **Not detected.** |
| Pore-lining helices | **Not detected.** |
| Helix count | **1** |
| N-terminal | **in** |
| Score | **-0.120768** |

## MEMSAT3 Prediction

**Summary of MEMSAT3 Topology Analysis**

| Number | Type | Direction | Score |
|---|---|---|---|
| 1 | helix | + | 14.316 |
| 1 | helix | - | 0.641 |

**MEMSAT3 Prediction**

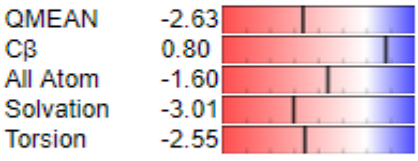
| Segment | Range | Score |
|---|---|---|
| 1 | (in) 346-365 | -2.06 |

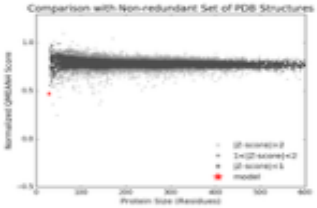
Fig 12 : Transmembrame helixes in our Query protein

**Homology Modelling:** The homologous model generated through SWISS MODEL was found to be 39.29% identical with METALLOTHIONEIN ISOFORM II with no ligands and QMEAN value of -2.63, GMQE value of 0.02, Cβ value of 0.80, Solvation value of -3.01 and torsion value of -2.55.



| Model #01 | File | Built with | Oligo-State | Ligands | GMQE | QMEAN |
|---|---|---|---|---|---|---|
|  | PDB | ProMod3 Version 1.1.0. | monomer (matching prediction) | None | 0.02 | -2.63 |



| Template | Seq Identity | Oligo-state | QSQE | Found by | Method | Resolution | Seq Similarity | Range | Coverage | Description |
|---|---|---|---|---|---|---|---|---|---|---|
| 4mt2.1.A | 39.29 | homo-dimer | | HHblits | X-ray | 2.00Å | 0.47 | 1406 - 1434 | 0.02 | METALLOTHIONEIN ISOFORM II |

Fig 13: Homologous model generated of our Query protein

**ultiple sequence alignment:** The results of multiple sequence alignment show the percentage identity matrix among 10 matching protein sequences with our query protein.

Percent Identity Matrix - created by Clustal2.1

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1: BAB10866.1 | 100.00 | 14.60 | 14.60 | 17.34 | 15.83 | 15.65 | 17.88 | 16.67 | 15.08 | 11.36 |
| 2: NP_199313.3 | 14.60 | 100.00 | 100.00 | 73.01 | 81.88 | 75.14 | 77.23 | 78.50 | 30.07 | 20.51 |
| 3: BAD95061.1 | 14.60 | 100.00 | 100.00 | 73.01 | 81.88 | 75.14 | 77.23 | 78.50 | 30.07 | 20.51 |
| 4: XP_009113881.1 | 17.34 | 73.01 | 73.01 | 100.00 | 96.25 | 75.00 | 80.90 | 81.73 | 30.62 | 21.43 |
| 5: XP_013610122.1 | 15.83 | 81.88 | 81.88 | 96.25 | 100.00 | 88.89 | 86.71 | 85.62 | 33.12 | 19.27 |
| 6: XP_019097648.1 | 15.65 | 75.14 | 75.14 | 75.00 | 88.89 | 100.00 | 86.67 | 86.50 | 29.37 | 18.89 |
| 7: XP_010441802.1 | 17.88 | 77.23 | 77.23 | 80.90 | 86.71 | 86.67 | 100.00 | 96.86 | 29.70 | 19.85 |
| 8: XP_019088165.1 | 16.67 | 78.50 | 78.50 | 81.73 | 85.62 | 86.50 | 96.86 | 100.00 | 31.00 | 19.26 |
| 9: XP_015159498.1 | 15.08 | 30.07 | 30.07 | 30.62 | 33.12 | 29.37 | 29.70 | 31.00 | 100.00 | 19.10 |
| 10: AED98370.1 | 11.36 | 20.51 | 20.51 | 21.43 | 19.27 | 18.89 | 19.85 | 19.26 | 19.10 | 100.00 |

Fig 14: Percentage identity matrix of all the proteins that match our query protein

**otein Protein Interaction:** The results of Protein Protein interaction show that our Query protein gets interacted with 10 other proteins with highest score of .632 with AT4G16970 and lowest score of .435 with AT1G58602.

