ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Structural and Functional Analysis of Dormancy Causing PDK4 Gene from Human and *Spermophilus Tridecemlineatus*

P Krishnamoorthy[1*], PK Praveen Kumar[2], P Rajasulochana[1] and D Krishna Priya [1]

[1]Dept .Bioinformatics., Bharath University, Chennai, India.
[2]Assistant Professor, Dept. Of Biotechnology, S.V.C.E.College,Sriperumbudur, Chennai, India.

ABSTRACT

This paper presents the details of analytical studies carried out by using advanced bioinformatics tools on classifying all proteins which are involved in causing dormancy in human & *Spermophilus tridecemlineatus* a selected animal (Squirrel) and analyzing their structures, keeping their active domains in major focus. The study primarily focused on (i) Identifying genes which are related to dormancy (ii) modeling PDK4 from different organisms (iii) structural analysis to find out the structural similarity and domain details and (iv) PDK4 pathway studies to give a detailed insight into mechanism of action of PDK4. From the studies, it is observed from the studies that squirrel PDK4 shows dormancy as natural adaptation compared to human PDK4. Typical case study of Ramachandran plot of human and squirrel PDK4 has been explained in detail.

Keywords: PDK4, *Spermophilus Tridecemlineatus,* gene

Corresponding author

ISSN: 0975-8585

INTRODUCTION

To encounter unfavorable conditions naturally certain animals developed various strategies to withstand them. Extreme temperatures like far above and far below animals' optimum temperatures is one such unfavorable condition, which certain animals face periodically in their life cycle. Many of these animals prefer to enter into dormancy state to overcome these unfavorable conditions. Animals which enter into dormancy are usually caused by certain alterations that occur in their bio-chemical metabolism. Dormancy is a period in an organism's life cycle when growth, development, and (in animals) physical activity is temporarily stopped. It is helpful to organisms to withstand unfavorable environmental condition. PDK4 gene will bind to PDK; will cause the reduction of metabolic activity to conserve energy. But in human period of dormancy can lead to fibromyalgia, chronic fatigue, obesity, dementias including Alzheimer s disease. Structural information of human PDK4 is very important for studying mode of action of PDK4 in human dormancy syndrome. Evolutionary history of PDK4 protein is important to analyze the changes that are happened during the evolution.

The process of dormancy appears to have evolved independently in a large number of different species, and, as such, there are a wide number of mechanisms by which a period of dormancy can be entered, depending on the physiology of the organism [1-3]. In many species, dormancy is an essential part of the life cycle, enabling an organism to change hugely, while undergoing minimal impact. In some cases, it allows organisms to survive huge climatic changes, such as a pond or river drying up. However, the type of dormancy most commonly associated with the word is that encountered when an animal becomes dormant during a period of cold weather. Dormancy, in short, allows organisms to adapt better to their environment during periods of hardship. It enables species to take advantage of certain environmental niches, which would otherwise be impossible to make habitable [4-6].

Main objectives of this study are:

- ✓ Identifying genes which are related to dormancy
- ✓ Modeling PDK4 from different organisms.
- ✓ Structural analysis to find out the structural similarity and domain details.
- ✓ PDK4 pathway studies to give a detailed insight into mechanism of action of PDK4.

MATERIALS AND METHODS

The following resources have been used to conduct analytical studies

National Center for Biotechnology Information -The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

EMBL Nucleotide Sequence Database - The EMBL Nucleotide Sequence Database (also known as EMBL-Bank) constitutes Europe's primary nucleotide sequence resource. Main sources for

DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications. The database is produced in an international collaboration with GenBank (USA) and the DNA Database of Japan (DDBJ).

PIR - The Protein Information Resource (PIR) is an integrated public bioinformatics resource to support genomic, proteomic and systems biology research and scientific studies (Wu et al., 2003).

