# Protein Annotation & Advanced Insilico Based Mutation Hotspot Identification in Temperature Sensitive CDPK3 Protein of *Oryza sativa*

*[1]G. Shiva Prasad, [2]Dr. L.V. Subba Rao, [1]Dr. K.V. Radha Krishna, [2]U.Chaitanya and [3]Jyothsna Gundlapally

*1.Department of Genetics and Plant Breeding, College of Agriculture, ANGR Agricultural University, Rajendranagar, Hyderabad-500030, Andhra Pradesh, India*
*2. Directorate of Rice Research, Rajendranagar, Hyderabad-500030, Andhra Pradesh, India.*
*3. BioAxis DNA Research centre Pvt Ltd, Hyderabad*

***Abstract:*** *In the process of development of cold and heat tolerant rice varieties was done in field conditions. The current work is the extension of the same. In the current paper an extensive Insilico based research was carried out on the Mutational analysis and hot spot prediction of the protein CDPK3, Calcium Dependent Protein Kinases of Oriza Sativa. The work involves the retrieval of the protein sequence from the NCBI primary data base and identification of the Functional domains. Several sequences of the same protein from the other related plant species were collected and the phylogenetic study and the conservation prediction were performed. Tools like Protparam and SOPMA have been applied to analyze the Physico chemical parameters and the Structure prediction of the protein was carried out. The regions of disorder present in the protein sequence were identified using several tools including DISEMBL, GLOBPLOT, RONN etc. Based upon the results of RONN the major sites prone for mutations are identified. The effect of possible substitution mutations on the selected target site was analyzed. I mutant tool was employed to check the effect of all the substitution mutations on the stability of the protein sequence, PolyPhen was used to evaluate the effect of mutations on Functionality and the tolerance level is calculated using SIFT. The work concludes to provide the complete protein annotation of Cold tolerance protein CDPK with a focus on the mutation hot spot prediction.*

***Keywords:*** *Protein Annotation, Insilico Characterization, Mutation hotspots, CDPK, Cold tolerance*

## I. Introduction:

Rice being the staple crop of Middle and south India extensive research on the development of rice cultivation is a topic of demand. The current environment of extreme heat or extreme cold is causing a lot of stress on all the plants and crops cultivate. Though plants have their inherent mechanism to tolerate the adverse climatic changes sometimes due to the mutations and damages caused to the genetic makeup of the plats they lack this property of stress tolerance finally resulting in a huge damage to the crop cultivation. Calcium-dependent protein kinases (CDPKs) play an important role in rice signal transduction, but the precise role of each individual CDPK is still largely unknown [1]. A study revealed that OsCDPK13 gene expression and protein accumulation were enhanced in response to cold, but suppressed under salt and drought stress which depict the importance of this protein in the cold tolerance of the plant. Several works also inferred that the Over-expression of a single Ca 2+ -dependent protein kinase CDPK confers both cold and salt/drought tolerance on rice plants [2]. One of the research papers also presented the use of CDPK protein insertions and over expressions in Sorghum in order to impart tolerance to cold and freezing so as to expand the acreage for production [3]. CDPK protein super family consists of six types of protein kinases that differ in the regulatory domains they contain [4].

In view of the importance of the CDPK protein in the crop development and induction of resistant to the plants the current work aims to characterize and annotate the protein using insilico tools and software. The work aims to identify the mutation prone regions present in the CDPK 3 protein sequence of Oryza Sativa.

## II. Materials and Methods:

### I. Sequence Retrieval:

The major step in any of the Protein annotation works is to retrieve the Sequence of the protein of interest. The protein sequence is generally taken from the most reliable public data base, the NCBI [5] data base. The Boolean Operator search has been used for specifying the organism and the protein whose sequence is to be retrieved.

### II. BLAST Analysis for collection of the related protein sequence:

In order to collect the sequences of the proteins that share sequence similarity to the query protein BLSAT [6] tool has been applied. The BLAST is a local alignment search tool that would show the sequences that share some percentage of similarity. It is available at NCBI.

### III. Multiple Sequence Alignment and Phylogenetic Study:

All the sequences collected in the above step are subjected to alignment so as to analyze the sequence conservation and phylogenetic relation present in the collected sequences. CLUSTALW [7] is used to perform the multiple sequence alignment and design the dendrogram showing the evolutionary relation among the sequences. The CLUSTALW tool available at SDSC Biology Workbench.

