Clinical Proteomics & Bioinformatics

Research Article ISSN: 2398-3124

In silico: analysis of retinoblastoma gene & novel drug designing

Farhina Pasha*

Department of Biology, Faculty of science, University of Tabuk, Saudi Arabia

Abstract

Data mining has emerged as a very powerful tool to extract information. In the present study RB1 gene, which is a tumor suppressor gene has been studied. Data mining is done first at Ground Level Mining, in which relevant data sets are collected and then reduced to the minimum size possible through statistical representation. Chromosome Report is studied to determine the chromosome location of the gene under study. It regulates cell cycle as a check point for p53 and for other genes as well, specifying cell fate. The mRNA Report gives detailed information regarding the gene for its identification and transcript sequences. The Peptide Report reveals highly descriptive data regarding protein, SNP regions, position and alleles. A special emphasis is given to protein interactions and blastp result for study of homologous protein sequences. Phylogenetic tree is generated using homologous protein sequences for Phylogenetic Inference and homology with other taxa. The preclinical data of cyclosporine has suggested that it has a significant role in treating retinoblastoma malignancies but the binding of Actinomycin-D with Rb is far better as evident by the low e-total value signifying its greater affinity. Therefore the drug is suggested as a potential drug for future in treatment of retinoblastoma and associated retinoblastoma malignancies.

Introduction

RB1, the first tumor suppressor gene to be discovered has a crucial role in cell cycle as a major check point. More specifically, Rb family members have overlapping roles in promoting cellular growth arrest. Targeted inactivation of all the three RB-related genes in embryonic stem cells causes deregulated G1/S transition, G1 arrest and cellular immortalization. Most noteworthy is the fact that Rb is the only pocket protein known to exhibit features of a bonafide tumor suppressor. Accordingly, RB heterozygous mice are predisposed to the onset of pituitary and thyroid cancer and deregulation of the Rb signaling pathway or inactivation of RB itself is a hallmark of nearly all human tumors as reported by Hanahan *et al.* [1].

In addition to Rb is required for myogenesis p300, which is a regulatory factor for myogenic transcription and complete blood particulates (CBP) are in fact known to regulate the activity of numerous transcription factors [2]. It possesses conserved functions in cell growth, transformation and development as stated by Goodman & Smolik (2000). P/CAF has likewise known to play a role in cell differentiation and cause tumorigenesis [3,4]. Current models propose that p300/CBP and P/CAF can form a multimeric complex with MyoD, which recruits these co-activators to muscle-specific promoters [3,5,6]. P/CAF and p300/CBP are believed to hyperacetylate the surrounding nucleosomes, thus increasing the accessibility of additional transcription factors to MyoD target promoters, [7]. The proteins p300/CBP and P/CAF have now been demonstrated to acetylate a wide variety of non-histone proteins, including many DNA-binding proteins, also general transcription factors, cytoplasmic proteins and tumor suppressors [8]. Recently, the co-activator p300 was shown to acetylate pRb in vitro [9].

Retinoblastoma is a rapidly developing cancer which develops in the cells of the retina, the light detecting tissues of the eye. In the developed world, more than nine out of every ten sufferers survive into adulthood with cure to retinoblastoma [10]. They also described two forms of the disease; a genetic, heritable form and a non-genetic, non-heritable form. Approximately 55% of children with retinoblastoma have the non-genetic form. If there is no history of the disease within the family, the disease is labeled "sporadic", but this does not necessarily indicate that it is the non-genetic form. In humans, the protein is encoded by the RB1 gene located on 13q arm 14.1-q14.2 band position. It is a 4840 bp long gene having 54 K Single Nucleotide Polymorphism. If both alleles of this gene are mutated early in life, the protein is inactivated and results in development of retinoblastoma cancer, hence the name Rb. Rb is an associated protein, which is 928 residues long having 106,159.11 Da molecular weight. Its isoelectric point is 8.04. Rb has two domains: Domain A and Domain B separated by a spacer and has a pocket region that refers to the region where regulatory factors like E1A binds.

Material and methods

Data mining

It uses computational techniques from statistics, machine learning and pattern recognition. These techniques are an automated means of reducing the complexity of data in large bioinformatics databases and of discovering meaningful, useful patterns and relationship in data. It included the following steps**:** *Data Characterization*, *Consistency Analysis, Domain Analysis*, *Data Enrichment, Frequency and Distribution Analysis*, *Normalization* and *Missing Value Analysis*

Correspondence to: Dr. Farhina Pasha, Department of Biology, Faculty of science, University of Tabuk, Saudi Arabia, Tel: +966 44273055; Fax: +966 44251124; E-mail: fbasha@ut.edu.sa

Key words: *retinoblastoma, Actinomycin-D, Data mining, RB1 gene*

Received: September 02, 2016; **Accepted:** October 07, 2016; **Published:** October 13, 2016

Data mining methods

The process of data mining is concerned with extracting patterns from the data by using techniques such as: Classification which involves mapping of data into one of several predefined or newly discovered classes.

Evaluation

The patterns identified by the data mining analysis are interpreted and typical evaluation ranges from simple statistical analysis and complex numerical analysis of sequences and structures to determining the clinical relevance of the findings.

Visualization: Visualization of evaluation results can range from simple pie charts to 3-D virtual reality displays that can be manipulated by haptic (force feedback) controllers.

Multiple Sequence alignments of protein sequences are important tools in studying sequences. The basic information they provide is identification of conserved sequence regions. Sequences can be aligned across their entire length (global alignment) or only in certain regions (local alignment). Global alignments need to use gaps (representing insertions/deletions) while local alignments can avoid them, aligning regions between gaps. CLASTALW is a fully automatic program for global multiple alignment of DNA and protein sequences. The alignment is progressive and considers the sequence redundancy.

Phylogenetic Tree can also be calculated from multiple alignments. Evolutionary relationships can be seen by viewing Cladogram and Phylogram.

Docking is the process of calculating the e-value, which is the pairing of hydrogen bonds of the protein and the drug molecules, so as to determine their active site affinity. Lower the e-value stronger is the bonding. "Hex" is an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to predict the affinity and activity of the small molecule. Hence docking plays an important role in rational drug designing and therefore is of biological and pharmaceutical significance.

Tools

The various tools and their URL's are:

Basic Local Alignment Search Tool (BLAST): URL: www.ncbi.nih.nlm.gov/blast

CLUSTAL W: URL: http://www.ebi.ac.uk/clustalw , URL: http://www.align.genome.jp/

Open Reading Frame Finder (ORF): URL: http://www.ncbi.nlm.nih.gov/gorf/gorf.html

Resources

The various data bases and software's used through NCBI are as under with their unique location addresses as URL's: (Box 1)

Box 1

S.No:	Data Base	URL
1	Map View	http://www.ncbi.nlm.nih.gov/mapview
2	Pub Med	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed
3	Ensembl	http://www.ensembl.org/
4	Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/
5	Universal Protein Resource (UniProt)	http://www.ebi.uniprot.org/
6	Swissprot	http://www.ebi.ac.uk/swissprot/
7	Pfam	http://www.sanger.ac.uk/Software/Pfam/
8	Protein Data Bank (PDB)	http://www.pdb.org/

Methodology

Ground level mining

To perform ground level mining, log on to www.ncbi.nlm.nih.gov/mapview, search *Homo sapiens* for RB1. Procure genomic view and query results for RB1 of *Homo sapiens*. Go to RB1 for Celera gene, which is the references verified by the NCBI and is supported by Pub Med sequences. Obtain Master Map by clicking on Genes RB1 highlighted on Celera. Click on the highlighted RB1 in master map and collect detailed information on Summary, Functions and Genomic Regions. Gene Table is opended to get the number of exons and introns. Taxonomic Links are clicked to get taxonomic IDs and SNP Links for nucleotide information. Unigene IDs, Markers, Gene & genotype Links and pathways are downloaded from the main page. Note the references of abstracts and interactions from corresponding PubMed links. Figures and information about metabolism is extracted from Kyoto encyclopedia of gene and genome (KEGG) pathways and Reactome event from the respective links.