PDB - A Resource for Studying Biological Macromolecules. The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards. The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Swiss-Prot - Protein knowledgebase

TrEMBL - Computer-annotated supplement to Swiss-Prot

Molecular Modeling Database (MMDB) - Experimentally resolved structures of proteins, RNA, and DNA, derived from the Protein Data Bank (PDB), with value-added features such as explicit chemical graphs, links to literature, similar sequences, related 3D structures, information about chemicals bound to the structures, and more. These connections make it possible, for example, to find 3D structures for homologs of a protein sequence of interest, then interactively view the sequence-structure relationships, active sites, bound chemicals and journal articles.

BLAST - The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

DaliLite - DaliLite Pairwise comparison of protein structures. Compares your structure(first structure) to a reference structure(second structure). It is a structural analysis tool.

ClustalW - ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

PhyloDraw - PhyloDraw is a drawing tool for creating phylogenetic trees. PhyloDraw supports various kinds of multialignment programs (Dialign2, Clustal-W, Phylip format, and pairwise distance matrix) and visualizes various kinds of tree diagrams, e.g. rectangular cladogram,

slanted cladogram, phylogram, free tree, and radial tree. With PhyloDraw users can manipulate the shape of a phylogenetic tree easily and interactively by using several control parameters. This program can export the final tree layout to BMP(bitmap image format) and Postscript.

CPHmodels 3.0 Server - CPHmodels 3.0 is a protein homology modeling server. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions.

PROSITE - PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [More details / References / Disclaimer / Commercial users]. PROSITE is complemented by ProRule, a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids.

BRENDA - The Comprehensive Enzyme Information System. BRENDA has the EC-Number Enzyme Name .BRENDA is a advanced search database.

KEGG - Kyoto Encyclopedia of Genes and Genomes. A grand challenge in the post-genomic era is a complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. Towards this end we have been developing a bioinformatics resource named KEGG as part of the research projects of the Kanehisa Laboratories in the Bioinformatics Center of Kyoto University and the Human Genome Center of the University of Tokyo.

Swiss-PdbViewer - The Swiss Institute of Bioinformatics presents Swiss-PdbViewer. Swiss-PdbViewer (aka DeepView) is an application that provides a user friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface.

FASTA - FASTA/SSEARCH/GGSEARCH/GLSEARCH - Protein Similarity Search
Provides sequence similarity searching against protein databases using the FASTA and SSEARCH programs. SSEARCH does a rigorous Smith-Waterman search for similarity between a query sequence and a database. GGSEARCH compares a protein or DNA sequence to a sequence database producing global-global alignment (Needleman-Wunsch). GLSEARCH compares a protein or DNA sequence to a sequence database. FASTA can be very specific when identifying long regions of low similarity especially for highly diverged sequences. You can also conduct sequence similarity searching against nucleotide databases or complete proteome/genome databases using the FASTA programs.

ISSN: 0975-8585

Methodology Adopted

In execution of the project, the following step by step procedure has been followed

- Identify the gene related to dormancy in human & suirrel
- Retrieve sequences
- Collect homologous genes from other organisms.
- Sequence analysis:
    - ✓ Pairwise sequence alignment
    - ✓ Multiple sequence alignment
    - ✓ Pattern searching
- Phylogenetic analysis
- Structure prediction – homology modeling
    - ✓ Template identification
    - ✓ Identify the % similarity of raw sequence to template
    - ✓ Download the tertiary structure
    - ✓ Align raw sequence to template
    - ✓ Submit modeling request
- Detailed structural validation
    - ✓ Global energy details
    - ✓ Comparison with non redundant set of PDBs
    - ✓ Estimation of local model quality
    - ✓ Ramachandran plot analysis
- Domain analysis.
    - ✓ Identify the domains.
    - ✓ Find out secondary structure details of domain
    - ✓ Figure out the exact location of domain on structure
- Classify the protein according to domain structure.
- Metabolic pathway studies.