**Nodes:**

Network nodes represent proteins

*splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.*

Node Color

*colored nodes: query proteins and first shell of interactors*

*white nodes: second shell of interactors*

Node Content

*empty nodes: proteins of unknown 3D structure*

*filled nodes: some 3D structure is known or predicted*

**Edges:**

Edges represent protein-protein associations

*associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.*

Known Interactions

*from curated databases*

*experimentally determined*

Predicted Interactions

*gene neighborhood*

*gene fusions*

*gene co-occurrence*

Others

*textmining*

*co-expression*

*protein homology*

**Your Input:**

| | | Neighborhood | Gene Fusion | Cooccurence | Coexpression | Experiments | Databases | Textmining | Homology | Score |
|---|---|---|---|---|---|---|---|---|---|---|
| RLM3 | *RESISTANCE TO LEPTOSPHAERIA MACULANS 3; TIR-NB-LRR receptor-like protein that confers resistance to the pathogens Leptosphaeria maculans (blackleg disease), Botrytis cinerea, Alternaria brassicicola and Alternaria brassicae. Required for efficient callose deposition downstream of RLM1 during infection with L.maculans (796 aa)* | | | | | | | | | |

**Predicted Functional Partners:**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| AT4G16970 | *Cell division control protein 7 (889 aa)* | | | | | | | • | | 0.632 |
| AT4G23440 | *Disease resistance protein (TIR-NBS class) (964 aa)* | | | | | | | • | | 0.606 |
| AT4G17000 | *Uncharacterized protein (674 aa)* | | | | | | | • | | 0.590 |
| AT1G64065 | *Late embryogenesis abundant hydroxyproline-rich glycoprotein (214 aa)* | | | | | | | • | | 0.579 |
| MAKR2 | *Uncharacterized protein (411 aa)* | | | | | | | • | | 0.556 |
| AT5G66900 | *CC-NBS-LRR class disease resistance protein; Probable disease resistance protein (809 aa)* | | | | • | | | • | | 0.541 |
| CRK22 | *Cysteine-rich receptor-like protein kinase 22 (660 aa)* | | | | • | | | • | | 0.531 |
| CRK40 | *Cysteine-rich receptor-like protein kinase 40 (654 aa)* | | | | • | | | • | | 0.526 |
| AT4G16980 | *Arabinogalactan family protein (164 aa)* | | | | • | | | • | | 0.522 |
| AT1G58602 | *LRR and NB-ARC domain-containing disease resistance protein; Potential disease resistance protein (1138 aa)* | | | | • | | | • | | 0.485 |

**Your Current Organism:**

Arabidopsis thaliana

*NCBI taxonomy Id: 3702*

*Other names: A. thaliana, Arabidopsis thaliana, Arabidopsis thaliana (L.) Heynh., mouse-ear cress, thale cress, thale-cress*

Fig 15 : Interaction among various proteins that are homologous to our query protein

**Phylogenetic Analysis:** The results of phylogenetic analysis show that our query protein shows most likelihood with Arabidopsis thaliana Disease resistance protein (TIR-NBS-LRR class) family mRNA

**(NM_123868.4)** and least likelihood with PREDICTED: Solanum tuberosum uncharacterized LOC102590743 (LOC102590743), mRNA **(XM_015304012.1).**

