IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 4, April 2015.
www.ijiset.com
ISSN 2348 – 7968

In Silico Drug Designing Of Tripeptide-S Against Blm Protein For Treatment Of Bloom Syndrome.

Raj Kumar Ahirwar[1], Shradha Choudhary[2], Archana Singh[3] & Dr. Smt. Rakesh Mehta[4,],
[4]Head of Botany & Biotech Department, Govt. M.G.M. P.G. College, Itarsi (M. P.) India,

Abstract: - Genetic polymorphisms impairing the DNA repair capacity can disrupt the genomic integrity and potentially modulate individual's susceptibility to various cancers. Mutations in the *blm* gene cause human bloom syndrome (bs), an autosomal recessive disorder of growth retardation, immunodeficiency and cancer predisposition, leukemia. Bloom's syndrome is associated with chromosomal aberrations, and affected individuals have an increased incidence of leukemia and solid tumors. The treatment of the very rare and chronic leukemia is done through the computer aided drug design (cadd) approach. The present work deals with *in silico* designing of the most potent ligand tripeptide-s by screening the pubchem database and further optimizing the target receptor (blm) on argus lab so that the receptor and the drug can dock perfectly yielding the maximum efficacy. The toxicity and tolerability of the ligand was also studied in detail to rule out the chances of serious side effects.

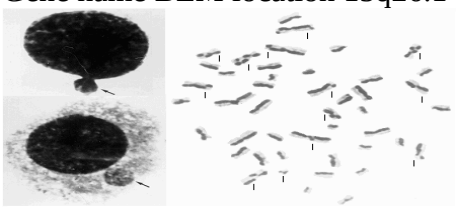
Keyword: - *blm protein -bloom syndrome protein,*

Abbreviation:- *cadd-computer aided drugs design, bs-bloom syndrome, BLAST (Basic Local Alignment Search Tool), PDB (Protein Data Bank), HNN (Hierarchical Neural Network), SOPMA (self-optimized prediction method with alignment) , PHYRE- (protein homology/ analog Y Recognition Engine), KEGG (Kyoto Encyclopedia of genes and Genomes)*

Introduction: - Bloom syndrome is an inherited disorder characterized by a high frequency of breaks and rearrangements in an affected person's chromosomes. People with Bloom syndrome are much smaller than average, and often have a high-pitched voice and characteristic facial features including a long, narrow face; small lower jaw; and prominent nose and ears. They tend to develop pigmentation changes and dilated blood vessels in the skin, particularly in response to sun exposure. These changes often appear as a butterfly-shaped patch of reddened skin on the face. The skin changes may also affect the hands and arms. Other features of the disorder may include intellectual disabilities, chronic lung problems, diabetes, and immune deficiency that lead to recurrent pneumonia and ear infections. Men with Bloom syndrome usually do not produce sperm, and as a result are unable to father children (infertile). Women with the disorder generally experience menopause earlier than usual. Chromosome instability in Bloom syndrome results in a high risk of cancer in affected individuals. Affected individuals develop

the full range of cancers found in the general population, but the cancers arise unusually early in life. People with Bloom syndrome may be first diagnosed with cancer at about 25 years old.

Gene name BLM location 15q26.1



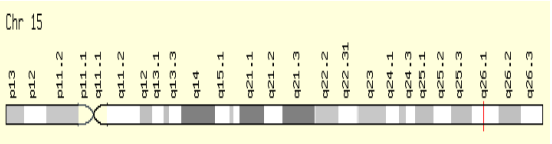
Protein

Description	1417 amino acids; contains one ATP binding site, one DEAH box, and two putative nuclear localization signals
Expression	accumulates to high levels in S phase of the cell cycle, persists in G2/M and sharply declines in G1; hyperphoshorylated in mitosis
Localisation	nuclear (PML nuclear bodies and nucleolus)
Function	· 3-5 DNA helicase; probable role in DNA replication and double-strand break repair · Preferred substrates: G-quadruplex DNA, D-loops structures and X-junctions. Recombinant protein promotes ATP-dependent branch migration of Hollyday junctions. · participates in a supercomplex of BRCA1-associated proteins named BASC (BRCA1-Associated genome Surveillance Complex) and in a complex named BRAFT (BLM, RPA, FA, Topoisomerase IIIalpha) containing five of the Fanconia Anemia (FA) complementation group proteins (FANCA, FANCG, FANCC, FANCE and FANCF). · Interacts physically and/or functionally with p53, 53BP1,WRN, MLH1, RAD51, TRF2, ligase IV, FEN1 · Associated with telomeres and ribosomal DNA repeats. · Phosphorylated in mitotic cells through the cdc2 pathway, and in response to DNA damaging agents.

Results:-

Disease Name: Bloom Syndrome, **Symptoms**: Leukemia, Blood cancer, Gene: BLM. Alternative gene names: RECQ2, RECQL3. **Locus**: 15q26.1 Chromosomal position: Protein: BLM (Bloom syndrome protein) **Size**: 1417 amino acids; 159000 Da.

Subcellular location: Nucleus. *Alternative name(s):* RecQ protein-like 3, DNA helicase, RecQ-like type 2, Short name=RecQ.



Function: Participates in DNA replication and repair. Exhibits a magnesium-dependent ATP- dependent DNA-helicase activity that unwinds single- and double-stranded DNA in a 3'-5' direction. Protein sequence: >sp|P54132|BLM_HUMAN Bloom syndrome protein OS=Homo sapiens GN=BLM PE=1 SV=1

Material and Methods

TOOL AND DATABASE

GENECARDS: - Genecards is a database of human genes that provides genomic, proteomic, transcriptomic, genetic and functional information on all known and predicted human genes it is being developed and maintained by the crown human genome center at the Weizmann institute of science.

BLAST: - It is the most widly used sequence similarity search programme. It

was develop by Eugene myers. David lipman and web Miller in 1990, it finds region of local similarity between.

PUBMED: - It is the free database accessing the MEDLINE database of citations, Abstract and some full text articles on life sciences and some biomedical topics. NLM at the NIH maintain pubmed as a part of Entrez information retrival system.

PDB: - It is repository for the 3-D structural data of large biological molecules, such as proteins and nucleic acids. Data is obtained by x-ray crystallography and spectroscopy.

PHYLOGENETIC ANALYSIS

BIOLOGY WORKBENCH: - The Biology Workbench is a web-based tool for biologists. It allows the biologists to search many popular protein and nucleic acid sequence database. Data base searching is integrated with access to wide variety of analysis and modeling tools. All with a point and click interface that elimates file format problems.