OUTPUT AND VALIDATION

Pairwise sequence alignment of human and squirrel PDK4 shows 96% sequence similarity (Fig. 1), eventhough they are from different organisms. According to phylogenetic analysis pan troglytis and homosapience are more relative than others and the evulutionary steps of them are more similar. More evolutionary changes are happened to gallus PDK4 compared to others. Spermophilus PDK4 has undergone more changes than human PDK4 during the course of evolution. By pattern analysis PDK4 domain belongs to histidine kinase family, and the human PDK4 domain lie between 228-351 and squirrel PDK4 histidine kinase domain lie between 265 -368 region. It is helpful for the functional analysis of PDK4 sequence. As per the protein structure prediction methods like Homology Modeling, Threading and Ab initio methods, we are supposed to find the template for our sequence of interest. While finding the template for the PDK4, We have looked for its % identity or similarity with template.

ISSN: 0975-8585

As per the modeling scenario, if the % identity is more than 60%, we should go for Homology modeling, if is is in the range of 25-60%; should go for threading method and if it is below 20-25%; should go for Ab Initio method. As per the % identity we have got with the template after sending template selection request either through Swiss PDB viewer or directly through the online Swiss model server, we have selected templates for squirrel PDK4 % of identity is above 90%. For modeling human PDK4, 2ZDX structure was selected as it showing above 85% similarity to the raw sequence and 2EOA was selected as template as it showing 80% sequence identity to our raw sequence, and we have got 3D structure for human PDK4 and squirrel PDK4 sequence and viewed by Rasmol. Domain analysis have shown that both the PDK4 are belongs to same family and both the proteins contains 16 helices, 13 strands and 30 turns. But the domain of both proteins are lying in different position in sequence; ie in squirrel PDK4 it is in between 245-368 position and in human PDK4 it is in 228- 351 position.

Squirrel PDK4:

found: 1 hit in 1 sequence

GI-3298574-GB-AAC40161-1- (412 aa)

MKAARFAMHSARSLSSVGLVPREVELFSRYSPSPLSMKQLLDFGSDNACERTSFSFLRQELPVRLA
NILKEIDVLPDRLTNTSSVQLVKSWYIQSLMELVEFHEKSPEDQKNLSDFVDTLIKVRNRHHNVVP
TMAQGILEYKDTCTVDPVTNQSLQYFLDRFYMNRISTRMLMNQHILIFSDSQTGNPSHIGSIDPKC
DVVAVIQDAFESSKMLCDQYYLTSPELKLTQVNGKFPGQPIHIVYVPSHLHHMLFELFKNAMRATV
ERQESWPSLTPVEVIVVLGKEDLTIKISDRGGGVPLRITDRLFSYMYSTAPTPVMDNSRNAPLAGF
GYGLPISRLYAKYFQGDLNLYSLSGYGTDAIIYLKALSSESVEKLPVFNKSAFKHYQMSSEADDWC
IPSREPRNLSKEKMAM

Human PDK4:

found: 1 hit in 1 sequence

gi-4505693-ref-NP_002603-1- (411 aa)

MKAARFVLRSAGSLNGAGLVPREVEHFSRYSPSPLSMKQLLDFGSENACERTSFAFLRQELPVRLA
NILKEIDILPTQLVNTSSVQLVKSWYIQSLMDLVEFHEKSPDDQKALSDFVDTLIKVRNRHHNVVP
TMAQGIIEYKDACTVDPVTNQNLQYFLDRFYMNRISTRMLMNQHILIFSDSQTGNPSHIGSIDPNC
DVVAVVQDAFECSRMLCDQYYLSSPELKLTQVNGKFPDQPIHIVYVPSHLHHMLFELFKNAMRATV
EHQENQPSLTPIEVIVVLGKEDLTIKISDRGGGVPLRIIDRLFSYTYSTAPTPVMDNSRNAPLAGF
GYGLPISRLYAKYFQGDLNLYSLSGYGTDAIIYLKALSSESIEKLPVFNKSAFKHYQMSSEADDWC
IPSREPKNLAKEVAM