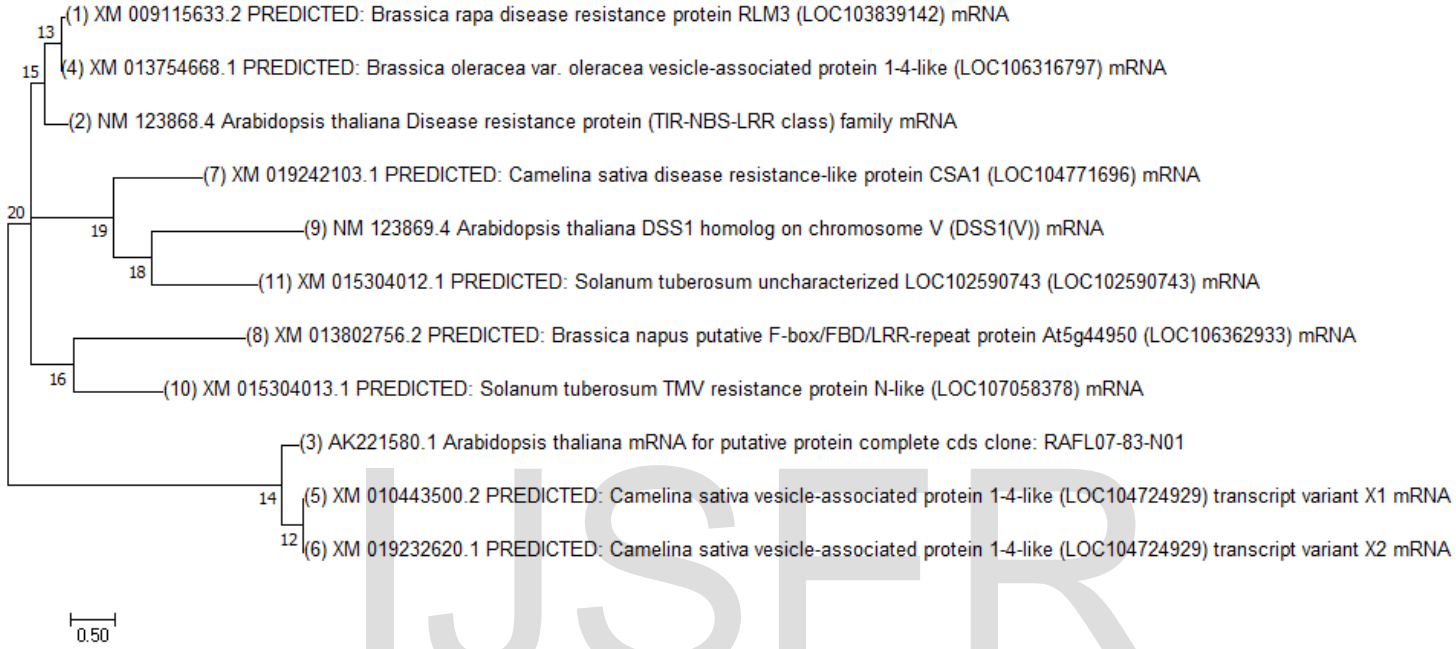





Fig: 16 Phylogenetic tree generated for our query protein

**EFERENCES:**


- Benkert, Pascal, Marco Biasini, and Torsten Schwede. "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics* 27.3 (2010): 343-350.
- Bertoni, Martino, et al. "Modeling protein quaternary structure of homo-and hetero-oligomers beyond binary interactions by homology." *Scientific reports* 7.1 (2017): 10480.
- Bienert, Stefan, et al. "The SWISS-MODEL Repository—new features and functionality." *Nucleic acids research* 45.D1 (2016): D313-D319.
- Bokor, A., Barbetti, M. J., Brown, A. G. P., MacNish, G., and Wood, P. M. (1975). Blackleg of rapeseed. Journal Department of Agriculture Western Australia 16,7-10.
- Booth, C., et al. "CMI descriptions of pathogenic fungi and bacteria." *CMI Descriptions of pathogenic fungi and bacteria*Set 33; Set 34 (1972)..
- Buchan, Daniel WA, et al. "Scalable web services for the PSIPRED Protein Analysis Workbench." *Nucleic acids research* 41.W1 (2013): W349-W357.