CLUSTAL-W: - Clustal is widely used multiple sequence alignment computer programs. W stand for weighted matrix there is two main variations: A clustal w – command line interface B clustal X- this version has a graphical user interface. It is available for Window, macos and linux/ unix. Clustal W is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignment of divergent sequence. It calculates the best match for the selected sequence and then lines them up sp that the identities, similarities and the differences can be seen. Evolusionary relationship can be viewed via cladograms or phylograms/ dandograms

DENDROGRAM: - The dendrogram is a graphical representation of the results of hierarchical cluster analysis. This is a tree-like plot where each step of hierarchical clustering is represented as a fusion of two branches of the tree into a single one. The branches represent clusters obtained on each step of hierarchical clustering.

TEXSHADE:- It is based on local alignment of sequence. It is alignment-shading software. In addition to common shading algorithms, it provides special shading modes showing functional aspects, e.g. charge or hydropathicity and a wide range of commands for handling shading colors, text styles, labels, legends.

BOXSHADE: - This program generates publication quality (or editable) graphics from multiple alignment files. It works by global alignment of all sequence. Many different input and output formats are supported.

Conserved and/or similar residues are emphasized by various degrees of shading. It is also possible to specify a similarity threshold that has to be reached for shading a residue.

CLUSTAL DISTANCE MATRIX: - The matrix contains data that allows relationship between a given set of elements (DNA and protein sequence). Values in the matrix file show distance, similarity or identity between different sequences.

STRUCTURAL ANALYSIS

PROTPARAM: - Protparam (References/ documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in swiss-prot or TrEMBL or for a user entered sequence. The computed parameters include the molecular weight, theoretical pl, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and

grand average of hydropathicity (GRAVY)

HNN: - The HNN (Hierarchical Neural Network) prediction method can be seen as an improvement on the famous classifier developed by Qian and Sejnowski, and derived from the system NETtalk (Guermeur). As its predecessor, it is made up of two networks: a sequence-to-structure network and a Structure-to-structure network. The prediction is thus only based on local information.

SOPMA: - SOPMA is a secondary structure prediction method. SOPMA (self-optimized prediction method with alignment) is an improvement of SOPMA method. These methods are based on the homologue method of Levin et al. SOPMA correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha- helix, beta sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins. Joint prediction with SOPMA and a neural network method (PHD) correctly predicts 82.2% of residues for 74% of co-predicted amino acids.

CPHMODELS: - CPHModels- 3.0 is a web predicting protein 3D structure by use of single template homology modeling. The CPHModels server predicts structure from amino acid sequence with respect to distance Constraints. Cphmodels is protein structure prediction using comparative (homology) modeling; distance predicts contacts between C-alpha atoms from the amino acid sequence using a neural network based method: and the RedHom tool finds a subset with low sequence similarity in a database.

HHPRED: - HHPred is a free server for protein function, structure and detection of remotely related sequence.

PHYRE: - The models produced by phyre are based on finding a sequence alignment to a known structure, copying the coordinates are relabeling the residues according to your sequence (based on the alignment). It can detect remotely homologous structures that can't be found by sequence and the sequence of the known structure. Phyre performs a profile-profile matching algorithm together with predicted secondary structure matching.

CONFORMATORY TOOLS

KEGG: - KEGG (Kyoto Encyclopedia of genes and Genomes) is a collection of online database dealing with genomes, enzymatic pathways, and biological chemicals, the Pathway database records network of molecular interaction in the cells, and variants of them specific to particular organisms.

PFAM: - Pfam is a database of protein families that includes their annotations and multiple sequence alignments generated using hidden Markov models. The pfam database contains information about protein domains and families. Pfam-A is the manually curated protein if the database that contains over 10,000 entries. For each entry a protein sequence alignment and a Hidden Markev model is stored. Because the entries in pfam-A do not cover all known proteins, an automatically generated supplement is provided called pfam-B. Pfam-B contains a large number of small families derived from clusters produced by an algorithm called ADDA. Although of lower quality, pfam-B families can be useful when no pfam-A families are found.

PRODOM: - Prodom is a comprehensive database of protein domain families from the global comparison of all available protein sequence.

PROSITE: - Prosite is a database of protein families and domains. It consists of entries describing the domains, families and functional sites as well as amino acid patterns, signatures, and profiles in them. Prosite's uses include identifying possible functions of newly discovered proteins and analysis of known proteins for previously undetermined activity. Properties from well-studied genes can be propagated to biologically related organisms, and different or poorly known genes biochemical functions can be predicted from similarities. Prosite offers tools for protein sequence analysis and motif detection (see sequence motif, prosite petterns). It is part of the ExPASy proteomics analysis servers.

VISUALIZATION TOOLS

RASMOL: - Rasmol is computer program written for molecular graphics visualization intended and used primarily for the depiction and exploration of biological macromolecule structure such as those found in the protein data bank. It was originally developed by Roger Sayle in the early 90's. Rasmol includes a language for selecting for visualization from the research collaborator for structural Bioinformatics (RCSB) bank.

DRUG DESIGNING

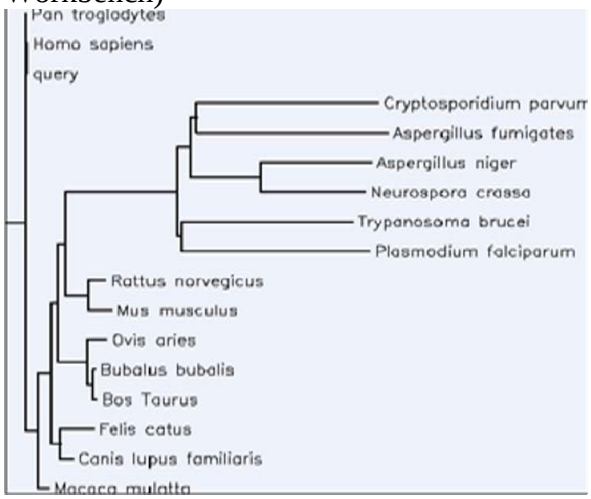
ARGUS LAB:- Argus lab is a molecular modeling program that runs on windows 98, NT, and 2000. It consists of a user interface that supports open GL graphics display of molecule structure and runs quantum mechanical calculation using the Argus compute server.

DRUG BANK: - The drugBank database available at the University of Alberta is a bioinformatics and chemoinformaics resource that combines detailed drug (i.e. chemical,pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains nearly 4800 drug entries including. >1480 FDA approved small molecule drugs 128 FDA approved biotech (protein/peptide) drugs >71 nutraceuticals >3200experimental drugs More than 2500 protein (i.e. drugs target, non-redundant) sequence are linked to these drug entries. Each drugcard entry contain more than 100 data fields with half of the information being devoted to drug/chemical data and the half devoted to drug target or protein data.