### IV. Conservation Analysis using BOXSHDE:

The similarity among the collected sequences along with the percentage of conservation has been studied using BOXSHDE [8] tool by performing Global alignment of multiple sequences. The BOXSHADE tool is also available at SDSC biology Workbench.

### V. Functional Analysis for the identification of Domains using SMART:

To identify the total number of domains (Independent functional units) present in the sequence of CDPK and to identify the positions of these domains on the sequence SMART [9] has been used. The tool will also give the detailed annotation of the domain and other important elements present on the input sequence.

### VI. Analysis of the Physico chemical properties of the protein:

To study the inherent properties of the protein that would be necessary for designing the experimental protocol PROTPARAM [10] was used. The tool is available at EXPASY SERVER. The Protparam tool would give an insight to the chemical and physiological properties of the protein

### VII. Secondary and Tertiary structure prediction using SOPMA and CPH tools:

To detect the secondary structural confirmations within the protein SOPMA [11] can be used. It would assign the conformational pattern for each residue separately and also summarize to show the total percentage of different confirmations present in the sequence. This data can be used to predict the regions present on the surface and those in the interior of the protein. To get the three dimensional structure of the protein one has to use RCSB Protein Data BANK [12]. However each protein would have a unique identity in PDB which must be known to retrieve the structure of the protein. CPH tool has been employed to get the PDB ID of the study protein. Once the ID is obtained the structure can be downloaded from PDB and visualized using the RASMOL [13] software.

### VIII. Prediction of the probable disorder regions present in the query sequence:

To identify the regions of disorder present in the user entered sequence several online tools are available each working on a different algorithm and principle. GLOB PLOT [14], DISEMBL [15] and RONN [16] were employed to predict the regions of possible disorder. The results of RONN would provide the additional information about the disorder probability for each amino acid residue displayed separately. These results can be summarized to identify the sites on the sequence that have maximum disorder probability.

### IX. Analysis of the effect of Substitution mutations on Stability and Functionality of the protein:

Once the mutational hot spots are identified the next step is to check for the effect of all possible substitution mutations on the stability of the protein structure. Also it is necessary to analyze the effect of the same on the functionality and disease occurrence.

The I Mutant [17] tool is used for the above purpose. To evaluate the effect of substitution of the functional ability and disease occurrence PolyPhen [18] was used. SIFT was used to evaluate the tolerance of the protein to the possible substitutions.

### X. Analysis of the Tolerability of the mutations on the protein structure:

To check the tolerance range of the protein towards the substitutions specified SIFT [19] has been used. The tool enables to check the effect of substitutions on the proteins acceptability range.

## III. Results and Discussion:

### I. Sequence Retrieval of CDPK3 from NCBI:

The protein sequence of the query protein CDPK3 from *Oryza Sativa* has been retrieved from NCBI using Boolean operators to make the search specified. The length of the sequence was found to be 527 amino acids.

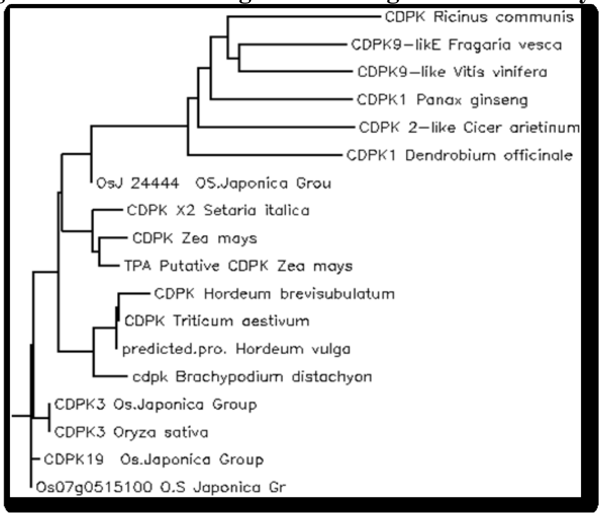
### II.BLAST for the collection of similar sequences:

Using the BLAST tool a total of 17 sequences of different organism were collected and used for further analysis. All the sequences belong to the common CDPK protein.