- Cai, Xiang, et al. "Effect of Water Flooding on Survival of Leptosphaeria biglobosa ‘brassicae’in Stubble of Oilseed Rape (Brassica napus) in Central China." *Plant Disease* 99.10 (2015): 1426-1433.
- Cuthbert, Patricia Anne. "Genetic analysis of blackleg and white rust resistance in Brassica rapa." (2000).
- Delourme, R., et al. "Major gene and polygenic resistance to Leptosphaeria maculans in oilseed rape (Brassica napus)." European Journal of Plant Pathology 114.1 (2006)
- Denby, Katherine J., Pavan Kumar, and Daniel J. Kliebenstein. "Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana." The Plant Journal 38.3 (2004)
- Dusabenyagasani, M., and Fernando, W. G. D. 2008. Development of a SCAR marker to track canola resistance against blackleg caused by Leptosphaeria maculans pathogenicity group 3. Plant Dis. 92:903-908.
- Elliott, Candace E., and Barbara J. Howlett. "Mutation of a gene in the fungus Leptosphaeria maculans allows increased frequency of penetration of stomatal apertures of Arabidopsis thaliana." Molecular plant 1.3 (2008)
- Elliott, Candace E., and Barbara J. Howlett. "Overexpression of a 3-ketoacyl-CoA thiolase in Leptosphaeria maculans causes reduced pathogenicity on Brassica napus." Molecular plant-microbe interactions 19.6 (2006)
- Elliott, Candace E., et al. "The cross-pathway control system regulates production of the secondary metabolite toxin, sirodesmin PL, in the ascomycete, Leptosphaeria maculans.(2011)
- Feng, Jie, et al. "The LmSNF1 gene is required for pathogenicity in the canola blackleg pathogen Leptosphaeria maculans." PLoS One 9.3 (2014)
- Fernando, W.G.D. 2010. Managing Blackleg Resistance Breakdown and Trade Barriers through Blackleg Resistance Stewardship in Canola. MB Agronomists Conference, December. University of Manitoba.
- Fernando, WG Dilantha, Yu Chen, and Kaveh Ghanbarnia. "Breeding for blackleg resistance: the biology and epidemiology." Advances in Botanical Research 45 (2007)
- Finn, Robert D., et al. "InterPro in 2017—beyond protein family and domain annotations." *Nucleic acids research* 45.D1 (2016): D190-D199.
- Finn, Robert D., et al. "Latest publications." *Nucleic Acids Research* (2017).
- Fitt, B.D.L. ,H. Brun, M.J. Barbetti and S.R. Rimmer. 2006. World-wide importance of phoma stem canker (Leptosphaeria maculans and L. biglobosa) on oilseed rape (Brassica napus). European Journal of Plant Pathology, 114: 3-15.
- Fitt, Bruce DL, et al. "World-wide importance of phoma stem canker (Leptosphaeria maculans and L. biglobosa) on oilseed rape (Brassica napus)." Sustainable strategies for managing Brassica napus (oilseed rape) resistance to Leptosphaeria maculans (phoma stem canker). Springer, Dordrecht, 2006
- Franceschini, Andrea, et al. "STRING v9. 1: protein-protein interaction networks, with increased coverage and integration." *Nucleic acids research* 41.D1 (2012): D808-D815.

