The Pharma Innovation Journal 2019; 8(2): 142-149

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; 8(2): 142-149
© 2019 TPI
www.thepharmajournal.com
Received: 01-12-2018
Accepted: 04-01-2019

Divya V
A) Department of Chemistry, M.S.M College, Kayamkulam, Kerala, India
B) Department of Chemistry, Sree Narayana College, Kollam, Kerala, India

VL Pushpa
Department of Chemistry, Sree Narayana College, Kollam, Kerala, India

Sunitha VR
Department of Chemistry, M.S.M College, Kayamkulam, Kerala, India

S Sarithamol
Department of Chemistry, Sree Narayana College, Kollam, Kerala, India

Suchitra Surendran
Department of Chemistry, Sree Narayana College, Kollam, Kerala, India

Manoj KB
A) Department of Chemistry, Sree Narayana College, Kollam, Kerala, India
B) Department of Chemistry, Sree Narayana College, Cherthala, Kerala, India

Correspondence
VL Pushpa
Department of Chemistry, Sree Narayana College, Kollam, Kerala, India

A Pharmacogenomic approach for the identification and validation of genetic markers behind CDK4 amplification leading to Cancer

Divya V, VL Pushpa, Sunitha VR, S Sarithamol, Suchitra Surendran and Manoj KB

Abstract
Cyclin Dependent Kinase 4 gene has been identified as the key regulator in many types of cancerous tumors. Identification of genetic involvement behind the expression of this gene is therefore a promising field while developing suitable treatment measures for cancers. Pharmacogenomic study of Cyclin Dependent Kinase 4 has revealed the genes associated with CDK4 expression and different epigenetic as well as Metagenomic influences that encourage CDK4 expression. This helps to identify the deleterious single nucleotide polymorphisms related to this gene. The study also focused to understand the structure of corresponding CDK4 protein and developed a high quality homology model of the protein so as to facilitate structure based drug discovery of CDK4 inhibitor.

Keywords: Pharmacognomics, single nucleotide polymorphisms, genetic signatures, epigenomics

Introduction
In the modern world, cancer is the most dreaded disease [1, 2] characterized by invasive, metastatic and angiogenesis properties. Cancer can be caused by genome level mutations by severe exposure to environmental factors such as exposure to UVB radiations, chemical substances, tobacco smoke, alcohols, reproductive life, diet, chemotherapeutic drugs etc. Conventional chemotherapeutic drugs have both cytostatic and cytotoxic effects [3] and its major drawback of these drugs is that they have no differentiation between rapidly dividing normal cells and cancer cells. Therefore, due to their cytotoxic effect they develop severe DNA damage to both cancer cells and normal cells. This leads to the development of new cancer in other parts of the body. To overcome this situation, we need more specific drugs that can specifically target the damaged cells. This influenced the development of personalized medicines.

The advent in the field of genomics and proteomics provided a fine distinction between the normal cells and cancerous cells. This leads to the designing of more specific drugs for specific cancers by identifying the specific changes accompanied in the genome level. The discovery of uniqueness of an individual's genome profile leads to the concept of personalized medicine. International Cancer Genome Consortium (ICGC) [6] has been constituted for collecting, organizing and characterizing different cancer genomes worldwide.

Hyperactivation of Cyclin Dependent Kinase 4 (CDK4) and its regulatory subunit, Cyclin D is reported in many types of cancers. Pharmacogenomic study of CDK4, in this regard, can be an effective way to design new molecular species for regulating Cyclin D- CDK4 pathway in tumor genesis. Pharmacogenomics deals with the effect of an individual's genetic inheritance with the body's response to drugs. Determination of *genetic markers* or Single Nucleotide Polymorphisms (SNPs) [7] is used to understand the mechanism of over expression of Cyclin Dependent Kinase 4. These SNPs plays a direct role in disease by affecting the gene function. An individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases can be predicted using SNPs. SNPs can also be used to track the inheritance of disease genes within families.

In addition to above factors certain epigenetic factors consisting of 847 attribute classified into 10 classes also contribute to the mutations. The ten classes are chromosome organization, DNA sequence, DNA structure, Epigenome and chromatin structure, Evolution History, Genes, Population variation, Regulatory regions, Repetitive DNA and Transcriptome.

In the present study the methylation possibility of CDK4 gene was studied using EpiGRAPH, a web based SVM tool [8]. The structure and characteristics of CDK4 protein is necessary for designing new molecular species. The size of active site and the chemical nature of the residues present in the active site region of CDK4 play a very crucial role in proposing a new molecule.

Altered expression or mutations of CDK4 can result in tumerogenesis since it incites cell cycle dysregulation. Because of the incidence of CDK4 over expression in various types of cancers such as breast cancers, lung cancers, tumors etc. are very common; CDK4 belongs to the class of most potential targets in cancer research.

Materials and methods

There are many genetic mutations which cause the development of cancer in which out of controlled growth of abnormal cells occurs. The present study deals with the identification of various mutations in CDK4 gene. The characteristics of this gene was found out using National center for Biotechnology Information (www.ncbi.nlm.nih.gov) [9] data base and Gene Cards: the Human Gene Data base (www.genecards.org) [10]. Mutations in the associated genes of the gene under consideration also contribute to the development of cancer. The associated genes of CDK4 gene were identified using STRING database (www.string-db.org) which consists of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, knowledge transfer between organisms, and interactions collected from other (primary) databases.