### III.Phylogenetic analysis using CLUSTALW Dendrogram:

To check the sequence similarity among the selected sequences MSA was performed using CLUSTALW and the phylogenetic tree was constructed.

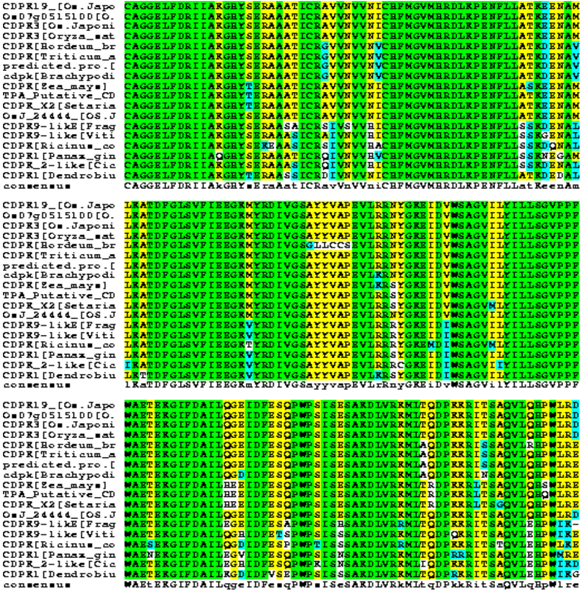
**Fig1: Phylogenetic Tree / Dendrogram showing the evolutionary relationship:**


CDPK Ricinus communis
CDPK9-likE Fragaria vesca
CDPK9-like Vitis vinifera
CDPK1 Panax ginseng
CDPK 2-like Cicer arietinum
CDPK1 Dendrobium officinale
OsJ 24444 OS.Japonica Grou
CDPK X2 Setaria italica
CDPK Zea mays
TPA Putative CDPK Zea mays
CDPK Hordeum brevisubulatum
CDPK Triticum aestivum
predicted.pro. Hordeum vulga
cdpk Brachypodium distachyon
CDPK3 Os.Japonica Group
CDPK3 Oryza sativa
CDPK19 Os.Japonica Group
Os07g0515100 O.S Japonica Gr

The above dendrogram shows the evolutionary relation between the selected 17 organisms with the query. It can be inferred that all the selected organisms are evolving from the same common branch which indicates that the sequences are homologues.