Chromosome report

Log on to www.ensembl.org and click on *Homo sapiens* icon. Search gene for *RB1* and select the gene under study.

Gene Report is taken from side menu of NCBI main page. The transcript sequence is taken directly from transcript information of Ensembl. The details regarding exons and introns are collected from Gene Table of NCBI. ORF Finder tool is used to predict ORFs. List of SNPs are obtained by integrating information from Ensembl and NCBI. The transcript structure is again collected from transcript info of Ensembl. Orthologues prediction table is directly procured from Ensembl.

Peptide report

The description, function & feature table about Rb is collected from Uniprot. Peptide statistic and sequence are extracted from Ensembl. Interactions are studied in detail by referring to BIND and PubMed Links. http://www.pdb.org/explore.do?structureId=2AZE.

Phylogenetic inference

Multiple sequence alignment is carried out using ClustalW to predict RB homologous sequences in different organisms and to generate phylogenetic tree such as Cladogram, Phylogram and Dendrogram.

Docking

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second and is performed as under: Rb image is downloaded as 'jpg' image. Drug structures are downloaded as crystalline structures. Protein and drugs are subjected to docking via Hex software, to obtain the e-values.

Results

Genome view

RB1 is located on 13th chromosome as visualized in the above karyotype (Figure 1).

Ideogram

The details of 13th chromosome are illustrated as under, with the retinoblastoma affected sequence highlighted as a red band (Figure 2).

Map 1: Celera Genes On Sequence (Celera) Table View		
Region Displayed		29,800K-30,240K bp
Total Celera Genes On Chromosome		718
Celera Genes Labelled		9
Total Celera Genes in Region		9
Map 2: Homo sapiens UniGene Clusters (Celera) Table View		
Region Displayed: 29,800K-30,240K bp		29,800K-30,240K bp
Total Transcript alignments On Chromosome		3172
UniGene Clusters Labelled		25
Total Transcript alignments in Region		39
Histogram Data	Tick Width	597bp/pixel
	Max Height	846 transcripts
Map 3: Genes On Sequence (Celera) Table View		
Region Displayed		29,800K-30,240K bp
Total Genes On Chromosome		642
Genes Labelled		5
Total Genes in Region		5

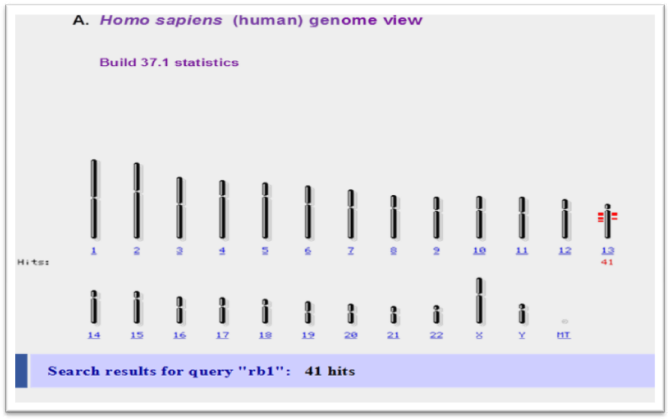


Figure 1. Homo sapiens (Human) genome view.

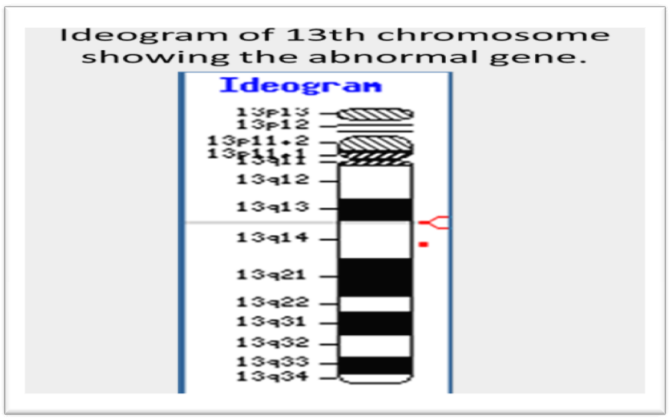


Figure 2. Ideogram of 13th chromosome showing the abnormal gene.

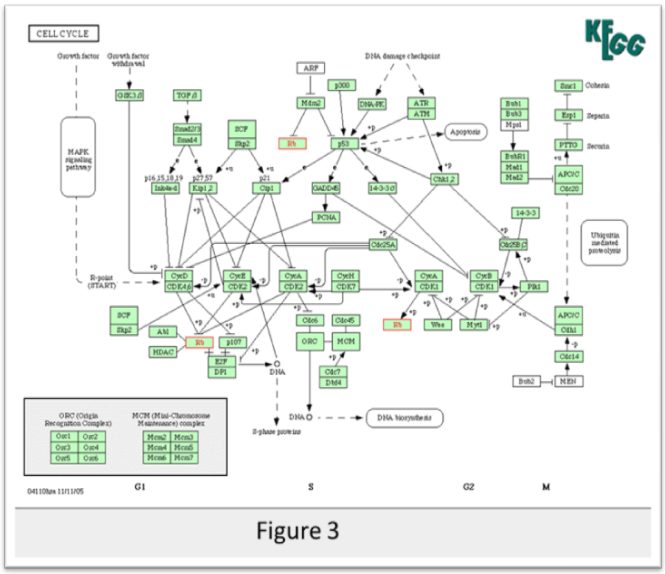


Figure 3. Metabolic pathway shows the check point p53 and its associated pathways.

Gene report

GENE DESCRIPTION-The RB1 regulates proliferation, cell fate specification and differentiation in the all cell types. It inhibits cell proliferation in the absence of Rb which is the expression part of RB1 and is linked to be the main cause of retinoblastoma, as the role of Rb in normal cells is to suppress tumor formation. Rb is found in all cells of the body, under normal conditions it acts as a brake on the cell division cycle by preventing certain regulatory proteins from triggering DNA replication. If Rb is missing a cell can replicate itself over and over in an uncontrolled manner, resulting in tumor formation (Figure 4).

Peptide report

Peptide sequence (Figure 5).

Description

1.1. Protein Name: Retinoblastoma-associated protein

1.2. Synonyms: (PP110) (P105-RB) (RB).

1.3. Protein ID Uniprot : P06400

1.4. No. of residues: 928 aa

1.5. Molecular weight: 106,159.11 Daltons

Phylogenetic inference

Homology search is performed for RB1 with entrez and generation of Phylogenetic Tree for RB1as Dendrogram (Figure 6).