PUBCHEM: - Pubchem is a database of chemical molecules. The system is maintained by NCBI, pubchem can be accessed for free through a web user interface. Millions of compound structure and

PHYLOGENETIC ANALYSIS:-

Dendrogram By Phylip (Biology Workbench)



Inference:

The above dendrogam shows that the query has been derieved from *Homo sapiens as both of them are on the same branch.* Pan troglodytes are the closest to query as it is on the same root. The organisms showing the highrical arrangement are on the lower side showing horizontal chain transfer where

organism in upper side showing vertical sequence and the most derived species among them are *aspergillus fumigates*.

CLUSTAL W:-

Homo_sapiens (1) 0.000 0.000 0.000 0.057 0.174 0.107 0.152 0.159 0.196 0.193 0.207 0.824 0.786 0.871 0.864 0.920 0.914

query (2) 0.000 0.000 0.000 0.057 0.174 0.107 0.152 0.159 0.196 0.193 0.207 0.824 0.786 0.871 0.864 0.920 0.914

Pan_troglodytes (3) 0.000 0.000 0.000 0.054 0.167 0.102 0.152 0.159 0.193 0.184 0.197 0.833 0.790 0.871 0.864 0.924 0.912

Macaca_mulatta (4) 0.057 0.057 0.054 0.000 0.167 0.095 0.166 0.172 0.200 0.177 0.197 0.833 0.783 0.871 0.878 0.931 0.918

Felis_catus (5) 0.174 0.174 0.167 0.167 0.000 0.111 0.194 0.201 0.208 0.215 0.229 0.844 0.815 0.882 0.882 0.931 0.910

Canis_ familiar (6) 0.107 0.107 0.102 0.095 0.111 0.000 0.172 0.179 0.214 0.204 0.211 0.841 0.804 0.891 0.878 0.931 0.905

Bubalus_bubalis (7) 0.152 0.152 0.152 0.166 0.194 0.172 0.000 0.014 0.069 0.241 0.262 0.831 0.787 0.862 0.883 0.924 0.897

Bos_Taurus (8) 0.159 0.159 0.159 0.172 0.201 0.179 0.014 0.000 0.069 0.248 0.269 0.838 0.794 0.862 0.883 0.931 0.897

Ovis_aries (9) 0.196 0.196 0.193 0.200 0.208 0.214 0.069 0.069 0.000 0.255 0.283 0.846 0.801 0.862 0.903 0.931 0.910

Rattus_norveg (10) 0.193 0.193 0.184 0.177 0.215 0.204 0.241 0.248 0.255 0.000 0.095 0.833 0.783 0.891 0.884 0.945 0.925

Mus_musculus (11) 0.207 0.207 0.197 0.197 0.229 0.211 0.262 0.269 0.283 0.095 0.000 0.819 0.790 0.891 0.871 0.931 0.912

Aspergillus niger (12) 0.824 0.824 0.833 0.833 0.844 0.841 0.831 0.838 0.846 0.833 0.819 0.000 0.527 0.913 0.923 0.897 0.921

Neurospcrassa (13) 0.786 0.786 0.790 0.783 0.815 0.804 0.787 0.794 0.801 0.783 0.790 0.527 0.000 0.909 0.901 0.905 0.916

Trypan_brucei (14) 0.871 0.871 0.871 0.871 0.882 0.891 0.862 0.862 0.862 0.891 0.891 0.913 0.909 0.000 0.916 0.950 0.910

Plasm_falciparu (15) 0.864 0.864 0.864 0.878 0.882 0.878 0.883 0.883 0.903 0.884 0.871 0.923 0.901 0.916 0.000 0.934 0.927

Cryptospo_parv (16) 0.920 0.920 0.924 0.931 0.931 0.931 0.924 0.931 0.931 0.945 0.931 0.897 0.905 0.950 0.934 0.000 0.901

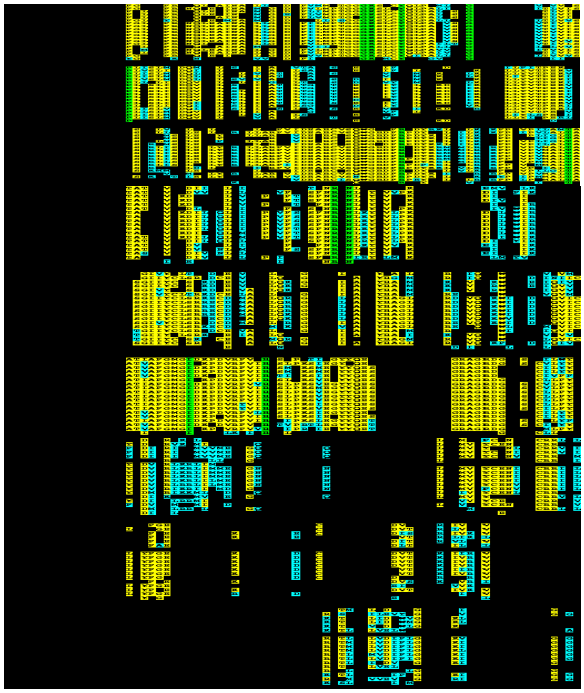
Aspergfumigate (17) 0.914 0.914 0.912 0.918 0.910 0.905 0.897 0.897 0.910 0.925 0.912 0.921 0.916 0.910 0.927 0.901 0.000

Inference:

Clustal distance matrix is an tool which show the evolution of species with time in relation with the dendogram. The *Homo sapiens* and the query are the same position, and the *Pan troglodytes* changing from the third position, and the *Macaca_mulatta* changing from fourth position, and the *Felis catus* changing from the fifth position, and the *Canis familiaris* changing from the six position, and the *Bubalus bubalis* changing the seventh position, and the *Bos_Taurus* changing from the eighth position, and the *Ovis_aries* changing from the ninth position, And the *Rattus_norvegicus* changing from the tenth position, and the *Mus_musculus* changing from the

eleventh position, and the *Aspergilusl niger* changing from the twelfth position, and the *Neurospora crassa* changing from the thirteenth position, and the *Trypan_brucei* changing from fourthteenth position, and the *Plasmodium_falciparum* changing from fifthteenth position, and the *Cryptospo_parv* changing from sixthteenth position, and the *Aspergilus fumigate* changing from seventeenth position,

BOX SHADE:-



Inference:

Boxshed shows the conserved regions,similar regions and is just a program for pretty- printing multiple alingment output. In boxshade yellow color shows the conserved domins,cyan color shows the Identical regions.