All the human pathogenic missense SNPs were collected from NCBI database, a depository of biomedical and genomic information. Deleterious SNPs were collected from GWAS[11] Central. SNPs were considered as deleterious when their SIFT score is in the deleterious region and PolyPhen-2 lies in the probably/possible damaging region. The effect of amino acid substitution on protein function can be predicted by using SIFT. The prediction is based on the degree of conservation of aminoacid residues in sequence alignments derived from closely related sequences collected through PSI-BLAST. PolyPhen-2 (Polymorphism Phenotyping v2) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Database of CpG islands and Analytical Tool (DBCAT)[11] is used for finding out the various CpG islands within the DNA sequence. EpiGRAPH [11] is another software tool for genome and epigenome analysis. EpiGRAPH analysis, which is based on Support Vector Machines, predicts the various genomic attributes that are directly predictive of DNA methylation and which are only predictive via indirect correlations. Both linear and nonlinear calculations were done by using these tools [11].

The contributions of other microbes towards the development of cancer were studied by using a technique called BLAST - Basic Local Alignment Search Tool. BLAST finds the regions of similarity between the selected gene sequence and that of various sequences of Microbes presents in the sequence database and calculates its statistical significance.

The proteins associated with this gene were collected from genecards and NCBI. For characterizing the protein molecule, sequence-wise and structure wise analysis is important. The primary sequence data of the human CDK4 protein molecule in the fasta format was collected from UniProtKB (UniProt id: P11802). PROTPARAM tool of Expasy was used for identifying the physico-chemical properties of CDK4. Secondary structure was analyzed by using SOPMA tool of Expasy. The topology of the protein molecule was identified by using this method. It describes the number of alpha helix, Beta Turn, Bend region, Random Coil etc.

Since no inhibitor bound CDK4 protein crystal structure was reported, a homology model of CDK4 was constructed using the crystal structure of CDK2 (1H1Q; resolution 2.5 Å, R-value 0.238 and R-value Free 0.332) (www.rscb.org)[12],[13] complexed with a CDK4 inhibitor, NU6094. FASTA sequence of CDK4 (P11802) was downloaded from Uniprot. Prime module of Schrodinger is used for building the model. An energy based model building method was adopted by which insertions were constructed and gaps were closed with segments from known structures and returned to a single model of the structure. The inhibitor molecule bound with the template protein structure was also selected while building the model. The reliability of the developed homology model was checked with both by analyzing the Ramchandran plot and Verify 3D (http://services.mbi.ucla.edu/Verify_3D) [14, 15] and found to be appropriate.

Result and Discussions

Cyclin Dependent Kinase 4 (CDK4) belongs to Serine-threonine protein kinase family. Expression of CDK4's is regulated by Cyclin D. G1/S cell cycle transition is favored by the activity of these proteins. Activity of CDK4 is chiefly controlled by Cyclin D1 and CDK4 inhibitor, p16- a tumor suppressor protein (INK4a). CDK4 form a holoenzyme with Cyclin D1 which phosphorylates Retinoblastoma protein (pRb) and thus releases transcription factors (E2Fs) which promotes gene transcription. p16 inhibits the interaction between CDK4 and Cyclin D1 and thus produce cell cycle arrest. Mutation of these proteins alters the cell cycle progression and its effect is observed in many types of cancers [16–23].

CDK4 gene is located in the chromosome region 12:57747727bp - 57756013bp (8287 bases). In the uncomplexed form, it is present in the cytoplasmic area and after the formation of holoenzyme with Cyclin D, it moves towards the nucleus membrane and enters into the nucleus as the cell cycle progress from G1 to S phase.

Associated genes of CDK4

Mutations present in the associated genes of CDK4 may also contribute to the development of cancer. The list of associated genes which have interaction score 0.90 and its chromosome location were presented in figure.1. Interaction scores were determined from the active sources like text mining, experiments, databases, co-expression neighborhood interactions, gene fusions and co-occurrences. The scores do not indicate the strength or the specificity of the interaction. Instead they are indicators of confidence. RB1, CCND1, CCND2 and CCND3 have interaction scores 0.999.

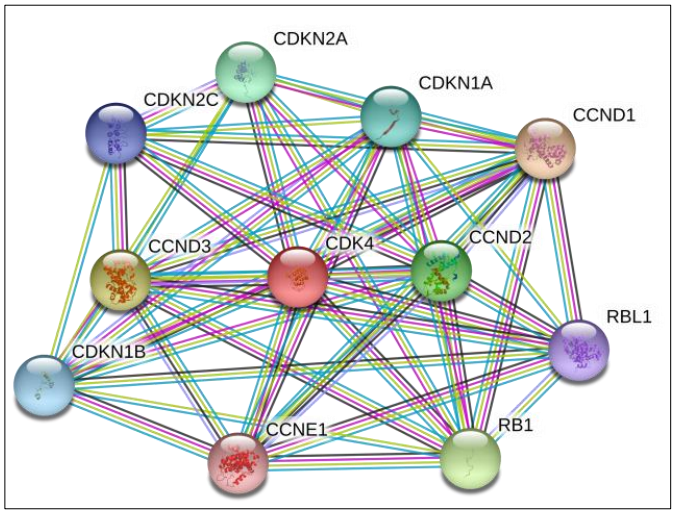

Fig 1: Associated genes of CDK4 (https://string-db.org. version 10.5)