Docking

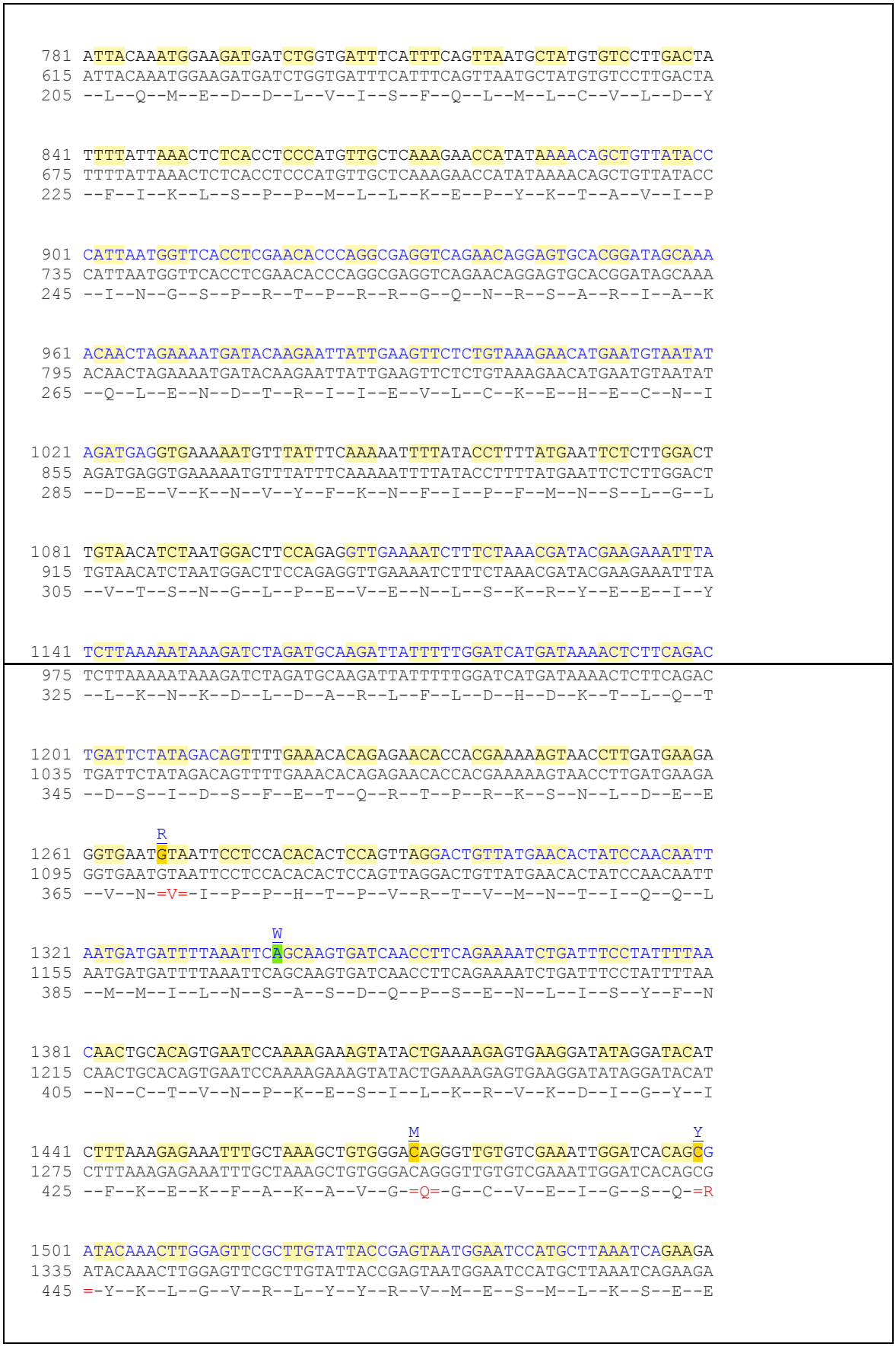
The affinity between the Rb protein and drugs is seen in docking and the bond strength is calculated as e-value. Docking is done with 'hex' inserting Rb as receptor and drug as ligand in the software. The result is given below: Box 3

a. Binding of cyclosporine with Rb as shown in Figure 7

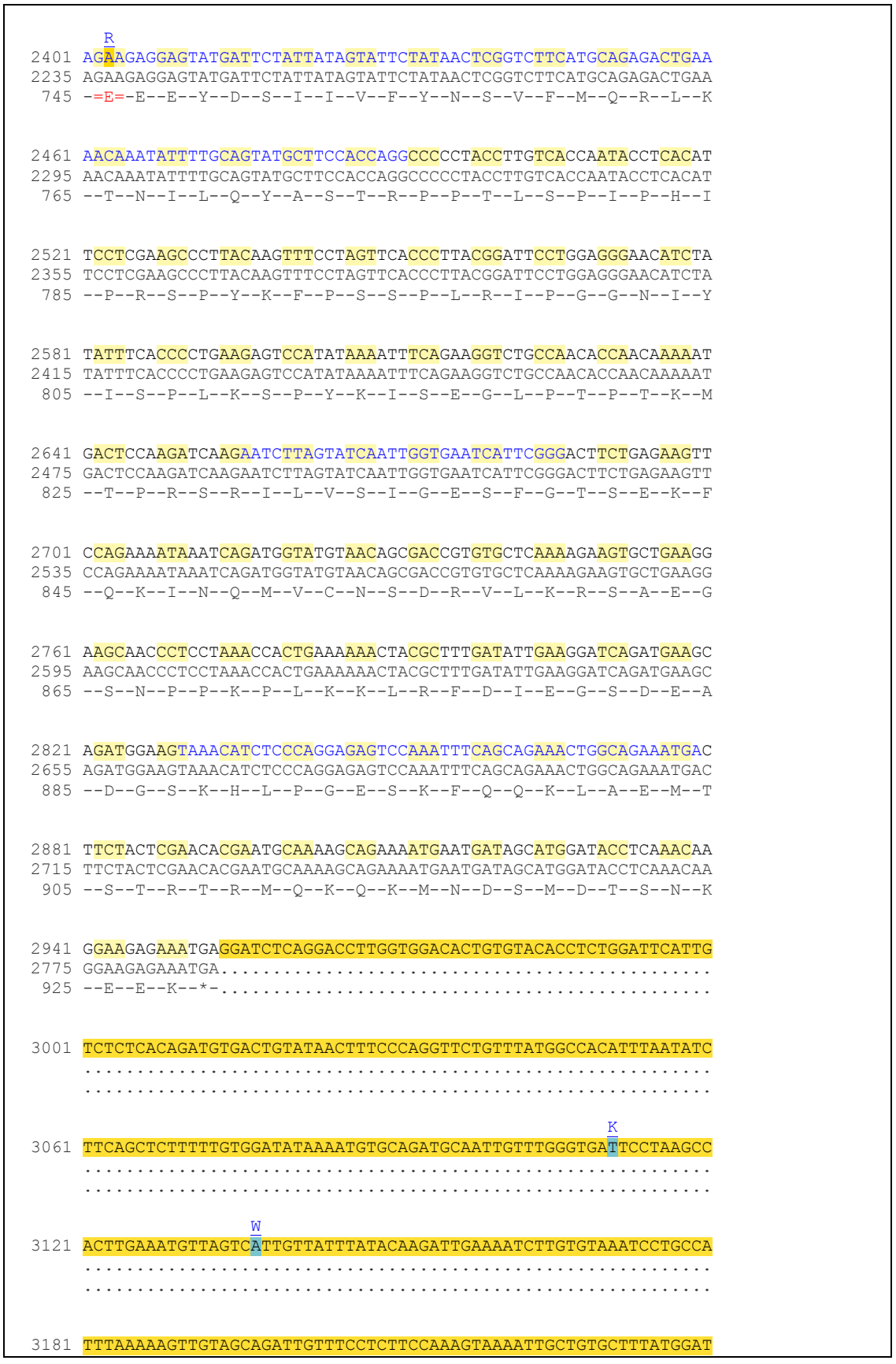
b. Binding of actinomycin-D with Rb as shown in Figure 8