STRUCTURAL ANALYSIS

PROTPARAM:-

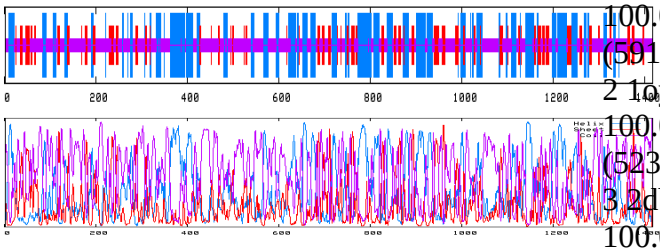
Number of amino acids: 1417
Molecular weight: 159000.4
Theoretical pI: 7.33

Amino acid composition: Ala (A) 75 5.3% CSV format
Arg (R) 65 4.6%
Asn (N) 72 5.1%
Asp (D) 103 7.3%
Cys (C) 30 2.1%
Gln (Q) 59 4.2%
Glu (E) 85 6.0%
Gly (G) 61 4.3%
His (H) 35 2.5%
Ile (I) 68 4.8%
Leu (L) 126 8.9%
Lys (K) 123 8.7%
Met (M) 28 2.0%
Phe (F) 57 4.0%
Pro (P) 66 4.7%
Ser (S) 157 11.1%
Thr (T) 90 6.4%
Trp (W) 9 0.6%
Tyr (Y) 34 2.4%
Val (V) 74 5.2%
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): 188
Total number of positively charged residues (Arg + Lys): 188
Atomic composition:
Carbon C 6948
Hydrogen H 11063
Nitrogen N 1945
Oxygen O 2206
Sulfur S 58
Formula:
C6948H11063N1945O2206S58
Total number of atoms: 22220
Extinction coefficients:
Extinction coefficients are in units of M-1 cm-1, at 280 nm measured in water.
Ext. coefficient 102035
Abs 0.1% (=1 g/l) 0.642, assuming all pairs of Cys residues form cystines
Ext. coefficient 100160

Abs 0.1% (=1 g/l) 0.630, assuming all Cys residues are reduced
Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).
Instability index:
The instability index (II) is computed to be 44.57
This classifies the protein as unstable.
Aliphatic index: 73.83
Grand average of hydropath city (GRAVY): -0.602
Inference:
It can be inferred from the above result that the protein is a hydrophilic molecule present on the surface since the GRAVY value is negative. It is a stable protein. Number of amino acids: 1417 Molecular weight: 159000.4, Theoretical pI: 7.33, Aliphatic index: 73.83 and it also shows atomic compound.

HNN:-

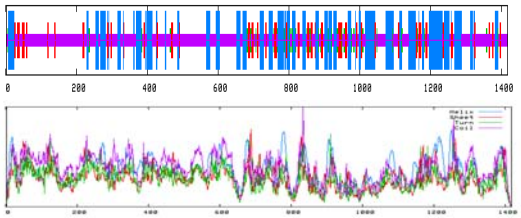
Hierarchical Neural Network result for: UNK_102410
Sequence length: 1417
Alpha helix (Hh): 416 is 29.36%
310 helix (Gg): 0 is 0.00%
Pi helix (Ii): 0 is 0.00%
Beta bridge (Bb): 0 is 0.00%
Extended strand (Ee): 227 is 16.02%
Beta turn (Tt): 0 is 0.00%
Bend region (Ss): 0 is 0.00%
Other states 0 is 0.00%



Inference:
The result of HNN shows that it contain high amount of alpha helix i.e. 29.36%, there is low amount of beta sheet, with an amino acid sequence of 1417, and it is also hydrophilic in nature.

SOPMA

SOPMA result for: UNK 103440
Sequence length : 1417
Alpha helix (Hh) : 434 is 30.63%
310 helix (Gg): 0 is 0.00%
Pi helix (Ii): 0 is 0.00%
Beta bridge (Bb): 0 is 0.00%
Extended strand (Ee): 160 is 11.29%
Beta turn (Tt): 56 is 3.95%
Bend region (Ss): 0 is 0.00%
Random coil (Cc): 767 is 54.13%
Ambigous states (?): 0 is 0.00%
Other states: 0 is 0.00%

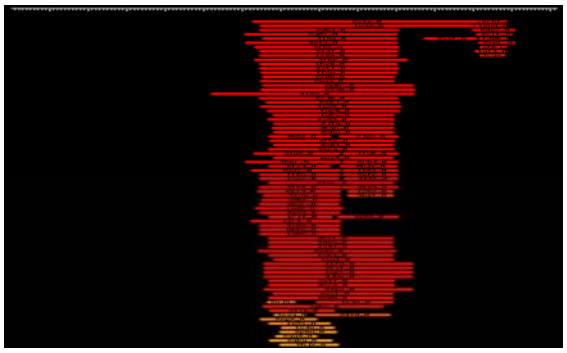


Parameters:
Window width: 17
Similarity threshold: 8
Number of states: 4
Inference:
The result of SOPMA shows that the protein contains 30.63% of alpha helix, with an amino acid sequence of 1417 residues

HHPRED:-

No Hit Prob E-value P-value ScoreSS ColsQueryHMM Template HMM
1 2v1x_A ATP-dependent DNA helic 100.0 0 0 967.4 51.3 561 627-1214 4-566 (591)
2 1oyw_A RECQ helicase, ATP-depe 100.0 0 0 916.3 52.6 513 645-1200 3-521 (523)
3 2db3_A ATP-dependent RNA helic 100.0 0 0 491.4 35.7 352 644-1013 58-425(434)

4 2j0s_A ATP-dependent RNA helic 100.0 0 0 468.9 37.5 340 644-1004 39-391 (410)
5 3i5x_A ATP-dependent RNA helic 100.0 0 0 483.0 40.2 336 652-1004 80-457 (563)
6 3fht_A ATP-dependent RNA helic 100.0 0 0 461.1 39.3 352 631-1006 14-389 (412)
7 3eiq_A Eukaryotic initiation f 100.0 0 0 466.0 34.6 382 603-1004 1-395 (414)
8 1hv8_A Putative ATP-dependent 100.0 0 0 452.8 38.3 336 644-1004 8-353 (367)
9 1s2m_A Putative ATP-dependent 100.0 0 0 458.9 36.8 339 648-1004 25-373 (400)
10 2i4i_A ATP-dependent RNA helic 100.0 0 0 460.9 36.1 346 644-1006 17-393 (417)
>2v1x_A ATP-dependent DNA helicase Q1; DNA strand annealing, mismatch repair, nucleotide-binding, DNA-binding, polymorphism, nuclear protein, ATPase; HET: ADP; 2.00A {Homo sapiens}
PDB: 2wwy _A*
Probab=100.00 E-value=0 Score=967.42 Aligned_cols=561 Identities=37%Similarity=0.667Sum_probs=0.0



Inference
From the above result, PDB id was chosen which 2v1x is. It shows the 3D structure of the protein.as it has max. Identities and lesser E value.
PHYRE:- To predict functional residues and GO classification,