### IV. Conservation study based on Multiple sequence alignment:

**Fig 2: MSA using BOXSHDE for conservation study**

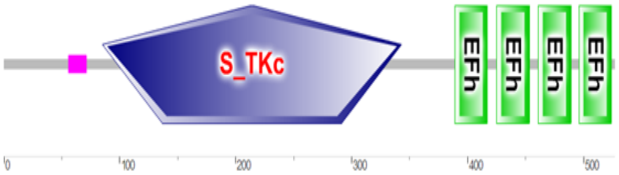


```
CDPK19_[Os.Japo CAGGELFDRIIAKGHYSERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
Os07g0515100[O. CAGGELFDRIIAKGHYSERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
CDPK3[Os.Japoni CAGGELFDRIIAKGHYSERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
CDPK3[Oryza_sat CAGGELFDRIIAKGHYSERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
CDPK[Hordeum_br CAGGELFDRIIAKGHYSERAAATICRGVVNVVNVCHFMGVMHRDLKPENFLLATKDENAV
CDPK[Triticum_a CAGGELFDRIIAKGHYSERAAATICRGVVNVVNVCHFMGVMHRDLKPENFLLATKDENAV
predicted.pro.[ CAGGELFDRIIAKGHYSERAAATICRGVVNVVNVCHFMGVMHRDLKPENFLLATKDENAV
cdpk[Brachypodi CAGGELFDRIIAKGHYSERAAATICRGVVNVVNVCHFMGVMHRDLKPENFLLATKDENAV
CDPK[Zea_mays]  CAGGELFDRIIAKGHYTERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLASKEENAM
TPA_Putative_CD CAGGELFDRIIAKGHYTERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
CDPK_X2[Setaria CAGGELFDRIIAKGHYTERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
OsJ_24444_[OS.J CAGGELFDRIIAKGHYSERAAATICRAVVNVVNICHFMGVMHRDLKPENFLLATKEENAM
CDPK9-likE[Frag CAGGELFDRIIAKGHYSERAASAICRSIVSVVHICHFMGVMHRDLKPENFLLSSKDENAL
CDPK9-like[Viti CAGGELFDRIIAKGHYSERAAASICRAIVNVVHICHFMGVMHRDLKPENFLLSSKGENAL
CDPK[Ricinus_co CAGGELFDRIIAKGHYSEKEAASICRSIVNVVHACHFMGVMHRDLKPENFLLSSKDQNAL
CDPK1[Panax_gin CAGGELFDRIIAQGHYSERAAATICRQIVNVVHVCHFMGVMHRDLKPENFLLSSKNEGAM
CDPK_2-like[Cic CAGGELFDRIIAKGHYSERAAASICRQIVNVVHICHFMGVMHRDLKPENFLLSSKDENAL
CDPK1[Dendrobiu CAGGELFDRIIAKGHYTERAASSICRDIVNVVNICHFMGVMHRDLKPENFLLATKDEDAL
consensus       CAGGELFDRIIAkGHYsEraAatICRavVnVVniCHFMGVMHRDLKPENFLLatKeenAm

CDPK19_[Os.Japo LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
Os07g0515100[O. LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
CDPK3[Os.Japoni LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
CDPK3[Oryza_sat LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
CDPK[Hordeum_br LKATDFGLSVFIEEGKMYRDIVGSGLLCCSEVLRRNYGKEIDVWSAGVILYILLSGVPPF
CDPK[Triticum_a LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
predicted.pro.[ LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
cdpk[Brachypodi LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLKRNYGKEIDVWSAGVILYILLSGVPPF
CDPK[Zea_mays]  LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLKRSYGKEIDVWSAGVILYILLSGVPPF