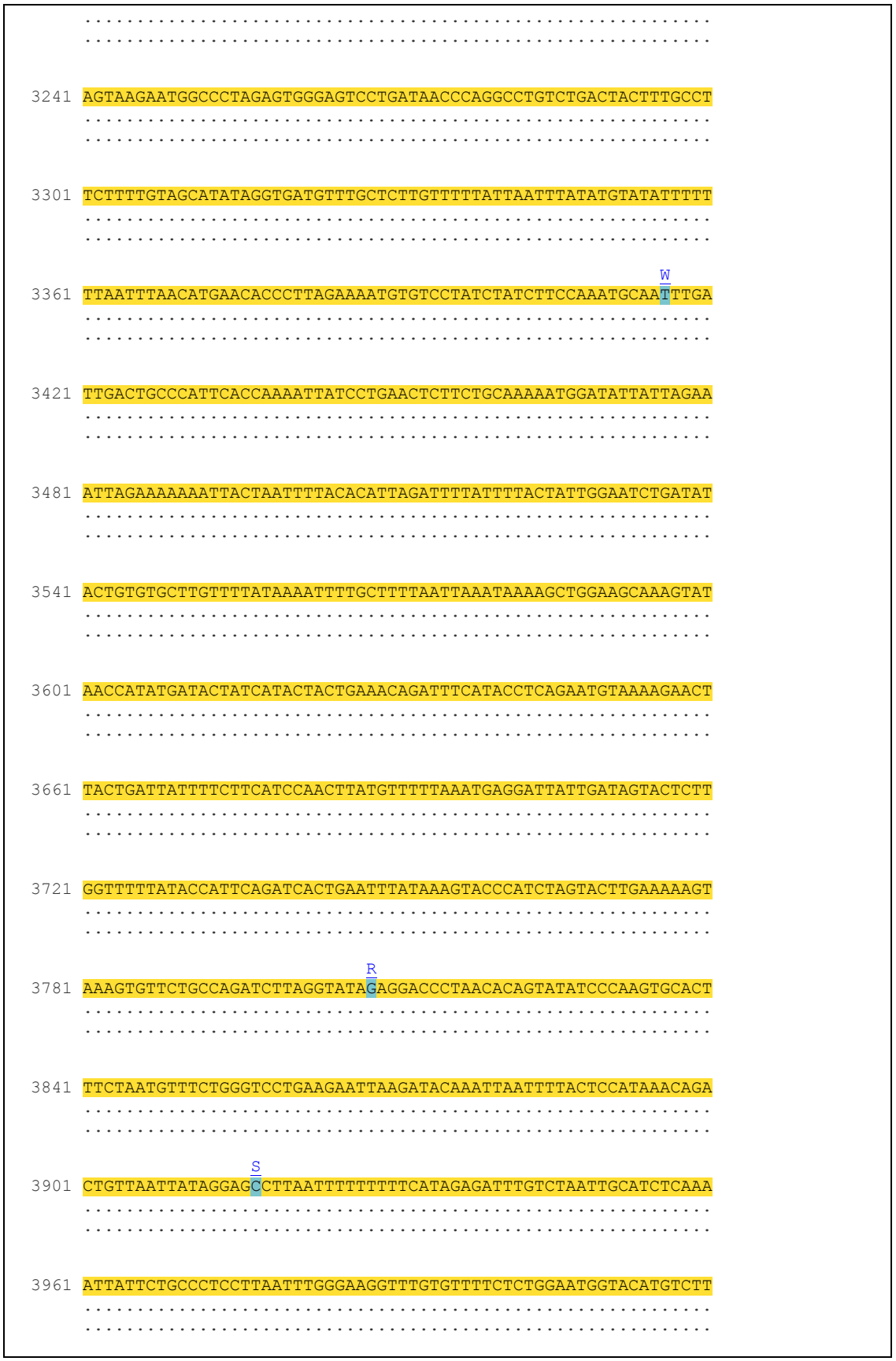
Box 3


1 GCTCAGTTGCCGGGCGGGGGAGGGCGCGTCCGGTTTTTCTCAGGGGACGTTGAAATTATT
61 TTTGTAACGGGAGTCGGGAGAGGACGGGGCGTGCCCCGACGTGCGCGCGCGTCGTCCTCC
121 CCGGCGCTCCTCCACAGCTCGCTGGCTCCCGCCGCGGAAAGGCGTCATGCCGCCCAAAAC
ATGCCGCCCAAAAC
-M--P--P--K--T
181 CCCCCGAAAAACGGCCGCCACCGCCGCCGCTGCCGCCGCGGAACCCCCGGCACCGCCGCC
15 CCCCCGAAAAACGGCCGCCACCGCCGCCGCTGCCGCCGCGGAACCCCCGGCACCGCCGCC
5 --P--R--K--T--A--A--T--A--A--A--A--A--E--P--P--A--P--P--P
241 GCCGCCCCCTCCTGAGGAGGACCCAGAGCAGGACAGCGGCCCGGAGGACCTGCCTCTCGT
75 GCCGCCCCCTCCTGAGGAGGACCCAGAGCAGGACAGCGGCCCGGAGGACCTGCCTCTCGT
25 --P--P--P--P--E--E--D--P--E--Q--D--S--G--P--E--D--L--P--L--V
301 CAGGCTTGAGTTTGAAGAAACAGAAGAACCTGATTTTACTGCATTATGTCAGAAATTAAA
135 CAGGCTTGAGTTTGAAGAAACAGAAGAACCTGATTTTACTGCATTATGTCAGAAATTAAA
45 --R--L--E--F--E--E--T--E--E--P--D--F--T--A--L--C--Q--K--L--K
361 GATACCAGATCATGTCAGAGAGAGAGCTTGGTTAACTTGGGAGAAAGTTTCATCTGTGGA
195 GATACCAGATCATGTCAGAGAGAGAGCTTGGTTAACTTGGGAGAAAGTTTCATCTGTGGA
65 --I--P--D--H--V--R--E--R--A--W--L--T--W--E--K--V--S--S--V--D
421 TGGAGTATTGGGAGGTTATATTCAAAAGAAAAAGGAACTGTGGGGAATCTGTATCTTTAT
255 TGGAGTATTGGGAGGTTATATTCAAAAGAAAAAGGAACTGTGGGGAATCTGTATCTTTAT
85 --G--V--L--G--G--Y--I--Q--K--K--E--L--W--G--I--C--I--F--I
481 TGCAGCAGTTGACCTAGATGAGATGTCGTTCACTTTTACTGAGCTACAGAAAAACATAGA
315 TGCAGCAGTTGACCTAGATGAGATGTCGTTCACTTTTACTGAGCTACAGAAAAACATAGA
105 --A--A--V--D--L--D--E--M--S--F--T--F--T--E--L--Q--K--N--I--E
M W
541 AATCAGTGTCCATAAATTCTTTAACTTACTAAAAGAATTGATACCAGTACCAAAGTTGA
375 AATCAGTGTCCATAAATTCTTTAACTTACTAAAAGAAATTGATACCAGTACCAAAGTTGA
125 --I--S--V--H--K--F--F-=N=--L--L--K-=E=-I--D--T--S--T--K--V--D
601 TAATGCTATGTCAAGACTGTTGAAGAAGTATGATGTATTGTTTGCACTCTTCAGCAAATT
435 TAATGCTATGTCAAGACTGTTGAAGAAGTATGATGTATTGTTTGCACTCTTCAGCAAATT
145 --N--A--M--S--R--L--L--K--K--Y--D--V--L--F--A--L--F--S--K--L
661 GGAAAGGACATGTGAACTTATATATTTGACACAACCCAGCAGTTCGATATCTACTGAAAT
495 GGAAAGGACATGTGAACTTATATATTTGACACAACCCAGCAGTTCGATATCTACTGAAAT
165 --E--R--T--C--E--L--I--Y--L--T--Q--P--S--S--S--I--S--T--E--I
721 AAATTCTGCATTGGTGCTAAAAGTTTCTTGGATCACATTTTTATTAGCTAAAGGGGAAGT
555 AAATTCTGCATTGGTGCTAAAAGTTTCTTGGATCACATTTTTATTAGCTAAAGGGGAAGT
185 --N--S--A--L--V--L--K--V--S--W--I--T--F--L--L--A--K--G--E--V