- Franceschini, Andrea, et al. "SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles." *Bioinformatics* 32.7 (2015): 1085-1087.
- Gabrielson, R. L. "Blackleg disease of crucifers caused by Leptosphaeria maculans (Phoma lingam) and its control." Seed Science and Technology (1983).
- Gasteiger, Elisabeth, et al. "Protein identification and analysis tools on the ExPASy server." *The proteomics protocols handbook*. Humana press, 2005. 571-607.
- Guex, Nicolas, Manuel C. Peitsch, and Torsten Schwede. "Automated comparative protein structure modeling with SWISS-MODEL and Swiss- PdbViewer: A historical perspective." *Electrophoresis* 30.S1 (2009).
- Guo, X.W. and W.G.D. Fernando. 2005. Seasonal and diurnal patterns of spore dispersal by Leptosphaeria maculans from canola stubble in relation to environmental conditions. Plant Dis. 89:97-104.
- HAMMOND, KIM E., B. G. Lewis, and T. M. Musa. "A systemic pathway in the infection of oilseed rape plants by Leptosphaeria maculans." Plant Pathology 34.4 (1985)
- HAMMOND, KIM E., and B. G. Lewis. "Variation in stem infections caused by aggressive and non-aggressive isolates of Leptosphaeria maculans on Brassica napus var. oleifera." Plant Pathology 36.1 (1987)
- Howlett, Barbara J. "Current knowledge of the interaction between Brassica napus and Leptosphaeria maculans." Canadian Journal of Plant Pathology 26.3 (2004)
- Howlett, Barbara J., et al. "Evolution of virulence in fungal plant pathogens: exploiting fungal genomics to control plant disease." Mycologia 107.3 (2015)
- H. R. Kutcher, F. Yu & H. Brun (2010) Improving blackleg disease management of Brassica napus from knowledge of genetic interactions with Leptosphaeria maculans, Canadian Journal of Plant Pathology, 32:1, 29-34
- Hwang, Sheau-Fang et al. “Blackleg (Leptosphaeria maculans) Severity and Yield Loss in Canola in Alberta, Canada.” Plants (2016).
- Huang, Yong-Ju, et al. "Effects of fungicide on growth of Leptosphaeria maculans and L. biglobosa in relation to development of phoma stem canker on oilseed rape (Brassica napus)." Plant Pathology 60.4 (2011)
- Hwang, Sheau-Fang, et al. "Blackleg (Leptosphaeria maculans) severity and yield loss in canola in Alberta, Canada." Plants 5.3 (2016)
- Janitha P. D. Wanasundara (2011) Proteins of Brassicaceae Oilseeds and their Potential as a Plant Protein Source, Critical Reviews in Food Science and Nutrition, 51:7, 635-677
- Jensen, Lars J., et al. "STRING 8—a global view on proteins and their functional interactions in 630 organisms." *Nucleic acids research* 37.suppl_1 (2008): D412-D416.

- Jones, David T. "Improving the accuracy of transmembrane protein topology prediction using evolutionary information." *Bioinformatics* 23.5 (2007): 538-544.
- Jones, David T. "Protein secondary structure prediction based on position-specific scoring matrices1." *Journal of molecular biology* 292.2 (1999): 195-202.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. "A model recognition approach to the prediction of all-helical membrane protein structure and topology." *Biochemistry* 33.10 (1994): 3038-3049.
- Jones, Philip, et al. "InterProScan 5: genome-scale protein function classification." *Bioinformatics* 30.9 (2014): 1236-1240
- Khangura, R. K., and M. J. Barbetti. "Prevalence of blackleg (Leptosphaeria maculans) on canola (Brassica napus) in Western Australia." Australian Journal of Experimental Agriculture 41.1 (2001): 71-80.
- Khashnobish, Aruna, and Carol A. Shearer. "Reexamination of some Leptosphaeria and Phaeosphaeria species, Passeriniella obiones and Melanomma radicans." *Mycological Research* 11.100 (1996): 1341-1354.
- Koch, E., H. M. A. Badawy, and H. H. Hoppe. "Differences Between Aggressive and Non-Aggressive Single Spore Lines of Leptosphaeria maculans in Cultural Characteristics and Phytotoxin Production." *Journal of phytopathology* 124.1 (1989): 52-62.
- Kumar, Arvind, et al. "Canola cultivation in India: scenario and future strategy." *16th Australian research assembly on Brassicas. Ballarat, Victoria* (2009): 0-5.
- Kumar, Sudhir, et al. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." *Molecular biology and evolution* 35.6 (2018): 1547-1549.
- Kumar, Sudhir, Glen Stecher, and Koichiro Tamura. "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets." *Molecular biology and evolution* 33.7 (2016): 1870-1874.
- Kutcher, H. R., et al. "Best management practices for blackleg disease of canola." Prairie Soils & Crops Journal 4 (2011
- Kutcher, H. R., et al. "Blackleg disease of canola mitigated by resistant cultivars and four-year crop rotations in western Canada." Canadian Journal of Plant Pathology 35.2 (2013)
- Kutcher, H.R. and S.S. Malhi. 2010. "Residue burning and tillage effects on diseases and yield of barley (Hordeum vulgare) and canola (Brassica napus)." Soil & Tillage Research, 109 (3), pp. 153-160.
- Kutcher, H. R., F. Yu, and H. Brun. "Improving blackleg disease management of Brassica napus from knowledge of genetic interactions with Leptosphaeria maculans." *Canadian Journal of Plant Pathology* 32.1 (2010): 29-34.
- Kutcher H. R., Keri M., McLaren D. L. and Rimmer S. R.. 2007. Pathogenic variability of Leptosphaeria maculans in western Canada. Canadian J of Plant Pathology 29(4):388-393.