TPA_Putative_CD LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRSYGKEIDVWSAGVILYILLSGVPPF
CDPK_X2[Setaria LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRSYGKEIDVWSAGVMLYILLSGVPPF
OsJ_24444_[OS.J LKATDFGLSVFIEEGKMYRDIVGSAYYVAPEVLRRNYGKEIDVWSAGVILYILLSGVPPF
CDPK9-likE[Frag LKATDFGLSVFIEEGKVYRDIVGSAYYVAPEVLRRRYGKEIDIWSAGVILYILLSGVPPF
CDPK9-like[Viti LKATDFGLSVFIEEGKVYRDIVGSAYYVAPEVLRRRYGKEIDIWSAGVILYILLSGVPPF
CDPK[Ricinus_co LKATDFGLSVFIEEGKTYRDIVGSAYYVAPEVLRRRYGKEMDIWSAGVMLYILLSGVPPF
CDPK1[Panax_gin LKATDFGLSVFIEEGKVYRDIVGSAYYVAPEVLRRSYGKEIDIWSAGVILYILLSGVPPF
CDPK_2-like[Cic IKATDFGLSVFIEEGKVYRDIVGSAYYVAPEVLRRRCGKEIDIWSAGVIIYILLSGVPPF
CDPK1[Dendrobiu LKTTDFGLSVFIEEGKVYRDIVGSAYYVAPEVLKRSYGKEIDIWSAGVILYILLSGVPPF
consensus       lKaTDFGLSVFIEEGKmYRDIVGSayyvapEVLrRnyGKEiDvWSAGVilYILLSGVPPF

CDPK19_[Os.Japo WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLTQDPKKRITSAQVLQHPWLRD
Os07g0515100[O. WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLTQDPKKRITSAQVLQHPWLRD
CDPK3[Os.Japoni WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLTQDPKKRITSAQVLQHPWLRD
CDPK3[Oryza_sat WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLTQDPKKRITSAQVLQHPWLRD
CDPK[Hordeum_br WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLAQDPKKRISSAQVLRHPWLRE
CDPK[Triticum_a WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLAQDPKKRISSAQVLQHPWLRE
predicted.pro.[ WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLAQDPKKRISSAQVLQHPWLRE
cdpk[Brachypodi WAETEKGIFDAILQGDIDFESQPWPSISESAKDLVRKMLAQDPKKRINSAQVLQHPWLRE
CDPK[Zea_mays]  WAETEKGIFDAILHEEIDFESQPWPSISESAKDLVRKMLTRDPKKRLTSAQVLQHSWLRE
TPA_Putative_CD WAETEKGIFDAILHEEIDFESQPWPSISESAKDLVRKMLTRDPKKRLTSAQVLQHQWLRE
CDPK_X2[Setaria WAETEKGIFDAILHEEIDFESQPWPSISESAKDLVRKMLTRDPKKRLTSGQVLQHPWLRE
OsJ_24444_[OS.J WAETEKGIFDAILQGEIDFESQPWPSISESAKDLVRKMLTQDPKKRITSAQVLQHPWLRD
CDPK9-likE[Frag WAETEKGIFDAILEGEIDFESAPWPSISHSAKDLVRRMLTQDPKKRITSAQVLEHPWIR-
CDPK9-like[Viti WAETEKGIFDAILQGHIDFETSPWPSISSSAKDLVRKMLTQDPQKRITSAQVLEHPWIRE
CDPK[Ricinus_co WAESEKGIFDAILQGDIDFESPPWPTISSSAKDLVRRMLTQDPKKRITSTQVLEHPWLRE
CDPK1[Panax_gin WAENEKGIFDAILEGVIDFESQPWPSISNSAKDLVRKMLTQDPRRRITSAQVLEHPWMRE
CDPK_2-like[Cic WAETEKGIFDAILEGHIDFESQPWPRISNSAKDLVRKMLIQDPKKRITSAQVLEHPWIRD
CDPK1[Dendrobiu WAETEKGIFDAILRGDIDFVSEPWPSISSSAKDLVRKMLTQDPRKRITSAQVLEHPWIRG
consensus       WAEtEKGIFDAILqgeIDFesqPWPsISeSAKDLVRkMLtqDPkkRitSaQVLqHpWlre
```