781 ATTACAAATGGAAGATGATCTGGTGATTTCATTTCAGTTAATGCTATGTGTCCTTGACTA
615 ATTACAAATGGAAGATGATCTGGTGATTTCATTTCAGTTAATGCTATGTGTCCTTGACTA
205 --L--Q--M--E--D--D--L--V--I--S--F--Q--L--M--L--C--V--L--D--Y
841 TTTTATTAAACTCTCACCTCCCATGTTGCTCAAAGAACCATATAAAACAGCTGTTATACC
675 TTTTATTAAACTCTCACCTCCCATGTTGCTCAAAGAACCATATAAAACAGCTGTTATACC
225 --F--I--K--L--S--P--P--M--L--L--K--E--P--Y--K--T--A--V--I--P
901 CATTAATGGTTCACCTCGAACACCCAGGCGAGGTCAGAACAGGAGTGCACGGATAGCAAA
735 CATTAATGGTTCACCTCGAACACCCAGGCGAGGTCAGAACAGGAGTGCACGGATAGCAAA
245 --I--N--G--S--P--R--T--P--R--R--G--Q--N--R--S--A--R--I--A--K
961 ACAACTAGAAAATGATACAAGAATTATTGAAGTTCTCTGTAAAGAACATGAATGTAATAT
795 ACAACTAGAAAATGATACAAGAATTATTGAAGTTCTCTGTAAAGAACATGAATGTAATAT
265 --Q--L--E--N--D--T--R--I--I--E--V--L--C--K--E--H--E--C--N--I
1021 AGATGAGGTGAAAAATGTTTATTTCAAAAATTTTATACCTTTTATGAATTCTCTTGGACT
855 AGATGAGGTGAAAAATGTTTATTTCAAAAATTTTATACCTTTTATGAATTCTCTTGGACT
285 --D--E--V--K--N--V--Y--F--K--N--F--I--P--F--M--N--S--L--G--L
1081 TGTAACATCTAATGGACTTCCAGAGGTTGAAAATCTTTCTAAACGATACGAAGAAATTTA
915 TGTAACATCTAATGGACTTCCAGAGGTTGAAAATCTTTCTAAACGATACGAAGAAATTTA
305 --V--T--S--N--G--L--P--E--V--E--N--L--S--K--R--Y--E--E--I--Y
1141 TCTTAAAAATAAAGATCTAGATGCAAGATTATTTTTGGATCATGATAAAACTCTTCAGAC
975 TCTTAAAAATAAAGATCTAGATGCAAGATTATTTTTGGATCATGATAAAACTCTTCAGAC
325 --L--K--N--K--D--L--D--A--R--L--F--L--D--H--D--K--T--L--Q--T
1201 TGATTCTATAGACAGTTTTGAAACACAGAGAACACCACGAAAAAGTAACCTTGATGAAGA
1035 TGATTCTATAGACAGTTTTGAAACACAGAGAACACCACGAAAAAGTAACCTTGATGAAGA
345 --D--S--I--D--S--F--E--T--Q--R--T--P--R--K--S--N--L--D--E--E
R
1261 GGTGAATGTAATTCCTCCACACACTCCAGTTAGGACTGTTATGAACACTATCCAACAATT
1095 GGTGAATGTAATTCCTCCACACACTCCAGTTAGGACTGTTATGAACACTATCCAACAATT
365 --V--N-=V=-I--P--P--H--T--P--V--R--T--V--M--N--T--I--Q--Q--L
W
1321 AATGATGATTTTAAATTCAGCAAGTGATCAACCTTCAGAAAATCTGATTTCCTATTTTAA
1155 AATGATGATTTTAAATTCAGCAAGTGATCAACCTTCAGAAAATCTGATTTCCTATTTTAA
385 --M--M--I--L--N--S--A--S--D--Q--P--S--E--N--L--I--S--Y--F--N
1381 CAACTGCACAGTGAATCCAAAAGAAAGTATACTGAAAAGAGTGAAGGATATAGGATACAT
1215 CAACTGCACAGTGAATCCAAAAGAAAGTATACTGAAAAGAGTGAAGGATATAGGATACAT
405 --N--C--T--V--N--P--K--E--S--I--L--K--R--V--K--D--I--G--Y--I
M
Y
1441 CTTTAAAGAGAAATTTGCTAAAGCTGTGGGACAGGGTTGTGTCGAAATTGGATCACAGCG
1275 CTTTAAAGAGAAATTTGCTAAAGCTGTGGGACAGGGTTGTGTCGAAATTGGATCACAGCG
425 --F--K--E--K--F--A--K--A--V--G-=Q=-G--C--V--E--I--G--S--Q-=R
1501 ATACAAACTTGGAGTTCGCTTGTATTACCGAGTAATGGAATCCATGCTTAAATCAGAAGA
1335 ATACAAACTTGGAGTTCGCTTGTATTACCGAGTAATGGAATCCATGCTTAAATCAGAAGA
445 =-Y--K--L--G--V--R--L--Y--Y--R--V--M--E--S--M--L--K--S--E--E