CCND1(chromosome region 11:69641105-69654474), CCND2 (chromosome region 12: 4273736-4305356) and CCND3 (chromosome region 6: 42048894-41934933) are the regulatory subunits of Cyclin D- CDK4 complex that regulates cell cycle progression through G1 to S phase by phosphorylation of retinoblastoma protein family members (pRb). As a result of phosphorylation pRb, dissociation of pRb/E2F complex occurs and thereby liberates E2Fs and proceeds to next phase of cell cycle. RB1 (chromosome region 13: 48303747-48481890) - tumor suppressor gene family member- acts as repressor of E2F (transcriptional factors) target genes. It promotes G0-G1 phase transition during the cell cycle when phosphorylated by CDK3/Cyclin C. CDKN1A and CDKN1B (interaction scores 0.998) are important Cyclin dependent kinase inhibitors-p27, Kip1 family members which acts either as an inhibitor or an activator CyclinD-CDK4 complexes depending on its phosphorylation states and/or stoichiometry. Cyclin dependent kinase inhibitor 2A (CDKN2A) acts as a negative regulator of the proliferation of normal cells by interacting strongly with CDK4 and CDK6 (interaction score 0.998). This inhibits their ability to interact with CyclinD and phosphorylate RB. Cyclin dependent kinase2C (CDKN2C, p18; interaction score 0.996) interacts strongly with CDK6, weakly with CDK4. CCNE1- Cyclin E1 (interaction score 0.994) is essential for the control of the cell cycle at the G1/S (start) transition. Retinoblastoma-like 1(p107) – RBL1 with CDK4 interaction score 0.996 is a key regulator into cell division. It directly involved in the heterochromatin formation by maintaining overall chromatin structure and in particular, that of constitutive heterochromatin by stabilizing histone methylation [24].

Characterization of genes

Human pathogenic missense mutations of CDK4 and its associated genes- CCND1, CCND2 and CCND3 with interaction scores 0.999 were identified. Deleterious SNPs of these genes were considered as the genetic signatures or genetic markers behind the development and progression of cancers. They are listed in the following table.1.

Table 1: Number of human pathogenic, missense and deleterious Single Nucleotide polymorphisms (SNPs) in CDK4, CCND1, CCND2, CCND3 and RB1genes

Gene Name	Chromosome Number	Number of SNPs	Number of missense SNPs	Number of pathogenic SNP	Number of Deleterious SNPs
CDK4	12	1700	174	1	1
CCND1	11	4007	169	NIL	NIL
CCND2	12	7430	129	5	5
CCND3	6	23345	254	NIL	NIL
RB1	13	33043	713	24	17

Single Nucleotide Polymorphisms (SNPs) is a common type of genetic variation occurring within a population in which a single nucleotide- A, T, C or G varies within the DNA between the genes. These SNPs were also known as biological markers or genetic signatures as it helps to locate those genes that are associated with a particular disease. It plays a direct role in disease by affecting the gene function. An individual's response to certain drugs, susceptibility to environmental factors such as toxins, and risk of developing particular diseases can be predicted using SNPs. SNPs can also be used to track the inheritance of disease affected genes within families.

SNPs in the coding regions are of two types synonymous SNPs and non-synonymous SNPs-of which synonymous SNPs do not affect the amino acid sequence in protein and non-synonymous SNPs change the amino acid sequence of the protein. Missense mutation is one form of non-synonymous SNPs in which a point mutation results in a codon that code for a different amino acid whereas nonsense mutation which is another form of non-synonymous SNP in which a codon is changed into a premature stop codon that result in the truncation of resulting protein. The SNPs potentially affect the protein structure and its functions and thereby results in human inherited diseases are called deleterious SNPs. Therefore these SNPs are called the genetic signature behind a particular disease. The number of human pathogenic missense SNPs and the deleterious SNPs were listed in the Table.2 given below. These non-synonymous single-nucleotide polymorphisms occurring in protein coding regions may alter the encoded amino acid; potentially affect protein structure and function, which results in human inherited diseases.

Table 2: Genetic signatures of CDK4, CCND2 and RB1

Gene	Deleterious SNP	Chromosome location	Allele Change	SIFT	Poly Phen
CDK4	rs11547328	12:57751648	G/A	deleterious	benign
CCND2	rs587777618	12:4299977	A/G	deleterious	possibly damaging
	rs587777620	12:4299978	C/A/T	deleterious	probably damaging
	rs587777621	12:4299980	C/T	deleterious	probably damaging
	rs587777622	12:4299981	C/G/T	deleterious	probably damaging
	rs777786993	12:4299990	T/C/G	deleterious	probably damaging
RB1	rs115596308	13:48411837	G/A/C	deleterious	probably damaging
	rs121434307	13:48411862	T/A	deleterious	probably damaging
	rs121434309	13:48411859	C/T	deleterious	probably damaging

	rs137853292	13:48452997	C/T	deleterious	probably damaging
	rs137853294	13:48459708	C/T	deleterious	probably damaging
	rs137853296	13:48463758	C/T	deleterious	probably damaging
	rs768305224	13:48360053	A/C/T	deleterious	probably damaging
	rs776534331	13:48367518	A/G/T	deleterious	probably damaging
	rs587778834	13:48459700	A/C/T	deleterious	probably damaging
	rs587778847	13:48379607	G/T	deleterious	probably damaging
	rs587778859	13:48459828	A/G	deleterious	probably damaging
	rs768305224	13:48360053	A/C/T	deleterious	probably damaging
	rs776534331	13:48367518	A/G/T	deleterious	probably damaging
	rs879255262	13:48412236	T/A	deleterious	probably damaging
	rs1131690851	13:48379606	G/A	deleterious	probably damaging
	rs1131690857	13:48459832	A/G	deleterious	probably damaging
	rs1131690884	13:48379633	G/T	deleterious	probably damaging