Form the above fig 2 it can be inferred that the Sequences aligned in the given MSA show maximum conservation and are evolutionarily closely related.

**Domain Identification using SMART:**

**Fig 3: The Picture showing total 5 domains one is the S_TKc domain and the other four are EFh domains.**



From the above picture it can be inferred that there are a total number of 5 domains in the CDPK sequence. One of it would be the S_TKc domain and the other four being the EFh domains. The S_TKc domain ranges from 85 to 343 in the query sequence.

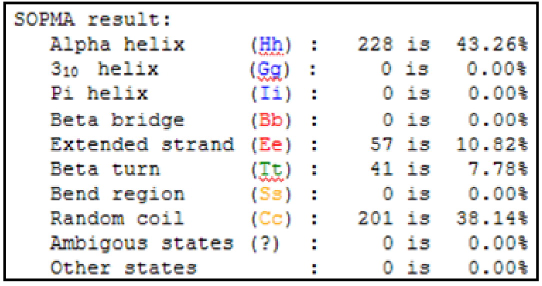
The EFh domains are present in the regions ranging from 389-417, 425-453, 461-489 and 496-524. These are the regions that are involved in the major functions of the protein. The function of the S_TKc domain is the Serine/Threonine protein kinases, It is the catalytic domain. The other domain EFh is involved in the binding of the protein to the calcium ion cofactors and thus is important in the normal functioning of the protein.

**Protparam for Physico chemical properties of the protein:**

To study the physic chemical properties of the protein PROTPARAM tool has been used which indicates that the protein in 527 amino acids in length with the isoelectric pH being 5.54 and the molecular weight is 58872.0. The protein is unstable with the instability index being 41.4. The hydropathicity of the protein was found to be -0.398 which indicates that the protein is polar and water soluble.