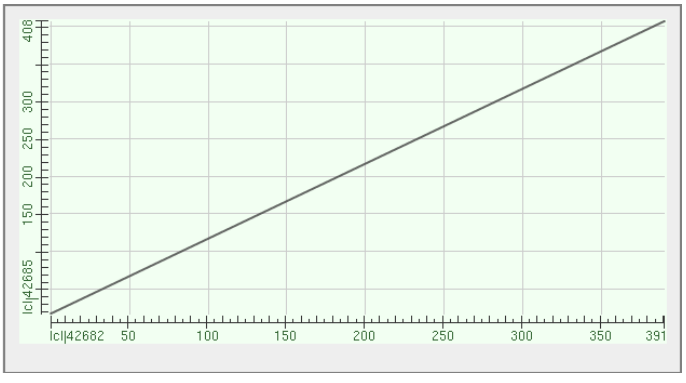


Fig. 1 Human PDK4 vs squirrel PDK4

ISSN: 0975-8585

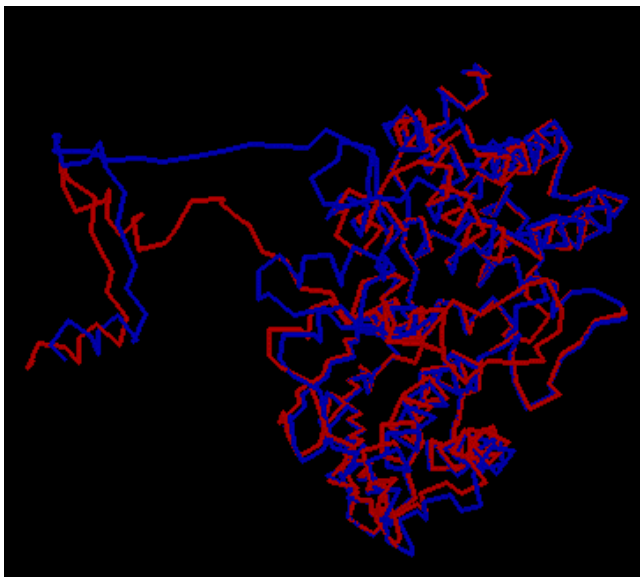
Alignment view:

```
 Score =  765 bits (1975),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 362/391 (92%), Positives = 378/391 (96%), Gaps = 0/391 (0%)

Query  1    GPVPREVEHFSRYSPSPLSMKQLLDFGSENACERTSFAFLRQELPVRLANILKEIDILPT  60
            G VPREVE FSRYSPSPLSMKQLLDFGS+NACERTSF+FLRQELPVRLANILKEID+LP
Sbjct  18   GLVPREVELFSRYSPSPLSMKQLLDFGSDNACERTSFSFLRQELPVRLANILKEIDVLPD  77

Query  61   QLVNTSSVQLVKSWYIQSLMDLVEFHEKSPDDQKALSDFVDTLIKVRNRHHNVVPTMAQG  120
            +L NTSSVQLVKSWYIQSLM+LVEFHEKSP+DQK LSDFVDTLIKVRNRHHNVVPTMAQG
Sbjct  78   RLTNTSSVQLVKSWYIQSLMELVEFHEKSPEDQKNLSDFVDTLIKVRNRHHNVVPTMAQG  137

Query  121  IIEYKDACTVDPVTNQNLQYFLDRFYMNRISTRMLMNQHILIFSDSQTGNPSHIGSIDPN  180
            I+EYKD CTVDPVTNQ+LQYFLDRFYMNRISTRMLMNQHILIFSDSQTGNPSHIGSIDP
Sbjct  138  ILEYKDTCTVDPVTNQSLQYFLDRFYMNRISTRMLMNQHILIFSDSQTGNPSHIGSIDPK  197

Query  181  CDVVAVVQDAFECSRMLCDQYYLSSPELKLTQVNGKFPDQPIHIVYVPSHLHHMLFELFK  240
            CDVVAV+QDAFE S+MLCDQYYL+SPELKLTQVNGKFP QPIHIVYVPSHLHHMLFELFK
Sbjct  198  CDVVAVIQDAFESSKMLCDQYYLTSPELKLTQVNGKFPGQPIHIVYVPSHLHHMLFELFK  257

Query  241  NAMRATVEHQENQPSLTPIEVIVVLGKEDLTIKISDRGGGVPLRIIDRLFSYTYSTAPTP  300
            NAMRATVE QE+ PSLTP+EVIVVLGKEDLTIKISDRGGGVPLRI DRLFSY YSTAPTP
Sbjct  258  NAMRATVERQESWPSLTPVEVIVVLGKEDLTIKISDRGGGVPLRITDRLFSYMYSTAPTP  317

Query  301  VMDNSRNAPLAGFGYGLPISRLYAKYFQGDLNLYSLSGYGTDAIIYLKALSSESIEKLPV  360
            VMDNSRNAPLAGFGYGLPISRLYAKYFQGDLNLYSLSGYGTDAIIYLKALSSES+EKLPV
Sbjct  318  VMDNSRNAPLAGFGYGLPISRLYAKYFQGDLNLYSLSGYGTDAIIYLKALSSESVEKLPV  377

Query  361  FNKSAFKHYQMSSEADDWCIPSREPKNLAKE  391
            FNKSAFKHYQMSSEADDWCIPSREP+NL+KE
Sbjct  378  FNKSAFKHYQMSSEADDWCIPSREPRNLSKE  408
```