The effect of amino acid substitution in protein function based on its sequence was studied by using SIFT calculations. All the five SNPs were deleterious as per SIFT analysis. PolyPhen-2 (Polymorphism Phenotyping v2) analysis showed that amino acid substitution probably damage the structure and functioning of the human protein in the case of rs587777620, rs587777621, rs587777622 and rs777786993. Allele change corresponding to the deleterious SNP, rs11547328 results in the residue change of Arginine residue which is a positively charged amino acid at 24th position of protein by Cysteine which is polar in nature. rs587777618 with an allele change Adenine -a purine base to Guanine- another purine base in the chromosome position 12:4299977codes for the residue change at 280 Threonine to Alanine. Another residue change of polar threonine to nonpolar isoleucine is observed at position 280 corresponding to SNP rs587777620. Allele change from Cytosine to Thymine results in a residue change of nonpolar Proline to polar Serine, at position 281 (rs587777621). In rs587777622, Proline is replaced by Leucine at position 281. When the allele change at 12:4299990 is Thymine to Cytosine, hydrophobic Valine at position 284 gets replaced by Alanine. When the allele changes to Guanine, a purine base, Valine at position 284 gets replaced by polar Glycine. These types of deleterious mutations results in dysregulation in cell cycle progression which in turn results in cancer generation and progression.

Epigenetics

Methylation, acetylation, phosphorylation, ubiquitylation etc. are few types of processes that alter the gene activity without changing the DNA sequence and resulting modifications can be transmitted to daughter cells. These epigenetic modifications regulate many cellular processes necessary for many organism functions. But improper modification can lead to adverse effects [25].

Three CpG islands have been located in CDK4 gene; proposing a high degree of mutation possibility in this region (Figure.2). They are 12:57747813 - 577484990, 12:57751257 - 57751477 and 12:57752217 – 57752427. Chromosome region from 12:57747813 to 12:57748499 of length 686 was identified as the densest CpG methylation regions of more than 70% GC population.

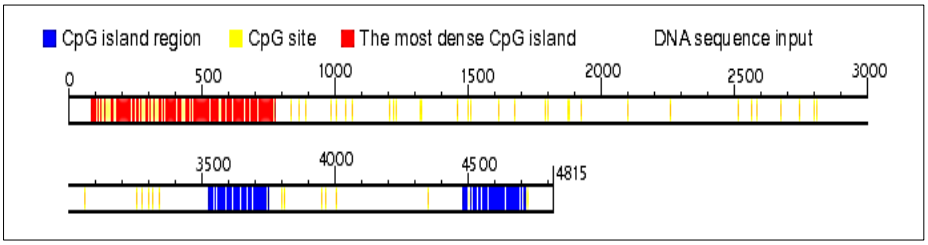

Fig 2: CpG island regions in CDK4 gene. Red colour region is the densest CpG island region and blue colour region is the CpG Island region (http://dbcat.cgm.ntu.edu.tw/)

We performed an EpiGRAPH analysis on our dataset. The methylation possibilities of the gene were studied on the basis of 507 attributes which were classified into 10 classes - chromosome organization, DNA sequence, DNA structure, Epigenome and chromatin structure, Evolution History, Genes, Population variation, Regulatory regions, Repetitive DNA and Transcriptome using EpiGRAPH which uses powerful Support Vector Machine (SVM) learning algorithms based on epigenome information. The summary of pairwise statistical comparisons between the positives and negatives calculated by using Wilcoxon's rank-sum test and Fischer's exact test were listed in the form of a Doughnut diagram given below (Figure.3). Of these ten classes, evolution history exhibits 100% correlation with CDK4 gene towards the methylation of the CpG islands (sensitivity, specificity and accuracy 100%) with least standard deviation (less than 0.1). Chemical modifications of histone proteins by methylation results in the variations in the chromatin structure which in turn influence the epigenetic changes in the CDK4 gene. EpiGRAPH predicted the relation of Epigenome -chromatin structure based on 81 attributes, with a correlation of 72.4% (sensitivity, 91.4%; specificity,80.5%; correlation 85.9%). DNA. Attributes associated with DNA structure with 51.6% correlation is also integrated with DNA mutation and histone modification. Repetitive DNA was likely to be related to gene expression (38.4%). Involvement of miRNA in mutation is understood by correlation of transcriptome with CDK4. The interdependence of Repetitive DNA and Epigenome and Chromatin structure is also clear since repetitive regions exhibit repressive chromatin structure by ignoring all other factors.