**Secondary structure Prediction using SOPMA:**

**Fig 4: Showing the secondary structural confirmations of the sequence as shown by SOPMA**



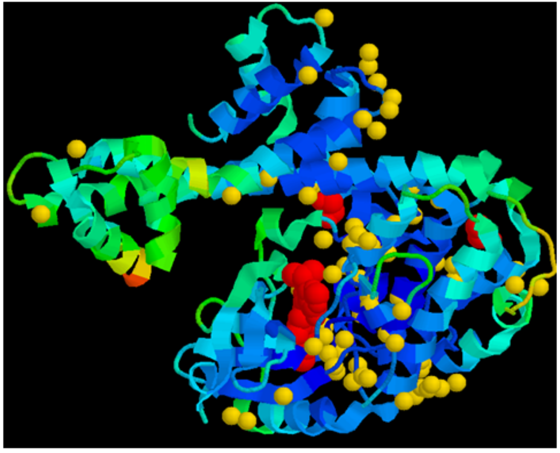
```
SOPMA result:
   Alpha helix      (Hh) :   228 is  43.26%
   310 helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    57 is  10.82%
   Beta turn        (Tt) :    41 is   7.78%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :   201 is  38.14%
   Ambigous states  (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

From the above figure 4 the total number and types of confirmations present in the given sequence have been identified. It can be inferred that the sequence contains 43.26% of alpha helical conformations, 38.14%random coil, 10.82% extended strand and 7.78% beta turn confirmations.

**Tertiary Structure Prediction CPH and Visualization using RASMOL:**

To get the 3D structure of the protein CPH tool has been employed and the ID 3SXF has been selected as the best id for the structure of the CDPK protein. The structure has been retrieved from the PDB data base and the visualization has been done in RASMOL software.

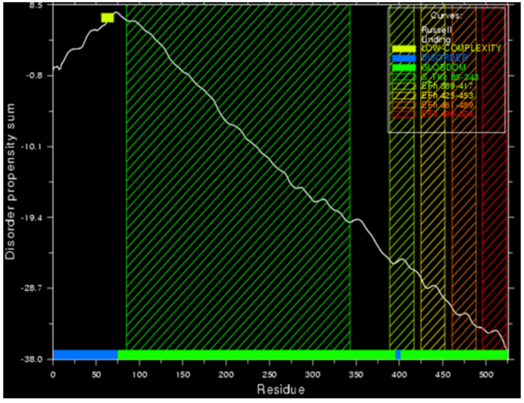
**Fig 5: The picture showing the 3D structure of CDPK as visualized by RASMOL software:**



**Disorder site identification using Insilico tools:**

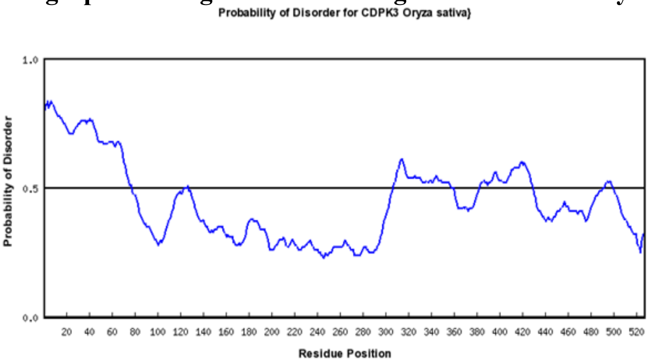
To identify the disorder regions present in the given sequence GLOBPLOT, DISEMBL and RONN have been used and the common regions are selected.

**Fig 6: The graph showing the results of GLOBPLOT**



The above fig 6 shows the disorder regions as indicated by the GLOBPLOT tool. From the results it can be explained that there are two regions of disorder in the given sequence which include: 1-74 and 396-401. To select the best site among the identified regions RONN can be used.

**Fig 7: The graph showing the Disorder regions as indicated by RONN.**



The above figure 7 shows the disorder regions as predicted by the RONN analysis. As per the results of RONN the disordered regions are 1 - 77, 127 - 127, 308 - 357, 383 - 429, and 493 – 499.

Based on the common mutation sites and the highest mutation probability the sites selected are 1-11 amino acids.

**Table 1: The table showing the summarized results of disorder prediction along with the evaluation of the hydropathicity index for each residue.**

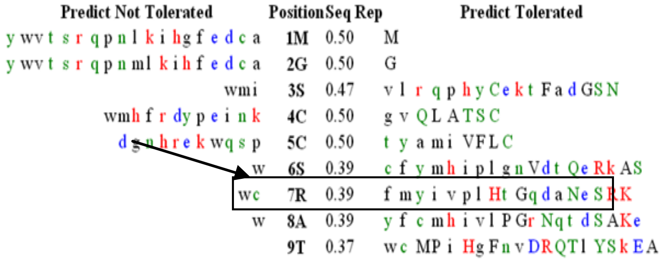
| Amino Acid | Position | Disorder probability | Hydropathicity Index / Polarity |
|---|---|---|---|
| M | 1 | 0.78 | 1.9 |
| G | 2 | 0.82 | -0.4 |
| S | 3 | 0.82 | -0.8 |
| C | 4 | 0.84 | 2.5 |
| C | 5 | 0.81 | 2.5 |
| S | 6 | 0.83 | -0.8 |
| R | 7 | 0.84 | -4.5 |
| A | 8 | 0.83 | 1.8 |
| T | 9 | 0.82 | -0.7 |
| S | 10 | 0.81 | -0.8 |
| P | 11 | 0.80 | -1.6 |

From the above table it can be explained that from among the amino acids 1-11 the amino acids R7 and P11 are supposed to be exposed on the surface due to their less hydropathicity index. Thus the amino acid R7 is identified as the mutation hotspot for the CDPK3 sequence from Oryza Sativa.

**Summarization of the effect of substitution mutations to the R7 position on the Stability of the structure:**

In the position 7R there are a total of 19 substitution mutations that can be possible and thus the analysis was carried out to detect the effect of these 19 mutations on the Stability of the protein using I Mutant, The results are summarized in the Table 2.

**Fig 8: The results of SIFT showing the Tolerance level for the mutations at R7.**