Fold Recognition

View Alignments	**SCOP Code**	**View Model**	**E-value**	**BioText**	**Fold/PDB descriptor**	**Superfamily**	**Family**	**(beta-test)**
c2v1xA_ (length:591) 31% i.d.	0	100 %	n/a	**PDB header:**hydrolase	**Chain:** A: **PDB Molecule:**atp-dependent dna helicase q1;	**PDBTitle:**crystal structure of human recq-like dna helicase		n/a
c1oyyA_ (length:523) 32% i.d.	0	100 %	n/a	**PDB header:**hydrolase	**Chain:** A: **PDB Molecule:**atp-dependent dna helicase;	**PDBTitle:**structure of the recq catalytic core bound to atp-gamma-s		n/a

c2j0qB_ (length:410) 15% i.d.	8.7e-41	100 %	n/a	**PDB header:**hydrolase	**Chain:** B: **PDB Molecule:**atp-dependent rna helicase	**PDBTitle:** the crystal structure of the exon junction complex at	n/a
c2z0mA_ (length:337) **16% i.d.**	3.1e-38	100 %	n/a	**PDB header:**rna binding protein	**Chain:** A: **PDB Molecule:**337aa long hypothetical atp-dependent rna	**PDBTitle:**crystal structure of hypothetical atp-dependent rna2 helicase from sulfolobus tokodaii	n/a
c1hv8A_ (length:367) **19% i.d.**	9.4e-38	100 %	n/a	**PDB header:**rna binding protein	**Chain:** A: **PDB Molecule:**putative atp-dependent rna helicase mj0669;	**PDBTitle:**crystal structure of a dead box protein from the2 hyperthermophile methanococcus jannaschii	n/a

Inference:
It is used to validate the PDB ID 2v1x. which is 2v1x because it has low e value and 100% estimate prediction.

CRYSTAL STRUCTURE OF HUMAN:-

RECQ-LIKE DNA HELICASE
DOI:10.2210/pd2v1x/pdb
Molecular Description Hide
Classification Hydrolase
Structure weight 136247.99
Molecule ATP-DEPENDENT DNA HELICASE Q1

INFERENCE: This tool is used for the molecular description ,which is hydrolase,source which should be homosapiens,and it is homosapiens,and ligand chemical component.

Cph model

Searching for template.

Round 0. Hits better than threshold: 0.000010
entry: 2WWY chain: A score: 378 E: 1e-104
entry: 2WWY chain: B score: 378 E: 1e-104
entry: 2V1X chain: A score: 373 E: 1e-102
entry: 2V1X chain: B score: 373 E: 1e-102
entry: 2RRD chain: A score: 195 E: 4e-49
Retrieving template...
Entry: 2wwy
Chain: A
Making profile-profile alignment...
Score: 478.0 bits
Identity: 38.5 %
Inference:
The result shows 3 D structure of protein in the form of PDB IDs. The PDB id with

the maximum score is chosen i.e. 2V1X and viewed in Rasmol.

RASMOL

RasMol Command Line

```
Classification ........ HYDROLASE
Secondary Structure ... PDB Data Records
Database Code ......... 2V1X
Experiment Technique .. X-RAY DIFFRACTION
Number of Chains ...... 6
Number of Groups ...... 1054 (332)
Number of Atoms ....... 8240 (390)
Number of Bonds ....... 83
Number of Helices ..... 49
Number of Strands ..... 34
Number of Turns ....... 0
Number of Bonds ....... 8485

RasMol> select all
8630 atoms selected!
RasMol> select helices
4245 atoms selected!
RasMol> color red
RasMol> select sheet
1514 atoms selected!
RasMol> color blue
RasMol> select his
246 atoms selected!
RasMol> color yellow
Warning: Unable to allocate shade!
RasMol> select ala
390 atoms selected!
RasMol> color green
RasMol> select coil
No atoms selected!
RasMol> select loop
No atoms selected!
```

Inference:

RASMOL is a visualization tool. The result shows the 3D structure of the protein. Here No. Of alpha helices are more as compare to beta sheet. This protein is hydrophilic in nature.

PDB:-

RecQ-like helicases, which include 5 members in the human genome, areimportant in maintaining genome integrity. We present a crystal structure of a truncated form of the human RECQ1 protein with Mg-ADP. The truncatedprotein is active in DNA fork unwinding but lacks other activities of the full-length enzyme: disruption of Holliday junctions and DNA strand annealing.The structure of human RECQ1 resembles that of Escherichia coli RecQ, withsome important differences. All structural domains are conserved, includingthe 2 RecA-like domains and the RecQ-specific zinc-binding and winged-helix(WH) domains. However, the WH domain is positioned at a differentorientation from that of the E. coli enzyme. We identify a prominent beta-hairpin of the WH domain as essential for DNA strand separation, which maybe analogous to DNA strand-separation features of other DNA helicases. Thishairpin is significantly shorter in the E. coli enzyme and is not required for itshelicase activity, suggesting that there are significant differences betweenthe modes of action of RecQ family members. **Keywords:** Adenosine Diphosphate, Adenosine Triphosphate, Amino Acid Motifs, Amino Acid Sequence, Conserved Sequence, DNA, Escherichia coli, Humans, Kinetics,Molecular Sequence Data, Mutant Proteins, Protein Binding, Protein Structure,Tertiary, RecQ Helicases, Sequence Alignment, Zinc **Related Structures:** Also Cited By: **2WWY**

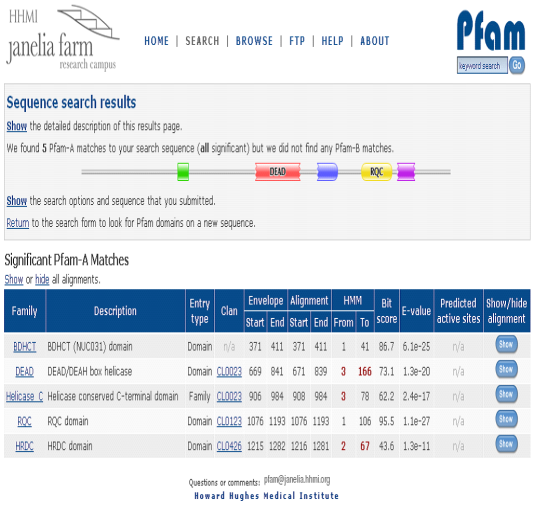
ORGANIZATIONAL AFFILIATION: Structural Genomics Consortium, Old Road Campus Research Building, Roosevelt Drive, University of Oxford, Oxford OX3 7DQ, United Kingdom.