Structure comparison- Dali lite:

No.	Second Structure & Chain	Z-Score	Aligned Residues	RMSD [Å]	Seq. Identity [%]	Structural Alignment	Superimposed C-alpha Traces
1	mol2	0.7	83	5.3	3	click here	CA 1.pdb

Super imposed C-alpha traces – Rasmol view:



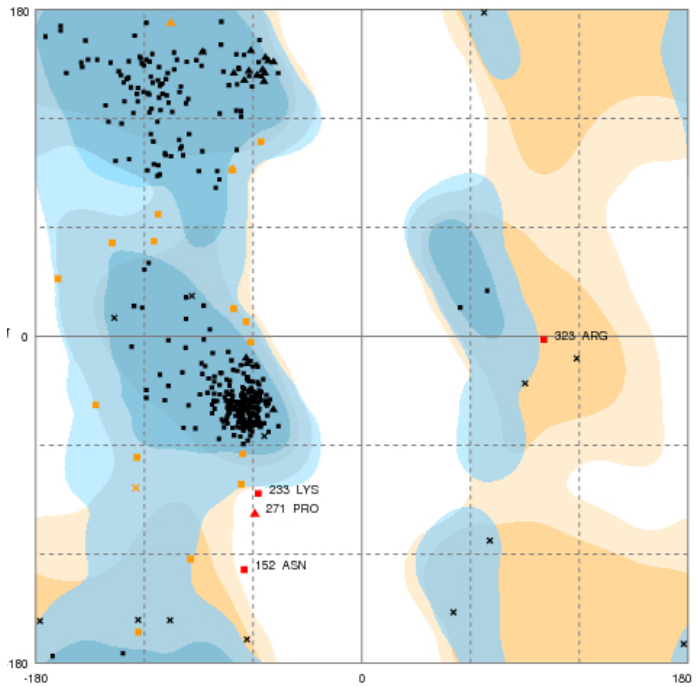
ISSN: 0975-8585

Each chain of mol1 is compared structurally to each chain of mol2 using the DaliLite program. The Dali method optimises a weighted sum of similarities of intramolecular distances. Sequence identity and the root-mean-square deviation of C-alpha atoms after rigid-body superimposition are reported for your information only; they are ignored by the structural alignment method. Suboptimal alignments do not overlap the optimal alignment or each other. Suboptimal alignments detected by the program are reported if the Z-score is above 2; they may be of interest if there are internal repeats in either structure. In the C-alpha traces, the chains of the first and second structure are renamed 'Q' and 'S', respectively. Z-Scores below 2 are not significant.