1561 AGAACGATTATCCATTCAAAATTTTAGCAAACTTCTGAATGACAACATTTTTCATATGTC
1395 AGAACGATTATCCATTCAAAATTTTAGCAAACTTCTGAATGACAACATTTTTCATATGTC
465 --E--R--L--S--I--Q--N--F--S--K--L--L--N--D--N--I--F--H--M--S
1621 TTTATTGGCGTGCGCTCTTGAGGTTGTAATGGCCACATATAGCAGAAGTACATCTCAGAA
1455 TTTATTGGCGTGCGCTCTTGAGGTTGTAATGGCCACATATAGCAGAAGTACATCTCAGAA
485 --L--L--A--C--A--L--E--V--V--M--A--T--Y--S--R--S--T--S--Q--N
S
1681 TCTTGATTCTGGAACAGATTTGTCTTTCCCATGGATTCTGAATGTGCTTAATTTAAAAGC
1515 TCTTGATTCTGGAACAGATTTGTCTTTCCCATGGATTCTGAATGTGCTTAATTTAAAAGC
505 --L--D--S--G--T--D--L--S--F--P--W--I--L--N--V--L--N--L--K-=A
1741 CTTTGATTTTTACAAAGTGATCGAAAGTTTTATCAAAGCAGAAGGCAACTTGACAAGAGA
1575 CTTTGATTTTTACAAAGTGATCGAAAGTTTTATCAAAGCAGAAGGCAACTTGACAAGAGA
525 =-F--D--F--Y--K--V--I--E--S--F--I--K--A--E--G--N--L--T--R--E
1801 AATGATAAAACATTTAGAACGATGTGAACATCGAATCATGGAATCCCTTGCATGGCTCTC
1635 AATGATAAAACATTTAGAACGATGTGAACATCGAATCATGGAATCCCTTGCATGGCTCTC
545 --M--I--K--H--L--E--R--C--E--H--R--I--M--E--S--L--A--W--L--S
W
1861 AGATTCACCTTTATTTGATCTTATTAAACAATCAAAGGACCGAGAAGGACCAACTGATCA
1695 AGATTCACCTTTATTTGATCTTATTAAACAATCAAAGGACCGAGAAGGACCAACTGATCA
565 --D--S--P-=L=-F--D--L--I--K--Q--S--K--D--R--E--G--P--T--D--H
R
1921 CCTTGAATCTGCTTGTCCTCTTAATCTTCCTCTCCAGAATAATCACACTGCAGCAGATAT
1755 CCTTGAATCTGCTTGTCCTCTTAATCTTCCTCTCCAGAATAATCACACTGCAGCAGATAT
585 --L--E--S--A--C--P--L--N--L--P--L--Q--N-=N=-H--T--A--A--D--M
1981 GTATCTTTCTCCTGTAAGATCTCCAAAGAAAAAAGGTTCAACTACGCGTGTAAATTCTAC
1815 GTATCTTTCTCCTGTAAGATCTCCAAAGAAAAAAGGTTCAACTACGCGTGTAAATTCTAC
605 --Y--L--S--P--V--R--S--P--K--K--K--G--S--T--T--R--V--N--S--T
2041 TGCAAATGCAGAGACACAAGCAACCTCAGCCTTCCAGACCCAGAAGCCATTGAAATCTAC
1875 TGCAAATGCAGAGACACAAGCAACCTCAGCCTTCCAGACCCAGAAGCCATTGAAATCTAC
625 --A--N--A--E--T--Q--A--T--S--A--F--Q--T--Q--K--P--L--K--S--T
2101 CTCTCTTTCACTGTTTTATAAAAAAGTGTATCGGCTAGCCTATCTCCGGCTAAATACACT
1935 CTCTCTTTCACTGTTTTATAAAAAAGTGTATCGGCTAGCCTATCTCCGGCTAAATACACT
645 --S--L--S--L--F--Y--K--K--V--Y--R--L--A--Y--L--R--L--N--T--L
2161 TTGTGAACGCCTTCTGTCTGAGCACCCAGAATTAGAACATATCATCTGGACCCTTTTCCA
1995 TTGTGAACGCCTTCTGTCTGAGCACCCAGAATTAGAACATATCATCTGGACCCTTTTCCA
665 --C--E--R--L--L--S--E--H--P--E--L--E--H--I--I--W--T--L--F--Q
S
2221 GCACACCCTGCAGAATGAGTATGAACTCATGAGAGACAGGCATTTGGACCAAATTATGAT
2055 GCACACCCTGCAGAATGAGTATGAACTCATGAGAGACAGGCATTTGGACCAAATTATGAT
685 --H--T--L--Q--N--E--Y--E--L--M--R-=D=-R--H--L--D--Q--I--M--M
2281 GTGTTCCATGTATGGCATATGCAAAGTGAAGAATATAGACCTTAAATTCAAAATCATTGT
2115 GTGTTCCATGTATGGCATATGCAAAGTGAAGAATATAGACCTTAAATTCAAAATCATTGT
705 --C--S--M--Y--G--I--C--K--V--K--N--I--D--L--K--F--K--I--I--V
2341 AACAGCATACAAGGATCTTCCTCATGCTGTTCAGGAGACATTCAAACGTGTTTTGATCAA
2175 AACAGCATACAAGGATCTTCCTCATGCTGTTCAGGAGACATTCAAACGTGTTTTGATCAA
725 --T--A--Y--K--D--L--P--H--A--V--Q--E--T--F--K--R--V--L--I--K


R
2401 AGAAGAGGAGTATGATTCTATTATAGTATTCTATAACTCGGTCTTCATGCAGAGACTGAA
2235 AGAAGAGGAGTATGATTCTATTATAGTATTCTATAACTCGGTCTTCATGCAGAGACTGAA
745 -=E=-E--E--Y--D--S--I--I--V--F--Y--N--S--V--F--M--Q--R--L--K
2461 AACAAATATTTTGCAGTATGCTTCCACCAGGCCCCCTACCTTGTCACCAATACCTCACAT
2295 AACAAATATTTTGCAGTATGCTTCCACCAGGCCCCCTACCTTGTCACCAATACCTCACAT
765 --T--N--I--L--Q--Y--A--S--T--R--P--P--T--L--S--P--I--P--H--I
2521 TCCTCGAAGCCCTTACAAGTTTCCTAGTTCACCCTTACGGATTCCTGGAGGGAACATCTA
2355 TCCTCGAAGCCCTTACAAGTTTCCTAGTTCACCCTTACGGATTCCTGGAGGGAACATCTA
785 --P--R--S--P--Y--K--F--P--S--S--P--L--R--I--P--G--G--N--I--Y
2581 TATTTCACCCCTGAAGAGTCCATATAAAATTTCAGAAGGTCTGCCAACACCAACAAAAAT
2415 TATTTCACCCCTGAAGAGTCCATATAAAATTTCAGAAGGTCTGCCAACACCAACAAAAAT
805 --I--S--P--L--K--S--P--Y--K--I--S--E--G--L--P--T--P--T--K--M
2641 GACTCCAAGATCAAGAATCTTAGTATCAATTGGTGAATCATTCGGGACTTCTGAGAAGTT
2475 GACTCCAAGATCAAGAATCTTAGTATCAATTGGTGAATCATTCGGGACTTCTGAGAAGTT
825 --T--P--R--S--R--I--L--V--S--I--G--E--S--F--G--T--S--E--K--F
2701 CCAGAAAATAAATCAGATGGTATGTAACAGCGACCGTGTGCTCAAAAGAAGTGCTGAAGG
2535 CCAGAAAATAAATCAGATGGTATGTAACAGCGACCGTGTGCTCAAAAGAAGTGCTGAAGG
845 --Q--K--I--N--Q--M--V--C--N--S--D--R--V--L--K--R--S--A--E--G
2761 AAGCAACCCTCCTAAACCACTGAAAAAACTACGCTTTGATATTGAAGGATCAGATGAAGC
2595 AAGCAACCCTCCTAAACCACTGAAAAAACTACGCTTTGATATTGAAGGATCAGATGAAGC
865 --S--N--P--P--K--P--L--K--K--L--R--F--D--I--E--G--S--D--E--A
2821 AGATGGAAGTAAACATCTCCCAGGAGAGTCCAAATTTCAGCAGAAACTGGCAGAAATGAC
2655 AGATGGAAGTAAACATCTCCCAGGAGAGTCCAAATTTCAGCAGAAACTGGCAGAAATGAC
885 --D--G--S--K--H--L--P--G--E--S--K--F--Q--Q--K--L--A--E--M--T
2881 TTCTACTCGAACACGAATGCAAAAGCAGAAAATGAATGATAGCATGGATACCTCAAACAA
2715 TTCTACTCGAACACGAATGCAAAAGCAGAAAATGAATGATAGCATGGATACCTCAAACAA
905 --S--T--R--T--R--M--Q--K--Q--K--M--N--D--S--M--D--T--S--N--K
2941 GGAAGAGAAATGAGGATCTCAGGACCTTGGTGGACACTGTGTACACCTCTGGATTCATTG
2775 GGAAGAGAAATGA.....
925 --E--E--K--*-.....
3001 TCTCTCACAGATGTGACTGTATAACTTTCCCAGGTTCTGTTTATGGCCACATTTAATATC
.....
.....
K
3061 TTCAGCTCTTTTTTGTGGATATAAAATGTGCAGATGCAATTGTTTGGGTGATTCCTAAGCC
.....
.....
W
3121 ACTTGAAATGTTAGTCATTGTTATTTATACAAGATTGAAAATCTTGTGTAAATCCTGCCA
.....
.....
3181 TTTAAAAAGTTGTAGCAGATTGTTTCCTCTTCCAAAGTAAAATTGCTGTGCTTTATGGAT