PRO-DOM:-

Position Pro-Dom domain Score E value
571-660 #PD358871 472 2e-46
PD358871 (Closest domain: BLM_HUMAN 571-660)
Number of domains in family: 4
Commentary (automatic):
SYNDROME NUCLEOTIDE-BINDING DNA ATP-BINDING HELICASE DNA-BINDING PHOSPHORYLATION NUCLEAR BLOOM 3.6.1.-

Length = 90
Score = 472(186.4 bits), Expect = 2e-46,Identities = 90/90 (100%),Positives = 90/90 (100%)
INFERENCE: I have chosen domain PD358871, since this is the domain which has less e value and less amino acids.

Pfam result page



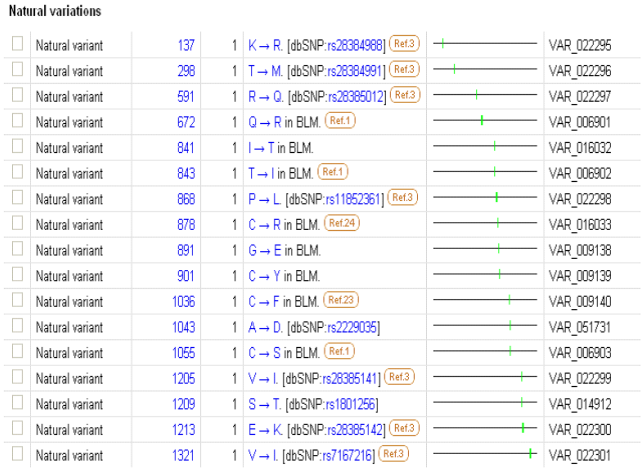
Inference:
The “pfam” tool is used to know the family and function of the proteins using the amino acid sequence.
The BLM protein belongs to the DEAD/DEAH box helicase, RQC domain, Helicase conserved C-terminal domain and
It is involved in the ATP binding, nucleic acid binding, DNA repair function.

Gene Functional Similarity Search Tool (GFSST)

GFSST is tool by which the help of we can get functional information with in gene ontology, the id present in the pfam were matched with the gfsst. Hence the id are present in the gfsst, thus, it’s near to the cancer.
Similar genes and gene products (protein) west on gene ontology is given by GFFST.

Cellular function in a biological function. Normally involve participation, multiplication of gene. In case of bloom syndrome many gene are involved but BLM gene is prominent in 100% of bloom syndrome disease.
Mutation that alter the function of BLM gene can protein cell increase the disease susctivility here by GFFST 3 things are highlighted on the gene invirolment
1. DNA Damage response
2. DNA Repair
3. Apoptosis (Cell Death)

Point Mutation



Involvement in disease:-

Defects in BLM are the cause of Bloom syndrome (BLM) [MIM: 210900]. BLM is an autosomal recessive disorder characterized by proportionate pre- and postnatal growth deficiency, sun-sensitive telangiectatic hypo- and hyper pigmented skin, predisposition to malignancy, and chromosomal instability.

Ligand Designing

Pub Chem:

AIM: To find the new drug and to work upon them.
According to this parameter, only the Tripetide S was matched and the rest all does not showed
Any resemblance .

ISSN 2348 – 7968

Ideal chemical compound parameters which we can use as a Drug .
è Molecular weight: Min 200 Max. 600 Other wise toxic
è Hydrogen Donor: Min. 0 Max. 6 Other wise toxic
è Hydrogen Acceptor : Min 0 Max 12 Other wise toxic
è Flexible bonds: Min 0 Max 15 Other wise toxic
è Total charge : Min 2 Max 2 Other wise toxic
è Log p : Min 2 Max 6

From the drug bank we found that, the drug which matched to the ID was omitted b/c we have to keep or design the new drug.

Tripeptide S Compound ID	9543465
Molecular Weight	458.64162 [g/mol]
Molecular Formula	C18H28N5O3S3+
XLogP3-AA	0.8
H-Bond Donor	4
H-Bond Acceptor	6

Canonical SMILES:

CC(C(C(=O)NCCC1=NC(=CS1)C2=NC(=CS2)C(=O)NCCC[S+](C)C)N)O Molecular Weight	458.64162 [g/mol]
Molecular Formula	C18H28N5O3S3+
XLogP3-AA	0.8
H-Bond Donor	4
H-Bond Acceptor	6
Rotatable Bond Count	11
Tautomer Count	8
Exact Mass	458.135427
Mono Isotopic Mass	458.135427
Topological Polar Surface Area	188
Heavy Atom Count	29
Formal Charge	1
Complexity	543
Isotope Atom Count	0
Defined Atom Stereo Center Count	2
Undefined Atom Stereo Center Count	0
Defined Bond Stereo Center Count	0
Undefined Bond Stereo Center Count	0
Covalently-Bonded Unit Count	1

Mobile ADME

C(c1n[c]([c]2nc([s]c2)CCNC(=O)C(C(O)C)N)[s]c1)(=O)NCCC[S+](C)C Toxic. PSA: 189.410000
List of compounds passing the filter: None
Detailled output for each compound:
MW: Molecular weight
Drs: Donors
Ars: Acceptors
FB: flexible bonds
RB: Rigid Bonds
#R: Ring Number
RL: Ring Length
C: Carbons
nC: non carbons
C/nC: ratio non carbons/carbons
#Chrg: number of charges
Chrg: Total Charge
LogP: logP (octanol / water)
PSA: Polar surface area
C(c1n[c]([c]2nc([s]c2)CCNC(=O)C(C(O)C)N)[s]c1)(=O)NCCC[S+](C)C : MW : 458.4 Drs : 4 Ars : 8 FB : 11 RB : 14 #R : 2 RL : 5 C : 18 nC : 11 C/nC : 0.611111 #Chrg : 1 Chrg : 1 LogP : -0.620000 PSA : 189.410000 104

PRODRG (Dundee PRODRG Server):

PRODRG> MOL mode detected.
PRODRG> WARNING: deleted hydrogen(s) from your input.
PRODRG> Molecule complexity index: 2.07.
PRODRG> 6 explicit hydrogen(s) added.
PRODRG> 36 bonds 3 ambiguous

ISSN 2348 – 7968

PRODRG> 51 bond angles 12 ambiguous
PRODRG> 22 improper dihedrals 2 ambiguous
PRODRG> 15 dihedrals 1 ambiguous
PRODRG> 17 partial charges 2 ambiguous
PRODRG> Net charge on molecule: 2.000
From the drug bank we found that, the drug which matched to the id, was omitted, because, we have to keep or design the new drug.