Ramachandran Plot analysis Result – Rampage

Ramachandran plot analysis is done for human and squirrel PDK4 structure by using rampage analysis tool and we have reached in following conclusion : 94.4% residues are fallen in favoured region for human PDK4, 4.3% residues are fallen in allowed region and 1.9% residues are fallen in outlined region. For squirrel PDK4 90.6% residues are coming in favoured region for human PDK4, 5.1% residues are coming in allowed region and 0.8% residues are fallen in outlined.

Ramachandran plot of human PDK4



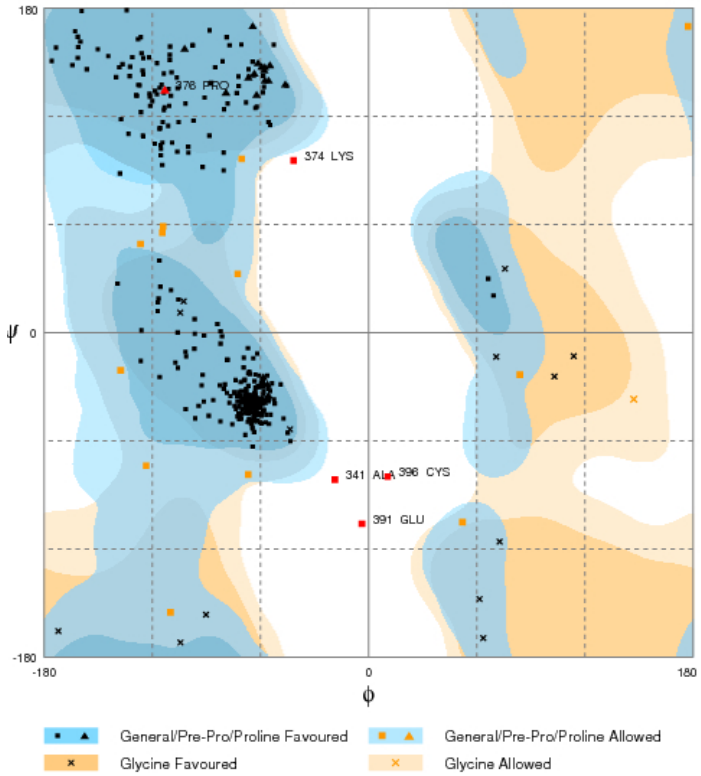
Evaluation of residues

Residue [26 :HIS] (-65.71, -64.78) in Allowed region
Residue [29 :ARG] (-63.79, 8.12) in Allowed region
Residue [43 :PHE] (-61.40, -3.24) in Allowed region
Residue [45 :SER] (-94.53,-122.65) in Allowed region
Residue [47 :ASN] (-71.43, 91.63) in Allowed region