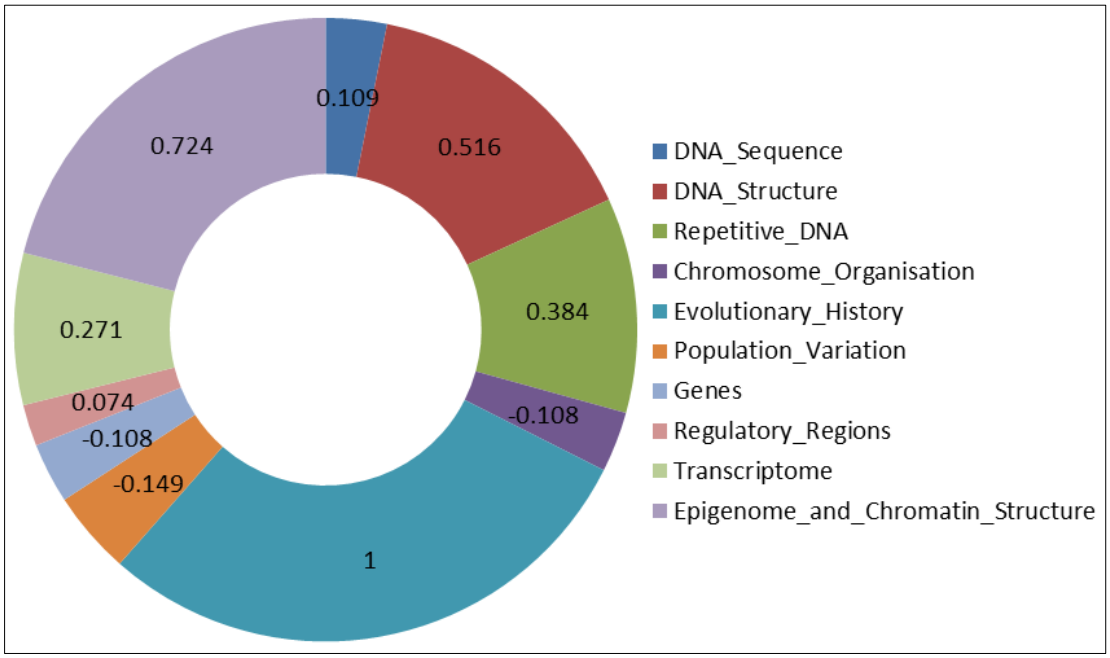

Fig 3: Results of linear kernel auto correlation analysis using Support Vector Machine (computed using Epi GRAPH; https://epigraph.mpi-inf.mpg.de)

Effect of microorganisms and influence of environmental factors

Methyloceanibacter marginalis strain R-67177 contig_56 exhibits 95% sequence identity with CDK4 gene. Species like Kurthia massiliensis strain JC30T, Nocardia brevicatena NBRC 12119, Sanguibacteroides justesenii strain OUH 308042 contig12 and Colwellia marinimaniae strain MTCD1 also have above 80% genome sequence identity with that of CDK4 and therefore, these microbes may influence the functions of CDK4. Sodium arsenite, (+)-JQ1 compound (thienotriazolodiazepine), resveratrol, Tetrachlorodibenzodioxin, Estradiol, Diethylnitrosamine, Naringin, Nicotine and Acetaminophen (http://ctdbase.org)[26] were cause variations in CDK4 gene expressions. Tetrachlorodibenzodioxin is produced as a byproduct during several organic syntheses and burning of organic compounds. This chemical has a tendency to amplify CDK4 expression and is known for its carcinogenicity. Another carcinogen, Diethylnitrosamine present in tobacco smoke also induces CDK4 gene expression and thereby induces cancer cell growth. Nicotine on the other hand has a tendency to diminish CDK4 expression [27–29]. Estradiol, the major female sex hormone has shown to have a tendency to express CDK4 expression [30–34]. Even though Acetaminophen, commonly known as paracetamol, has many beneficial effect in the therapeutic fields, cases were reported that it has a tendency to enhance CDK4 activity [35]. Resveratrol present in grapes, blueberries, raspberries and mulberries has a tendency to suppress the activity of CDK4 [36–38]. Sodium arsenite, (+)-JQ1 compound (thienotriazolodiazepine) and Naringin also lowers the expression of CDK4. Palbociclib, Ribociclib and Abemaciclib are the three specific inhibitors that inhibit CDK4 and are approved by FDA for treating breast cancer [39] patients.

Protein characterization

Protein analysis plays a significant role in drug designing and helps to recognize the protein structure, properties and characteristics of the target. Understanding the characteristics of the corresponding protein has been achieved by defining primary and secondary structure of CDK4 protein.

For illustrating the protein molecule, the primary sequence data of CDK4 molecule in fasta format were collected. The different properties like Instability index, aliphatic index, pH, half-life period were analyzed by using Prot Param tool of EXPASY (Table.3). Secondary structure was also analyzed by using SOPMA tool of EXPASY (Table.4).