- Kutcher, H. R., M. H. Balesdent, S. R. Rimmer, T. Rouxel, A. M. Chèvre, R. Delourme, and H. Brun. 2010. 'Frequency of avirulence genes in Leptosphaeria maculans in Western Canada', Canadian Journal of Plant Pathology, 32: 1, 77 - 85.
- Kutcher, HR, WGD Fernando, TK Turkington and DL McLaren 2011. 'Best Management Practices for Blackleg Disease of Canola. Prairie Soils & Crops Journal. Volume 4.2011.
- Laskowski, R. A., M. W. MacArthur, and J. M. Thornton. "PROCHECK: validation of protein structure coordinates." *International tables of crystallography, volume F. Crystallography of biological macromolecules* (2001): 722-725
- Laskowski, Roman A., et al. "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." *Journal of biomolecular NMR* 8.4 (1996): 477-486.
- Laskowski, Roman A., et al. "PROCHECK: a program to check the stereochemical quality of protein structures." *Journal of applied crystallography* 26.2 (1993): 283-291.
- Li, Hua, et al. "Differences in the responses of stem tissues of spring-type Brassica napus cultivars with polygenic resistance and single dominant gene-based resistance to inoculation with Leptosphaeria maculans." Botany 85.2 (2007)
- Li, Hua, et al. "Germination and invasion by ascospores and pycnidiospores of Leptosphaeria maculans on spring-type Brassica napus canola varieties with varying susceptibility to blackleg." Journal of General Plant Pathology
- Li, Hua & Sivasithamparam, Krishnapillai & J. Barbetti, Martin & Kuo, John. (2004). Germination and invasion by ascospores and pycnidiospores of Leptosphaeria maculans on spring-type Brassica napus canola varieties with varying susceptibility to blackleg. Journal of General Plant Pathology.
- Li, Weizhong, et al. "The EMBL-EBI bioinformatics web and programmatic tools framework." *Nucleic acids research* 43.W1 (2015): W580-W584.
- Lüthy, Roland, James U. Bowie, and David Eisenberg. "Assessment of protein models with three-dimensional profiles." *Nature* 356.6364 (1992): 83.
- Marcroft, S. J., et al. "Factors affecting production of inoculum of the blackleg fungus (Leptosphaeria maculans) in south-eastern Australia." Australian Journal of Experimental Agriculture 43.10 (2003)
- McGee DC, Petrie GA, 1979. Seasonal patterns of ascosporedischarge by Leptosphaeria maculans in relation to blackleg of oilseed rape.
- McWilliam, Hamish, et al. "Analysis tool web services from the EMBL-EBI." *Nucleic acids research* 41.W1 (2013): W597-W600.