3241 AGTAAGAATGGCCCTAGAGTGGGAGTCCTGATAACCCAGGCCTGTCTGACTACTTTGCCT
3301 TCTTTTGTAGCATATAGGTGATGTTTGCTCTTGTTTTTATTAATTTATATGTATATTTTT
W
3361 TTAATTTAACATGAACACCCTTAGAAAATGTGTCCTATCTATCTTCCAAATGCAATTTGA
3421 TTGACTGCCCATTCACCAAAATTATCCTGAACTCTTCTGCAAAAATGGATATTATTAGAA
3481 ATTAGAAAAAAATTACTAATTTTACACATTAGATTTTATTTTACTATTGGAATCTGATAT
3541 ACTGTGTGCTTGTTTTATAAAATTTTGCTTTTAATTAAATAAAAGCTGGAAGCAAAGTAT
3601 AACCATATGATACTATCATACTACTGAAACAGATTTCATACCTCAGAATGTAAAAGAACT
3661 TACTGATTATTTTCTTCATCCAACTTATGTTTTTAAATGAGGATTATTGATAGTACTCTT
3721 GGTTTTTATACCATTCAGATCACTGAATTTATAAAGTACCCATCTAGTACTTGAAAAAGT
R
3781 AAAGTGTTCTGCCAGATCTTAGGTATAGAGGACCCTAACACAGTATATCCCAAGTGCACT
3841 TTCTAATGTTTCTGGGTCCTGAAGAATTAAGATACAAATTAATTTTACTCCATAAACAGA
S
3901 CTGTTAATTATAGGAGCCTTAATTTTTTTTTCATAGAGATTTGTCTAATTGCATCTCAAA
3961 ATTATTCTGCCCTCCTTAATTTGGGAAGGTTTGTGTTTTCTCTGGAATGGTACATGTCTT

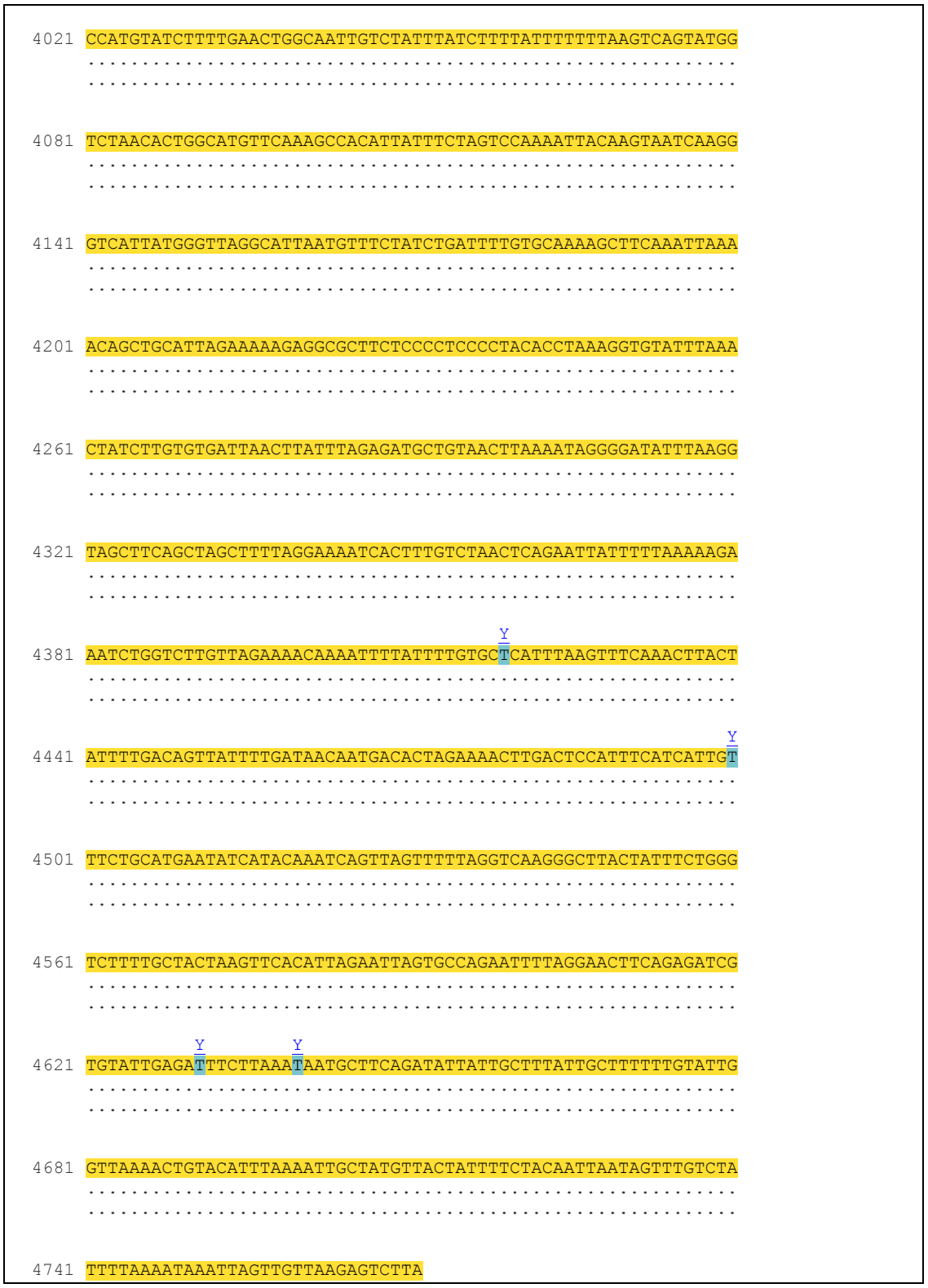


Figure 4. Transcript Sequence.

```
  1 MPPKTPRKTAATAAAAAAAEPPAPPPPPPPEEDPEQDSGPEDLPLVRLEFEETEEPDFTAL
 61 CQKLKIPDHVRERAWLTWEKVSSVDGVLGGYIQKKKELWGICIFIAAVDLDEMSFTFTEL
121 QKNIEISVHKFFNLLKEIDTSTKVDNAMSRLLKKYDVLFALFSKLERTCELIYLTQPSSS
181 ISTEINSALVLKVSWITFLLAKGEVLQMEDDLVISFQLMLCVLDYFIKLSPPMLLKEPYK
241 TAVIPINGSPRTPRRGQNRSARIAKQLENDTRIIEVLCKEHECNIDEVKNVYFKNFIPFM
301 NSLGLVTSNGLPEVENLSKRYEEIYLKNKDLDARLFLDHDKTLQTDSIDSFETQRTPRKS
361 NLDEEVNVIPPHTPVRTVMNTIQQLMMILNSASDQPSENLISYFNNCTVNPKESILKRVK
421 DIGYIFKEKFAKAVGQGCVEIGSQRYKLGVRLYYRVMESMLKSEEERLSIQNFSKLLNDN
481 IFHMSLLACALEVVMATYSRSTSQNLDSGTDLSFPWILNVLNLKAFDFYKVIESFIKAEG
541 NLTREMIKHLERCEHRIMESLAWLSDSPLFDLIKQSKDREGPTDHLESACPLNLPLQNNH
601 TAADMYLSPVRSPKKKGSTTRVNSTANAETQATSAFQTQKPLKSTSLSLFYKKVYRLAYL
661 RLNTLCERLLSEHPELEHIIWTLFQHTLQNEYELMRDRHLDQIMMCSMYGICKVKNIDLK
721 FKIIVTAYKDLPHAVQETFKRVLIKEEEYDSIIVFYNSVFMQRLKTNILQYASTRPPTLS
781 PIPHIPRSPYKFPSSPLRIPGGNIYISPLKSPYKISEGLPTPTKMTPRSRILVSIGESFG
841 TSEKFQKINQMVCNSDRVLKRSAEGSNPPKPLKKLRFDIEGSDEADGSKHLPGESKFQQK
901 LAEMTSTRTRMQKQKMNDSMDTSNKEEK
```

Figure 5. Peptide flow sequence.