Table 3: Primary structural features of CDK4 protein. (https://web.expasy.org/protparam)

Number of Amino acids	303
Molecular weight	33729.8
Theoretical pI	6.51
Estimated half life	30hrs
Instability Index	39.0
Aliphatic Index	89.74
GRAVY	-0.167

The relative volume occupied by aliphatic side chains is defined as aliphatic index of a protein. It considered as a positive factor for the increasing of thermo stability of globular protein. The thermodynamic stability of proteins increases as the aliphatic index increases. The results indicate that the CDK4 protein molecule is thermodynamically stable. The hydrophilicity of CDK4 was determined by GRand A Verage hydropath Y (GRAVY). The protein molecule under consideration has negative GRAVY value which clearly indicates the hydrophilic nature of it. Half-life period of CDK4 was 30hrs which shows mutated form is highly stable. Half -life period of the protein is the time required for the protein to reduce its activity to half of its initial activity. For measuring the stability of proteins, instability index is used. It is also used to identify whether it will be stable in a test tube or not. If the value of Instability index is less than 40, it is probably stable. If it is greater, it will not stable. CDK4 shows an instability index value 39 indicating its moderate stability in test tube. Existence of zwitter ionic form of proteins

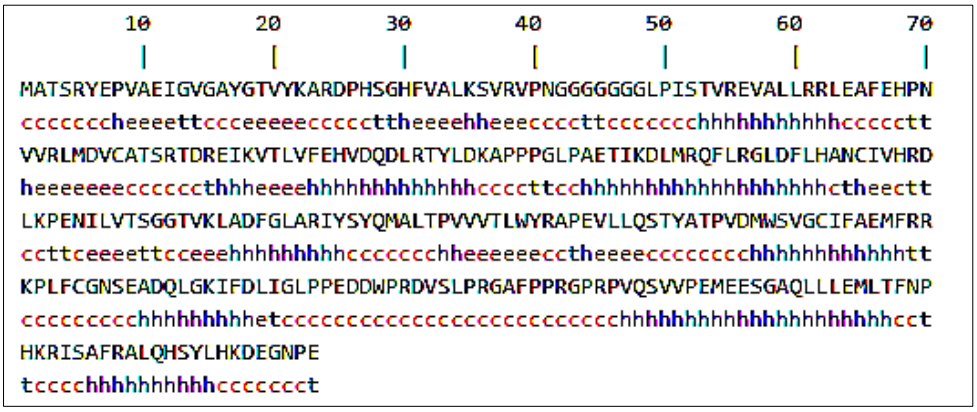

Fig 4: Exhibits higher the presence of random coil type of proteins which suggests a larger unfolding of protein or lack of stabilization. All the protein structures under consideration were stable because the percentage of random coil is comparatively low.

Table 4: Secondary structure results

Alpha Helix	116	38.28%
Extended strand	47	15.51%
Beta turn	25	8.25%
Random coil	115	37.95%

Homology modeling

It has been found that binding with the inhibitor molecules cause conformational changes in the active site of the protein from inactive (DFG-IN) to active (DFG–OUT). Since no inhibitor bound CDK4 protein crystal structure (active form) was reported, a homology model of CDK4 was constructed using the crystal structure of CDK2 (1H1Q; resolution 2.5 Å, R-value 0.238 and R-value Free 0.332) complexed with a CDK4 inhibitor, NU6094. NU6094 has a phenylpyrimidine moiety which is responsible for the hydrogen bonding interaction with the hinge region residue Val96. Validation of the modeled protein was checked by Ramachandran Plot (Figure.5). Presence of majority of residues in the favored region as indicated by the red coloured region and few of the residues in the yellow region confirms the usefulness of the model.

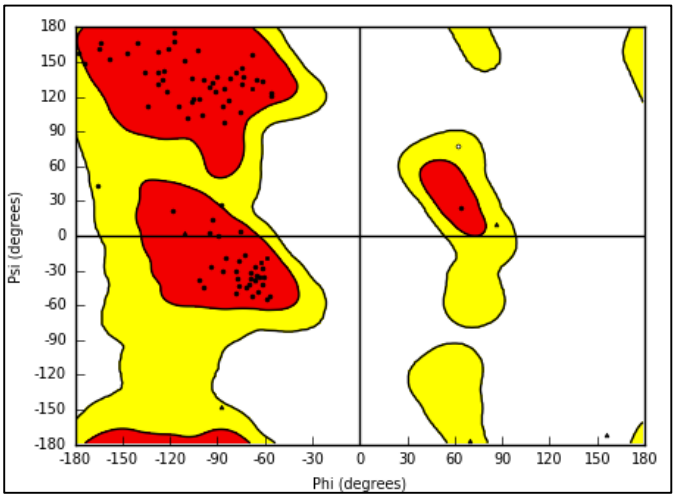

Fig 5: Ramachandran plot of homology modeled CDK4 protein

Evaluation of modeled protein structure using Verify_3D indicates that 92.03% of the residues have an averaged 3D-1D score >=0.2. The hinge region residues His 95, Val96 and Asp97 are having values 0.4, 0.41 and 0.43 respectively. These results substantiates that the modeled protein can be used for docking studies[40]. Protein preparation was finished by using Protein preparation Wizard of Schrodinger. Homology modeled CDK4 bound with ligand NU6094 was selected. The protein structure was preprocessed by assigning bond orders and by adding hydrogen atoms. Water molecules that bridge between ligand and the proteins were retained and other water molecules were removed outside 5.00Å from heterogroup. A restrained minimization is done by selecting RMSD tolerance of 0.30 Å and also lightened potential steric clashes. This structure was selected for further docking studies and 3D QSAR model generations.