- Mering, Christian von, et al. "STRING: a database of predicted functional associations between proteins." *Nucleic acids research* 31.1 (2003): 258-261.
- Morris, Anne Louise, et al. "Stereochemical quality of protein structure coordinates." *Proteins: Structure, Function, and Bioinformatics* 12.4 (1992): 345-364.
- Murray, G. M., and Stovold, G. E. (1970). Blackleg: a potential major disease of rape. Biological and Chemical Research Institute News 18, 9-10.
- Nugent, Timothy, and David T. Jones. "Transmembrane protein topology prediction using support vector machines." *BMC bioinformatics* 10.1 (2009): 159.
- Payal, P., & Panchal, H. (2016). Primary, secondary and tertiary structural analysis of disease resistance protein RGA4 of ZEA maize using bioinformatics tools. International Journal of Advanced Research in Computer and Communication Engineering, 5(6), 7–12.
- Pontius, Joan, Jean Richelle, and Shoshana J. Wodak. "Deviations from standard atomic volumes as a quality measure for protein crystal structures." *Journal of molecular biology* 264.1 (1996): 121-136.
- Rakow, G. "Species origin and economic importance of Brassica." Brassica. Springer, Berlin, Heidelberg, 2004. 3-11.
- Remy, Estelle, et al. "The Lmpma1 gene of Leptosphaeria maculans encodes a plasma membrane H+-ATPase isoform essential for pathogenicity towards oilseed rape." Fungal Genetics and Biology 45.7 (2008)
- Rimmer, S. R. "Chasing genes for resistance to blackleg and sclerotinia in Brassica napus." *Proceedings 12th International Rapeseed Congress, Wuhan, China*. 2007.
- Rimmer, S. Roger. "Resistance genes to Leptosphaeria maculans in Brassica napus." Canadian Journal of Plant Pathology 28.S1 (2006)
- Robin, Arif Hasan Khan, et al. "Leptosphaeria maculans Alters Glucosinolate Profiles in Blackleg Disease–Resistant and-Susceptible Cabbage Lines." *Frontiers in plant science* 8 (2017): 1769
- Rouxel, Thierry, et al. "Biological effects of sirodesmin PL, a phytotoxin produced by Leptosphaeria maculans.(1988)
- Roy, Ambrish, Alper Kucukural, and Yang Zhang. "I-TASSER: a unified platform for automated protein structure and function prediction." *Nature protocols* 5.4 (2010): 725.
- Roy, N. N. "Interspecific transfer of Brassica juncea-type high blackleg resistance to Brassica napus." *Euphytica* 33.2 (1984): 295-303.
- Saal, B., and D. Struss. "RGA-and RAPD-derived SCAR markers for a Brassica B-genome introgression conferring resistance to blackleg in oilseed rape." Theoretical and applied genetics 111.2 (2005)

- Salisbury, P. A., et al. "Blackleg disease on oilseed Brassica in Australia: a review." *Australian Journal of Experimental Agriculture* 35.5 (1995): 665-672.
- Salisbury, Phillip & J. Ballinger, D & Wratten, N & Plummer, Kim & J. Howlett, B. (1995). Blackleg disease on oilseed Brassica in Australia: A review. Australian Journal of Experimental Agriculture. 35. 665-672. 10.1071/EA9950665. F
- Sievers, Fabian, et al. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." *Molecular systems biology* 7.1 (2011): 539.
- Sippl, Manfred J. "Recognition of errors in three-dimensional structures of proteins." *Proteins: Structure, Function, and Bioinformatics* 17.4 (1993): 355-362.
- Smith, E. G., et al. "Economics of Shorting Canola Rotations." *Agriculture and Agri-Food Canada. Poster* (2008).
- Snel, Berend, et al. "STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene." *Nucleic acids research* 28.18 (2000): 3442-3444.
- Sosnowski, M., E. Scott, and M. Ramsey. "INVESTIGATING RESISTANCE AND EPIDEMIOLOGY OF BLACKLEG (LEPTOSPHAERIA MACULANS) IN CANOLA."
- Sosnowski, Mark Roman. *Studies on the epidemiology of blackleg (Leptosphaeria maculans) and mechanisms of resistance in canola/Mark R. Sosnowski*. Diss. 2002
- Sprague, S. J., et al. "Major gene resistance to blackleg in Brassica napus overcome within three years of commercial production in southeastern Australia." *Plant Disease* 90.2 (2006): 190-198.
- Staal, Jens, and Christina Dixelius. "RLM3, a potential adaptor between specific TIR-NB-LRR receptors and DZC proteins." *Communicative & integrative biology* 1.1 (2008): 59-61.
- Staal, Jens. Genes and mechanisms in Arabidopsis innate immunity against Leptosphaeria maculans. Vol. 2006. No. 69. 2006.
- Staal, Jens, et al. "Transgressive segregation reveals two Arabidopsis TIR-NB-LRR resistance genes effective against Leptosphaeria maculans, causal agent of blackleg disease." The Plant Journal 46.2 (2006)
- Szklarczyk, Damian, et al. "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored." *Nucleic acids research* 39.suppl_1 (2010): D561-D568.
- Szklarczyk, Damian, et al. "STRING v10: protein–protein interaction networks, integrated over the tree of life." *Nucleic acids research* 43.D1 (2014): D447-D452.
- Szklarczyk, Damian, et al. "The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible." *Nucleic acids research* (2016): gkw937
- Tsuda, Kenichi, et al. "Network properties of robust immunity in plants." PLoS genetics 5.12 (2009