```
          Predict Not Tolerated        PositionSeq Rep          Predict Tolerated
ywvt s r q p n l k i hg f e d c a    1M   0.50    M
ywvt s r q p n ml k i h f e d c a    2G   0.50    G
                              wmi    3S   0.47    v l r q p h y C e k t F a d GSN
          wmh f r dy p e i n k       4C   0.50    g v QL A T S C
              d g n h r e k wq s p   5C   0.50    t y a mi VFLC
                                w    6S   0.39    c f y mh i p l g n Vd t Qe Rk AS
                               wc    7R   0.39    f my i v p l Ht Gq d a Ne S RK
                                w    8A   0.39    y f c mh i v l P Gr Nqt d S AKe
                                     9T   0.37    wc MP i Hg Fn v DRQTl YSk EA
```

**Table 2: The summarized results for the effect of substitutions:**

| Wild Type | Mutant Type | Effect on stability (I Mutant) | Effect on functions | Tolerance level (SIFT) |
|---|---|---|---|---|
| 7R | A | Decrease | Benign | T |
| 7R | V | Decrease | Benign | T |
| 7R | N | Decrease | Benign | T |
| 7R | D | Decrease | Benign | T |
| 7R | C | Decrease | Possibly Damaging | NT |
| 7R | E | Decrease | Benign | T |
| 7R | Q | Decrease | Benign | T |
| 7R | G | Decrease | Benign | T |
| 7R | H | Decrease | Possibly Damaging | T |
| 7R | I | Decrease | Benign | T |
| 7R | L | Decrease | Benign | T |
| 7R | K | Decrease |  | T |
| 7R | M | Decrease | Possibly Damaging | T |
| 7R | F | Decrease | Possibly Damaging | T |
| 7R | P | Decrease | Benign | T |
| 7R | S | Decrease | Benign | T |
| 7R | T | Decrease | Benign | T |
| 7R | W | Decrease | Possibly Damaging | NT |
| 7R | Y | Decrease | Possibly Damaging | T |

From the above results of I Mutant it can be inferred that any of the substitution at the position R would definitely result in the decrease in the stability of the protein thus resulting in the inactive form or defective form of the protein.

**SIFT results based on damaging substitutions:**

| Substitution | Score |
|---|---|
| R7C | 0.904 |
| R7H | 0.755 |
| R7M | 0.904 |
| R7F | 0.728 |
| R7W | 0.973 |
| R7Y | 0.728 |

## IV. Conclusion:

The current work aims to analyze the complete protein sequence and functional parameters of CDPK which is a major protein involved in the stress mechanism and cold tolerance. The study aimed to identify the Disorder regions and the recognition of the mutational hotspot sites on the protein sequence.

The domain regions present within the sequence have been calculated using SMART tool. The work also included the evaluation of phylogenetic relationship among the selected plant species and designing of the Dendrogram for the same. The physicochemical parameters of the sequence were calculated using Protparam. The PDB id that can depict the 3D structure of the protein was found to be 3SXF which was visualized using RASMOL. The major step of the research is to evaluate the effect of the substitution mutation at the hot spot site on the stability, functionality disease occurrence and tolerability of the protein. These parameters were calculated using various insilico tools. The final summary of the work is that the hotspot site located on the CDPK3 sequence of Oryza sativa is R7 position which when substituted with any of the other possible 19 amino acids would result in a decrease in the stability of the protein as shown by I Mutant tool. As per the results of SIFT it can also be inferred that though all the substitutions result in a decrease in the stability of the structure the mutations R7C and R7W are not tolerated whereas the other substitutions may be tolerated. According to the results of PolyPhen 6substitutions among the total 19 were found to cause a possibly damaging effect. These include R7C, R7H, R7M, R7F, R7W, and R7Y whose scores were further evaluated and it was found that 3 of the 6 are having maximum damage than the other substitutions the three maximum damaging substitutions are R7C, R7M, R7W. From the complete analysis it can be clear that the amino acid substitutions of R7C and R7W are found to be damaging with respect to the results of all the three tool I Mutant, PolyPhen and SIFT. Thus the site R7 and the substitutions with C and W are found to be the most damaging mutations that could ultimately affect the cold tolerance of the plants.

## Acknowledgement:

Omerald Biosolutions Pvt ltd is gratefully acknowledged for providing the extensive Insilico based research dry lab facilities.

## References:

[1]. OsCDPK13, a calcium-dependent protein kinase gene from rice, is induced by cold and gibberellin in rice leaf sheath, Abbasi. F et. Al, Plant Mol Biol. 2004 Jul; 55(4):541-52. PMID: 15604699

[2]. Over-expression of a single Ca 2+ -dependent protein kinase confers both cold and salt/drought tolerance on rice plants , Yusuke Saijo et al, Plant Journal- vol. 23, no. 3, pp. 319-327, 2000

[3]. Expression of the rice CDPK-7 in sorghum: molecular and phenotypic analyses, Tejinder Kumar Mall et al, Plant Molecular Biology, March 2011, Volume 75, Issue 4-5, pp 467-479

[4]. The CDPK super family of protein kinases, Alice C. Harmon et.al, New phytologist – Vol . 151, No 1, July 2001

[5]. Primary database NCBI available at www.ncbi.nlm.nih.gov

[6]. NCBI-BLAST: Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

[7]. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) ClustalW and ClustalX version 2. Bioinformatics **23(21)**: 2947-2948.

[8]. BOXSHADE tool for MSA, from SDSC BIOLOGY WORK BENCH

[9]. SMART: online tool for domain analysis: Schultz et al. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5857-5864, Letunic et al. (2012) *Nucleic Acids Res* , doi:10.1093/nar/gkr931

[10]. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; *Protein Identification and Analysis Tools on the ExPASy Server;* (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607

[11]. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995 Dec;11(6):681-684 Geourjon C, Deleage G
Institut de Biologie et de Chimie des Proteines, UPR 412-CNRS, Lyon, France.

[12]. RCSB PDB: An Information Portal to Biological Macromolecular Structures An Information Portal to Biological Macromolecular Structures

[13]. Roger Sayle and E. James Milner-White. "RasMol: Biomolecular graphics for all", *Trends in Biochemical Sciences (TIBS)*, September 1995, Vol. 20, No. 9, p. 374.

[14]. Exploring protein sequences for globularity and disorder, Nucleic Acid Res 2003 - Vol. 31, No.13 (OpenAccess)

[15]. R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson and R.B. Russell ”Protein disorder prediction: implications for structural proteomics
Structure Vol 11, Issue 11, 4 November 2003

[16]. Yang,Z.R., Thomson,R., McMeil,P. and Esnouf,R.M. (2005) RONN: the bio-basis function neural network technique applied to the dectection of natively disordered regions in proteins Bioinformatics 21: 3369-3376

[17]. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W306-10., I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure

[18]. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. Nat Methods 7(4):248-249 (2010).

[19]. Journal of Nucleic acid research, SIFT: predicting amino acid changes that affect protein function Pauline C. Ng and Steven Henikoff*