ISSN: 0975-8585

Residue [141 :TYR] (-70.66, 15.32) in Allowed region
Residue [149 :PRO] (-71.50, 92.57) in Allowed region
Residue [151 :THR] (-112.42, 67.25) in Allowed region
Residue [197 :ASN] (-137.70, 51.54) in Allowed region
Residue [269 :ASN] (-66.49, -81.52) in Allowed region
Residue [294 :ARG] (-114.56, 52.35) in Allowed region
Residue [326 :PRO] (-105.40, 173.23) in Allowed region
Residue [330 :PHE] (-123.24,-162.92) in Allowed region
Residue [333 :GLY] (-124.68, -83.55) in Allowed region
Residue [354 :SER] (-55.62, 107.21) in Allowed region
Residue [356 :TYR] (-124.06, -66.72) in Allowed region
Residue [390 :SER] (-167.65, 31.69) in Allowed region
Residue [391 :GLU] (-146.75, -37.84) in Allowed region
Residue [152 :ASN] (-64.88,-128.52) in Outlier region
Residue [233 :LYS] (-57.13, -86.60) in Outlier region
Residue [271 :PRO] (-59.04, -97.12) in Outlier region
Residue [323 :ARG] (100.49, -1.85) in Outlier region
Number of residues in favoured region (~98.0% expected) : 368 (94.4%)
Number of residues in allowed region (~2.0% expected) : 18 (4.6%)
Number of residues in outlier region : 4 (1.0%)

Ramachandran plot for Squirrel PDK4:



Evaluation of residues
Residue [46 :ASP] (52.15,-105.14) in Allowed region
Residue [73 :ASP] (-72.64, 32.60) in Allowed region
Residue [74 :VAL] (-137.45, -20.86) in Allowed region

Residue [152 :ASN] (-70.35, 96.38) in Allowed region
Residue [197 :LYS] (-126.50, 49.22) in Allowed region
Residue [243 :TYR] (177.19, 169.92) in Allowed region
Residue [294 :ARG] (-114.38, 55.49) in Allowed region
Residue [310 :MET] (83.87, -23.32) in Allowed region
Residue [324 :ASN] (-66.75, -78.65) in Allowed region
Residue [329 :GLY] (147.09, -36.86) in Allowed region
Residue [356 :TYR] (-123.43, -73.80) in Allowed region
Residue [395 :TRP] (-109.92,-155.08) in Allowed region
Residue [411 :ALA] (-113.95, 58.87) in Allowed region
Residue [341 :ALA] (-18.68, -81.53) in Outlier region
Residue [374 :LYS] (-41.59, 95.46) in Outlier region
Residue [376 :PRO] (-113.03, 134.82) in Outlier region
Residue [391 :GLU] (-3.74,-106.05) in Outlier region
Residue [396 :CYS] (10.63, -79.80) in Outlier region
Residues in most favoured regions [A,B,L] 307 90.6%
Residues in additional allowed regions [a,b,l,p] 29 8.6%
Residues in generously allowed regions [~a,~b,~l,~p] 3 0.9%
Residues in disallowed regions 0 0.0%
---- ------
Number of non-glycine and non-proline residues 339 100.0%
Number of end-residues (excl. Gly and Pro) 5
Number of glycine residues (shown as triangles) 16
Number of proline residues 24

Total number of residues 384

CONCLUDING REMARKS

Structures were predicted for human and squirrel PDK4 by using CPH model server and SPDBV. Structures were selected based on minimum energy criterion. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor nto greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions in Ramachandran plot. So the qualities of predicted structures are good. The PDK4 sequence and structure of both the organisms are found to be very close similarity and its domain histidine kinase superfamily.

REFERENCES

[1] Barger JL, Brand MD, Barnes BM, Boyer BB. Am J Physiol Regul Integr Comp Physiol 2003; 284:R1306-1313
[2] Bocharova LS, Gordon RY, Arkhipov VI: Uridine uptake and RNA synthesis in the brain of torpid and awaken ground squirrels. Comp Biochem Physiol 101B:189-192, 1992
[3] French AR. J Comp Physiol 1985; 156B:13-19.
[4] Frerichs KU, Smith CB, Brenner M, DeGracia DJ, Krause GS, Marrone L, Dever TE, Hallenbeck JM. Proc Natl Acad Sci USA 1998; 95:14511-14516
[5] Hittel DS, Storey KB. Arch Biochem Biophys 2002; 401:244-254.
[6] Ivanov KP. J Therm Biol 2000; 25:467-479