- Von Mering, Christian, et al. "STRING: known and predicted protein–protein associations, integrated and transferred across organisms." *Nucleic acids research* 33.suppl_1 (2005): D433-D437.
- Von Mering, Christian, et al. "STRING 7—recent developments in the integration and prediction of protein interactions." *Nucleic acids research* 35.suppl_1 (2006): D358-D362.
- Van de Wouw, Anton Cozijnsen, Jo Rayner and Barbara Howlett. 2009. Monitoring of virulence in Australian populations of the blackleg fungus. School of Botany, University of Melbourne.
- Warwick, S. I., et al. "Genetic variation of Ethiopian mustard (Brassica carinata A. Braun) germplasm in western Canada." *Genetic Resources and Crop Evolution* 53.2 (2006): 297-312.
- Waterhouse, Andrew, et al. "SWISS-MODEL: homology modelling of protein structures and complexes." *Nucleic acids research* (2018).
- West JS, Biddulph JE, Fitt BDL, Gladders P, 1999. Epidemiology of Leptosphaeria maculans in relation to forecasting stem canker severity on winter oilseed rape in the UK. Annals of Applied Biology 135, 535-46.
- West, J. S., et al. "Leptosphaeria maculans causing stem canker of oilseed rape in China." *Plant Pathology* 49.6 (2000): 800-800.
- WestJS, Evansn. "Leptosphaeria maculans causing stem canker of oil seed rape in China." *PlantPathology* 49.6 (2000): 800
- West JS, Kharbanda PD, Barbetti MJ, Fitt BDL, 2001. Epidemiology and management of Leptosphaeria maculans (phoma stem canker) on oilseed rape in Australia, Canada and Europe.
- Wherrett, A. D., K. Sivasithamparam, and M. J. Barbetti. "Establishing the relationship of ascospore loads with blackleg (Leptosphaeria maculans) severity on canola (Brassica napus)." Australian Journal of Agricultural Research 55.8 (2004)
- Williams RH, Fitt BDL, 1999. Differentiating A and B groups of Leptosphaeria maculans, causal agent of stem canker of winter oilseed rape in the UK. Plant Pathology 46, 161-75.
- Yang, Jianyi, et al."The I-TASSER Suite: protein structure and function prediction."*Nature methods* 12.1 (2015): 7.
- Yu, Chin-Sheng, Chih- Jen Lin, and Jenn- Kang Hwang. ″ Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n- peptide compositions.″ *Protein science* 13.5 (2004): 1402-1406.
- Yu, Chin-Sheng, et al. ″Prediction of protein subcellular localization.″ *Proteins: Structure, Function, and Bioinformatics*64.3 (2006): 643-651
- Zhang Xuehua, Fernando W. G. Dilantha (2017) Insights into fighting against blackleg disease of Brassica napus in Canada. Crop and Pasture Science 69, 40-47

- Zhang, Xuehua. "Exploring disease resistance in the Brassica napus-Leptosphaeria maculans pathosystem." (2016).
- Zhang, Xuehua, and WG Dilantha Fernando. "Insights into fighting against blackleg disease of Brassica napus in Canada." *Crop and Pasture Science* 69.1 (2018): 40-47.
- Zhang, Xu, et al. "Leptosphaeria spp., phoma stem canker and potential spread of L. maculans on oilseed rape crops in China." *Plant Pathology* 63.3 (2014): 598-612

**uthor Details:**

**Corresponding author:**

**Sumit Joshi : G.B. Pant Institute of Engineering and Technology, Pauri , Uttarakhand**

**Joshisumit14@gmail.com**

**Shivani Banchariya[2] : National Institute of Immunology, Aruna Asaf Ali Marg New Delhi-110067**

**shivani.banchariya@gmail.com**