Conclusion

For designing a new drug molecule, the identification of a target protein is not an easy task. Understanding the genetic involvement behind each disease is very important. Cyclin Dependent Kinase 4 or CDK4 is one of the key regulator cell cycle is found to be overexpressed in many varieties of cancer. Therefore understanding the genome level significance of CDK4 is very significant. CDK4 gene mutations were found to play a vital role in carcinogenesis. Mutations in the associated genes may also influence the overexpression of CDK4. CCND1, CCND2, CCND3 and RB1genes are the associated genes of CDK4 with confidence score of 0.999. Human pathogenic missense SNPs of all these genes were analyzed and CDK4 has only one pathogenic, missense, deleterious SNPs, CCND2 has five pathogenic, missense, deleterious SNPs and RB1 has 17 pathogenic, missense and deleterious SNPs. This pathogenic, missense, deleterious SNPs are considered as the genetic signatures of carcinogenesis. In addition to evolutionary history, DNA structure and changes in the chromatin structure were also shown marked influence for the mutation of CDK4. Methyloceanibacter marginalis strain R-67177 contig_56, Kurthia massiliensis strain JC30T, Nocardia brevicatena NBRC 12119, Sanguibacteroides justesenii strain OUH 308042 contig12 and Colwellia marinimaniae strain MTCD1may cause mutations in CDK4 gene sequence since their genome possess more than 80% sequence identity with that of CDK4.

Both primary and secondary structural analysis of CDK4 protein using online tools in ExPASy revealed the stability of CDK4. For developing drug molecule, we need to have a crystal structure of a protein with a ligand in the active site. Due to the absence of such a crystal structure of CDK4, a homology model was also built using the prime module of Schrodinger software suit. Validation of the modeled protein was checked by Ramachandran Plot. Presence of majority of residues in the favored region as indicated by the red coloured

region and few of the residues in the yellow region confirms the usefulness of the model. The quality of the model was also confirmed by Verify_3D online software tool. The structure was then validated and was optimized for further studies using protein preparation wizard of Schrodinger software suit.

Acknowledgements
We thank Dr. P. K. Krishnan Namboori, Computer Engineering and Networking, AMRITA Vishwa Vidyapeetham, Coimbatore, India for his assistance while doing pharmacogenomic study. We extend our hearty gratitude to UGC for sanctioning Faculty Development Programme for completing this research work.

References

1. What Is Cancer? - National Cancer Institute. www. Cancer.gov/about-cancer/understanding/what-is-cancer.
2. What Is Cancer? https://www.cancer.org/cancer/cancer-basics/what-is-cancer.html.
3. Lauren Pecorino. Molecular Biology of Cancer Mechanisms, Targets and Therapeutics. third. Oxford; 2012.
4. Lynda Chin JNA. Cancer genomics: from discovery science to personalized medicine. Nat Med. 2011; 17:297–303.
5. Garraway LA, Jänne PA. Circumventing cancer drug resistance in the era of personalized medicine. Cancer Discov. 2012; 2(3):214-226.
6. Hudson TJ. The International Cancer Genome Consortium. Cancer Research. http://www.icgc.org. Published, 2009.
7. Namboori PK, Vineeth KV, Rohith V, Hassan I, Sekhar L, Sekhar A NM. The ApoE gene of Alzheimer's disease (AD). Funct Integr Genomics. 2011; 11(4):519-522.
8. Bock C, Halachev K, Büch J, Lengauer T. EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi) genomic data. 2009; 10(2):20-23.
9. Kim D, Pruitt Garth R, Brown Susan M. Hiatt Françoise Thibaud-Nissen Alexander Astashyn Olga Ermolaeva Catherine M. Farrell Jennifer Hart Melissa J. Landrum Kelly M. McGarvey Michael R. Murphy Nuala A. O'Leary Shashikant Pujar Bhanu Rajput Sanjida H. Rang. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014; 42(D1):D756-D763.
10. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S *et al*. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016; 54:1.
11. Bock C, Walter J, Paulsen M, Lengauer T. CpG Island Mapping by Epigenome Prediction. Plos Comput Biol. 2007; 3(6):e110.
12. Berman HM, Westbrook J, Feng Z *et al*. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235-242.
13. Davies TG, Bentley J, Arris CE *et al*. Structure-based design of a potent purine-based Cyclin- dependent kinase inhibitor. Nat Struct Biol. 2002; 9(10):745-749.
14. Ru C, Ru G. y-Randomization and Its Variants in QSPR / QSAR. Journel Chem Inf Model. 2007; 47:2345-2357.
15. Luthy R, Bowie JU ED. Assessment of protein models with three-dimensional profiles. Nature. 1992; 356(6364):83-85.
16. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CDK4.
17. https://www.ncbi.nlm.nih.gov/gene/1019#bibliography.
18. Michelle L Cook Sangar, Laura A Genovesi, Madison W Nakamoto, Melissa J. Davis, Sue E. Knobluagh, Pengxiang JI, Amanda Millar, Brandon J. Wainwright JMO. Inhibition of CDK4/6 by Palbociclib Significantly Extends Survival in Medulloblastoma Patient-Derived Xenograft Mouse Models. Clin Cancer Res, 2017, 112250.
19. Geoerger B, Bourdeaut F, Dubois SG *et al*. A Phase I Study of the CDK4 / 6 Inhibitor Ribociclib (LEE011) in Pediatric Patients with Malignant Rhabdoid Tumors, Neuroblastoma, and Other Solid Tumors. Clin Cancer Res, 2017, 2433-2442.
20. Pradhan JBR, Kumar AGR, Upadhyay VSAD. CDK4 in lung, and head and neck cancers in old age: evaluation as a biomarker. Clin Transl Oncol, 2016.
21. Shao Y, Shen Y, Li Y, Liang C, Zhang B. Direct repression of the oncogene CDK4 by the tumor suppressor miR-486-5p in non-small cell lung cancer. Oncotarget. 2016; 7(23):34011-34021.
22. Geer LY, Marchler-bauer A, Geer RC *et al*. The NCBI BioSystems database. Nucleic Acids Res. 2010; 38(August):492-496.
23. Dai M, Zhang C, Ali A *et al*. CDK4 regulates cancer stemness and is a novel therapeutic target for triple-negative breast cancer. Nat Publ Gr, 2016, 1-15.
24. https://string-db.org.
25. Iyer PM, Karthikeyan S, Sanjay Kumar P, Krishnan Namboori PK. Comprehensive strategy for the design of precision drugs and identification of genetic signature behind proneness of the disease-a pharmacogenomic approach. Funct Integr Genomics. 2017; 17(4):375-385.
26. Allan Peter Davis Cynthia J. Grondin Kelley Lennon-Hopkins Cynthia Saraceni-Richards Daniela Sciaky Benjamin L. King Thomas C. Wiegers Carolyn J. Mattingly. The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. Nucleic Acids Res. 2015; 43:D914-D920.
27. Liu Q, Zhang J, Zhu H, Qin C, Chen Q ZB. Dissecting the signaling pathway of nicotine-mediated neuroprotection in a mouse Alzheimer disease model. FASEB J. 2007; 21(1):61-73.
28. Shin VY, Jin HC, Ng EK, Yu J, Leung WK, Cho CH SJ. Nicotine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone induce cyclooxygenase-2 activity in human gastric cancer cells: Involvement of nicotinic acetylcholine receptor (nAChR) and beta-adrenergic receptor signaling pathways. Toxicol Appl Pharmacol. 2008; 233(2):254-261.
29. Frazer-Abel AA, Baksh S, Fosmire SP, Willis D, Pierce AM, Meylemans H, Linthicum DS, Burakoff SJ, Coons T, Bellgrau D MJ. Nicotine activates nuclear factor of activated T cells c2 (NFATc2) and prevents cell cycle entry in T cells. J Pharmacol Exp Ther. 2004; 311(2):758-769.
30. Wróbel AM GE. Differential effect of methyl-, butyl- and propyl paraben and 17β-estradiol on selected cell cycle and apoptosis gene and protein expression in MCF-7 breast cancer cells and MCF-10A non-malignant cells. J Appl Toxicol. 2014; 34(9):1041-1050.
31. Kuo KL, Lin WC, Ho IL, Chang HC, Lee PY, Chung YT *et al*. 2-methoxyestradiol induces mitotic arrest, apoptosis, and synergistic cytotoxicity with arsenic trioxide in human urothelial carcinoma cells. PLoS One. 2013; 8(8):e68703.