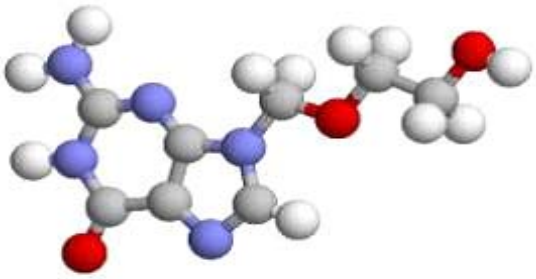
Lipinski´s Rule

In 1997 Christopher Lipinski from Pfizer found a simple mnemonic which he called the "Rule of 5" because the parameter cut-off values all contained 5's. Numerically there actually are only four rules. This was a major breakthrough for the chemInformatic society. The Rule of 5 stated:

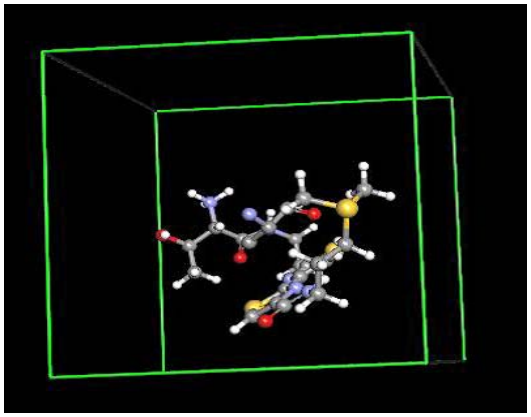
Poor absorption or permeation are more likely when there are: **1) More than 5 H-bond donors 2) The molecular weight is over 500 3) The CLog P is over 5 (or MLOGP is over 4.15) 4) The sum of N's and O's is over 10** Substrates for transporters and natural products are exceptions.

The rule is useful but it has some limitations. Some of today's blockbusters fail by using that rule. It is too simple to detect all patterns of drugs. For further development's, see expert systems for selecting molecules.

Example: Acyclovir H-donors: 4 MW: 225.21 MLOGP: -0.09 (exp. LogP: -1.56) H-acceptors : 8 Acyclovir would pass the Rule of 5.

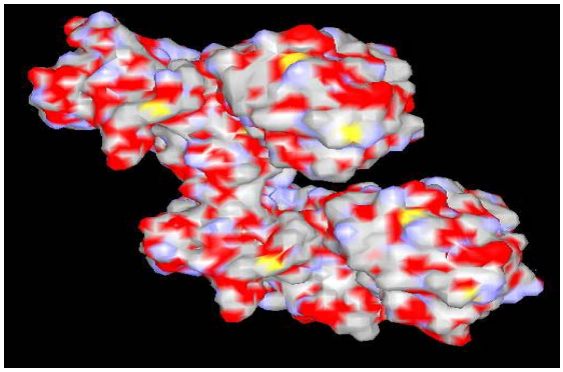


DOCKING IN ARGUS



INFERENCE: Docking is done so that the drug may find the site or the surface to bind the protein

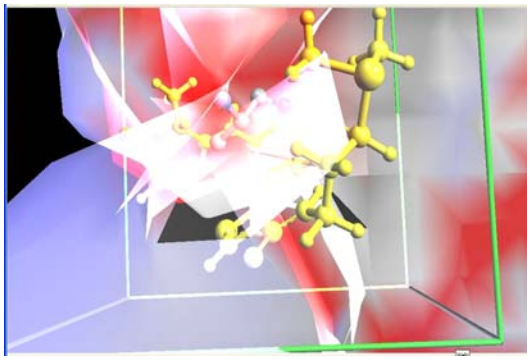
SOLVENT ACCEESSIBLE SURFACE



INFERENCE: This tool shows the binding of the compound, or the drug to the protein

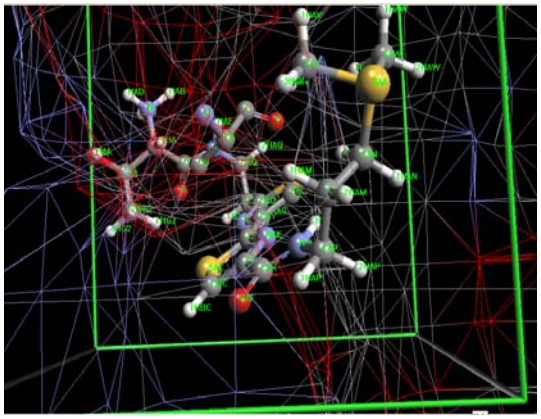
TRANSLUCENT:

ISSN 2348 – 7968



INFERENCE: Translucent is done to see the position of the ligand, where the protein gets bind

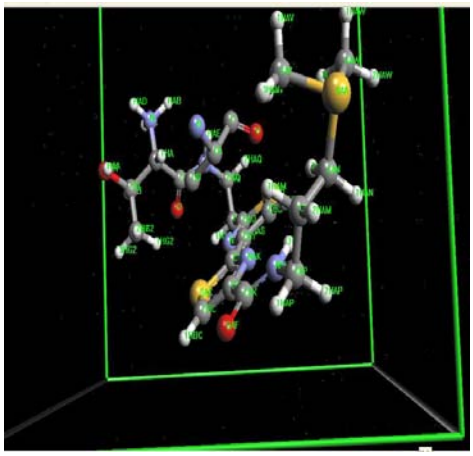
MESH



INFERENCE

The red color of grid are of carbon bonds, blue color grid of nitrogen and gray color grid are oxygen

DOT



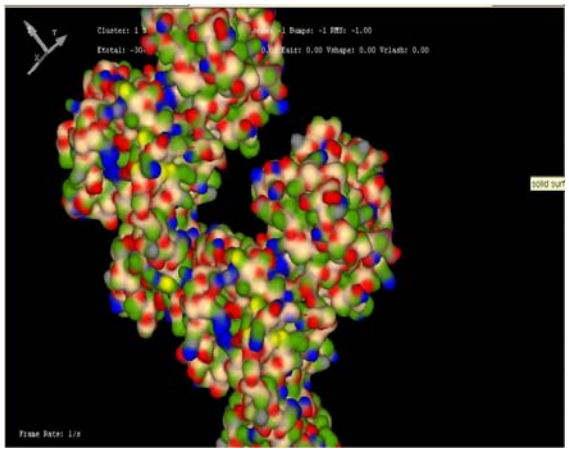
INFERENCE: The structure above shows the structure of the compound, in which red color dots shows presence of bonds, the blue color shows the presence of nitrogen, as well as the white color shows the presence of oxygen.

HEX 5.1
STARTING ENERGY
10 1989 000:000 -222.7 -222.7 0.0 0.0 0.0 0.0 -1 -1.00
10 1990 000:000 -222.7 -222.7 0.0 0.0 0.0 0.0 -1 -1.00
10 1991 000:000 -222.7 -222.7 0.0 0.0 0.0 0.0 -1 -1.00
10 1992 000:000 -222.7 -222.7 0.0 0.0 0.0 0.0 -1 -1.00
10 1993 000:000 -222.7 -222.7 0.0 0.0 0.0 0.0 -1 -1.00
10 1994 000:000 -222.7 -222.7 0.0 0.0 0.0 0.0 -1 -1.00
10 1995 000:000 -222.6 -222.6 0.0 0.0 0.0 0.0 -1 -1.00
10 1996 000:000 -222.6 -222.6 0.0 0.0 0.0 0.0 -1 -1.00
10 1997 000:000 -222.6 -222.6 0.0 0.0 0.0 0.0 -1 -1.00
10 1998 000:000 -222.6 -222.6 0.0 0.0 0.0 0.0 -1 -1.00
10 1999 000:000 -222.6 -222.6 0.0 0.0 0.0 0.0 -1 -1.00
10 2000 000:000 -222.6 -222.6 0.0 0.0 0.0 0.0 -1 -1.00
END ENERGY
1 1 000:000 -304.2 -304.2 0.0 0.0 0.0 0.0 -1 -1.00
1 2 000:000 -302.9 -302.9 0.0 0.0 0.0 0.0 -1 -1.00
1 3 000:000 -301.4 -301.4 0.0 0.0 0.0 0.0 -1 -1.00
1 4 000:000 -299.9 -299.9 0.0 0.0 0.0 0.0 -1 -1.00
1 5 000:000 -294.7 -294.7 0.0 0.0 0.0 0.0 -1 -1.00
1 6 000:000 -294.1 -294.1 0.0 0.0 0.0 0.0 -1 -1.00

ISSN 2348 – 7968

1 7 000:000 -293.4 -293.4 0.0 0.0 0.0 0.0 -1 -1.00
1 8 000:000 -290.8 -290.8 0.0 0.0 0.0 0.0 -1 -1.00
1 9 000:000 -290.6 -290.6 0.0 0.0 0.0 0.0 -1 -1.00
1 10 000:000 -288.6 -288.6 0.0 0.0 0.0 0.0 -1 -1.00

SOLID SURFACE



INFERENCE

Target protein inhibit among one of the ligand to from the least energy with negative score that may be indicate their stability. This ligand would be the better drug for further studies.

SIDE CHINS

INFERENCE

The picture above shows the ligand binds to the receptor

CONCLUSION

Leukemia is a cancer of the bone marrow and blood and is the most common type of childhood cancer. The bone marrow is the soft, spongy center of the bones and produces the three types of blood cells: white blood cells that fight infection, red blood cells that carry oxygen and platelets that help with blood clotting and bleeding.When a child has leukemia, the bone marrow begins to make blood cells that dont mature properly. The immature cells continue to reproduce, crowding out the healthy ones. Normal, healthy cells reproduce only when they have sufficient space. by the approach of the bioinformatics, I have been step ahead to prepare the drug which is TRIPEPTIDE-S. by the help of tools. Blast, pubmed, pdb, clustal-w,clustal distance matrix, boxshade, prot param, hnn, sopma, cph model, hhpred, phyre along with some of the confirmatory tool, KEGG, gene card, pfam, prodom, prosite, Rasmol, and for the Drug Designing, ArgusLab, Drug bank, pubchem, through which we got the appropriate result. And thus the drug was prepared

REFERENCES

1.Fiancette R, Amin R, Truffinet V, Vincent-Fabert C, Cogné N, Cogné M, Denizot Y. UMR CNRS 6101, Centre National de la Recherche Scientifique, Université de Limoges, 2 rue Dr. Marcland, 87025 Limoges, France. **A myeloma translocation-like model associating CCND1 with the immunoglobulin heavy-chain locus 3′ enhancers does not promote by itself B-cell malignancies**

2.Fabiani E, D'Alò F, Scardocci A, Greco M, Di Ruscio A, Criscuolo M, Fianchi L, Pagano L, Hohaus S, Leone G, Voso MT.

Istituto di Ematologia, Università Cattolica del Sacro Cuore, L go A Gemelli, Rome, Italy **Polymorphisms of detoxification and DNA repair enzymes in myelodyplastic syndromes**

3. Han S, Lee KM, Park SK, Lee JE, Ahn HS, Shin HY, Kang HJ, Koo HH, Seo JJ, Choi JE, Ahn YO, Kang D. Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea. **Genome-wide association study of childhood acute lymphoblastic leukemia in Korea**

4. Huret JL, Schoenwald M, Brizard A, Guilhot F, Vilmer E, Tanzer J.
Département d'Hématologie et Oncologie Médicale, CNRS URA 223, Hôpital Jean Bernard, Poitiers, France. **Chromosome 6p rearrangements appear to be secondary changes in various haematological malignancies**

5. Janson CH, Tehrani M, Wigzell H, Mellstedt H.
Department of Oncology, Karolinska Hospital, Stockholm, Sweden. **Rational use of biological response modifiers in hematological malignancies—A review of treatment with interferon, cytotoxic cells and antibodies**

6. Ramanarayanan J, Dunford LM, Baer MR, Sait SN, Lawrence W, McCarthy PL.
Department of Medicine, Roswell Park Cancer Institute, State University of New York, Buffalo, NY 14263, USA. **Chronic myeloid leukemia after treatment of lymphoidmalignancies: Response to imatinib mesylate and favorable outcomes in three patients**

7. Ribatti D, Crivellato E, Molica S.
Department of Human Anatomy and Histology, University of Bari Medical School, Piazza G. Cesare, 11 Policlinico, 70124 Bari, Italy. ribatti@anatomia.uniba.it **Mast cells and angiogenesis in haematologicalmalignancies**

8. Specchia G, Buquicchio C, Albano F, Liso A, Pannunzio A, Mestice A, Rizzi R, Pastore D, Liso V. Department of Haematology, University of Bari, Policlinico, Piazza Giulio Cesare 11, 70124 Bari, Italy. emadhba@cimedoc.uniba.it **Non-treatment-related chronic myeloid leukemia as a second malignancy**

9. Tumer TB, Yilmaz D, Tanrikut C, Sahin G, Ulusoy G, Arinç E. Department of Biological Sciences, Middle East Technical University, 06531 Ankara, Turkey. tumertb@gmail.com **DNA repair XRCC1 Arg399Gln polymorphism alone, and in combination with CYP2E1 polymorphisms significantly contribute to the risk of development of childhood acute lymphoblastic leukemia**

10.Warren M, Chung YJ, Howat WJ, Harrison H, McGinnis R, Hao X, McCafferty J, Fredrickson TN, Bradley A, Morse HC 3rd.
Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. mvw@pathologydiagnostics.com
Irradiated *Blm*-deficient mice are a highly tumor prone model for analysis of a broad spectrum of hematologic malignancies.