32. Wang W, Smith R 3RD SS. Aryl hydrocarbon receptor-mediated antiestrogeni city in MCF-7 cells: modulation of hormone-induced cell cycle enzymes. Arch Bio sChem Biophys. 1998; 356(2):239-248.
33. Gupta RK, Singh JM, Leslie TC, Meachum S, Flaws JA YH. Di-(2-ethylhexyl) phthalate and mono-(2-ethylhexyl) phthalate inhibit growth and reduce estradiol levels of antral follicles *in vitro*. Toxicol Appl Pharmacol. 2010; 242(2):224-230.
34. Zhu H, Xiao X, Zheng J, Zheng S, Dong K YY. Growth-promoting effect of bisphenol A on neuroblastoma *in vitro* and *in vivo*. J Pediatr Surg. 2009; 44(4):672-680.
35. Bhushan B, Chavan H, Borude P *et al*. Dual Role of Epidermal Growth Factor Receptor in Liver Injury and Regeneration after Acetaminophen Overdose in Mice. Toxicol Sci. 2017; 155(2):363-378.
36. Benitez DA, Pozo-Guisado E, Alvarez-Barrientos A, Fernandez-Salguero PM CE. Mechanisms involved in resveratrol-induced apoptosis and cell cycle arrest in prostate cancer-derived cell lines. J Androl. 2007; 28(2):282-293.
37. Roccaro AM, Leleu X, Sacco A, Moreau AS, Hatjiharissi E, Jia X *et al*. Resveratrol exerts anti proliferative activity and induces apoptosis in Waldenströms macro globulinemia. Clin Cancer Res. 2008; 14(6):1849-1858.
38. Bai Y, Mao QQ, Qin J, Zheng XY, Wang YB, Yang K *et al*. Resveratrol induces apoptosis and cell cycle arrest of human T24 bladder cancer cells *in vitro* and inhibits tumor growth *in vivo*. Cancer Sci. 2010; 101(2):488-493.
39. Xu H, Yu S, Liu Q *et al*. Recent advances of highly selective CDK4/6 inhibitors in breast cancer. J Hematol Oncol. 2017; 10(1):97. http://jhoonline.biomedcentral.com/articles/10.1186/s13045-017-0467-2.
40. Divya V, Pushpa VL, Sarithamol S, Manoj KB. Computational approach for generating robust models for discovering novel molecules as Cyclin Dependent Kinase 4 inhibitors. J Mol Graph Model. 2018; 82:48-58.