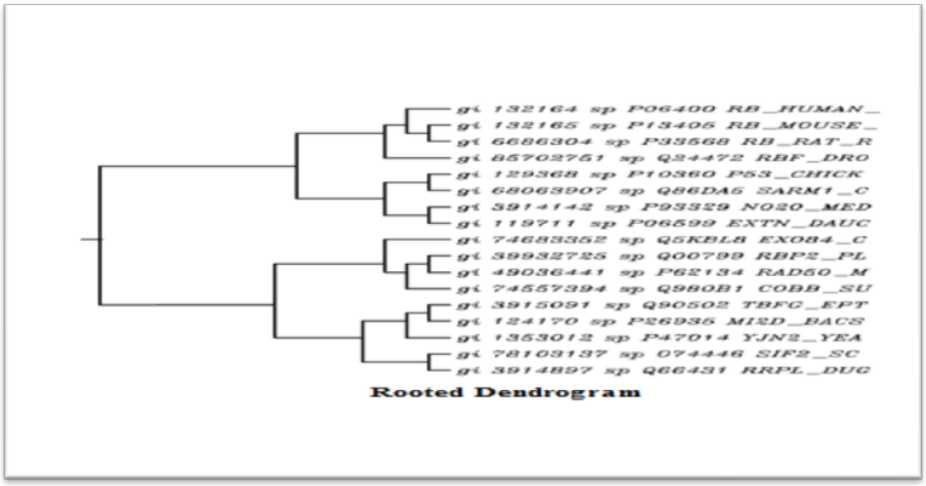


Figure 6. Rooted Dendrogram.

Docking summary report for Rb receptor and cyclosporine ligand										
	Clst	Soln	Models	Etotal	Eshape	Eforce	Vshape	Vclash	Lig(A,B,G)	Rec(B,G)
	1	1	000:000	-63.7	-63.7	0.0	0.0	0.0	281.250	44.316	83.044	124.494	180.000	42.00	-1	-1.00
Docking summary report for Rb receptor and actinomycin –d ligand										
	Clst	Soln	Models	Etotal	Eshape	Eforce	Vshape	Vclash	Lig(A,B,G)	Rec(B,G)
	1	1	000:000	-446.7	-446.7	0.0	0.0	0.0	348.750	58.283	108.000	104.273	126.293	27.50	-1	-1.00

Discussion

The process of data mining is concerned with the extraction of relevant information from the enormous amount of data available in various databases. In order to create a high-level description of the nature and the content of the data for the specified gene RB1, minning is performed. RB1 is a tumour suppressor gene and regulates other genes as well, specifying cell fate.

For the above purpose the data is mined from various websites such as NCBI, PubMed, KEGG, Ensembl, Uniprot, Swissprot, InterPro, Pfam, PDB and the tools used for extracting information are BLAST, CLUSTALW and ORF FINDER.

GROUND LEVEL MINING results in reduction of data to the minimum size possible through statistical representation, whereas the CHROMOSOME REPORT is studied through map viewer and the chromosome location of RB1 is visualized.

Lee *et al.,* [11] in their studies showed that RB tumor grown *in vitro* expresses highly specific photoreceptor cell genes. They did not find any marker genes specific to rod cells, but during the study of gene report it is evident that the specified gene regions are associated with marker gene as found through entrez search.

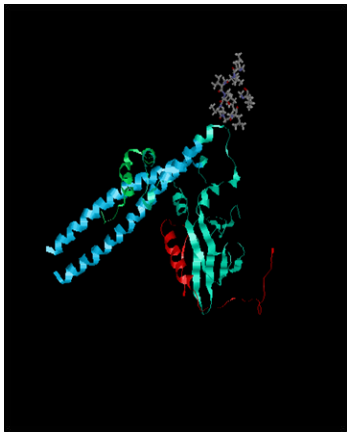


Figure 7. Rb binds with cyclosporine after docking.

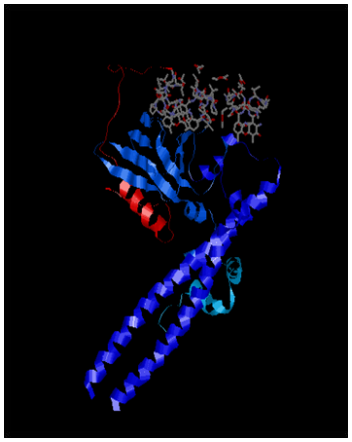


Figure 8. Rb and actinomycin–D in bond condition after docking.

Table 1. List of SNPs values according to region and position.

List Of SNPs								
Region	Contig. Position	mRNA Position	3D	Function	dbSNP allel	Protein Residue	Condon Pos	Amino Acid Pos
exon_1	29858050	139		start codon				1
exon_1	29858111	200		frame shift	G	Arg [R]	2	21
				contig ref	-		2	21
exon_4	29899233	535		Missense	C	His [H]	1	133
				contig ref	A	Asn [N]	1	133
	29899247	549		Missense	T	Asp [D]	3	137
				contig ref	A	Glu [E]	3	137
exon_11	29922716	1240		Missense	A	Ile [I]	1	368
				contig ref	G	Val [V]	1	368
exon_12	29927587	1311	Yes	synonymous	T	Ser [S]	3	391
			Yes	contig ref	A	Ser [S]	3	391
exon_13	29931145	1444	Yes	Missense	A	Lys [K]	1	436
			Yes	contig ref	C	Gln [Q]	1	436
exon_14	29933731		Yes	Nonsense	T	[Ter[*]]	1	445
			Yes	contig ref	C	Arg [R]	1	445
exon_16	29934327	1585		frame shift	A	Asn [N]	1	483
				contig ref	-		1	483
	29934328	1586		frame shift	A	His [H]	2	483
				contig ref	-		2	483
exon_17	29935459	1712	Yes	Missense	G	Gly [G]	2	525
			Yes	contig ref	C	Ala [A]	2	525
exon_18	30007141	1845	Yes	Missense	T	Phe [F]	3	569
			Yes	contig ref	A	Leu [L]	3	569
	30007230	1934		Missense	G	Ser [S]	2	599
				contig ref	A	Asn [N]	2	599
exon_20	30013955	2229		Missense	G	Glu [E]	3	697
				contig ref	C	Asp [D]	3	697
exon_22	30019160	2375		Missense	G	Gly [G]	2	746
				contig ref	A	Glu [E]	2	746

The mRNA REPORT is deciphered to gather detailed information regarding the gene for its identification, it revealed the exon and the transcription regions and PEPTIDE REPORT revealed highly descriptive data regarding Rb. The observations are in confirmation with Yokota *et al.,* who found markedly reduced amounts of Rb transcript in some small cell carcinomas. Special emphasis was given to protein interactions and blastp results to oblation homologous protein sequences.

Our results agree the fact as given by DeCaprio *et al.,* Buchkovich *et al.,* [12] and Chen *et al.,* [13] who demonstrated that the RB1 gene product has the properties of a cell cycle regulatory element and that its function is modulated by a phosphorylation/dephosphorylation mechanism during cell proliferation and differentiation, as evident from the KEGG pathway.

Phylogenetic tree is generated using homologous protein sequences for PHYLOGENETIC INFERENCE. Results show that RB homologous genes were found in various organisms varying from *Bacteria* to *Homo sapiens.* These organisms can be used as model organisms for *in vitro* and *in silico* studies related to novel drug discovery. Lee *et al.,* [14] used a rabbit antiserum against the RB for studing all cell lines expressing normal Rb mRNA. Sivakumaran *et al.,* also conducted a comprehensive survey of sequence variation in the RB1 gene in diverse

Table 2. Chromogram detail list.

Summary of the above Chromogram is as follows:		
Length	11,41,42,980	bps
Known Protein-coding Genes	361	
Novel Protein-coding Genes	31	
Pseudogene Genes	47	
miRNA Genes	18	
rRNA Genes	10	
snRNA Genes	28	
snoRNA Genes	40	
Misc RNA Genes	30	
SNPs	4,20,368	

human populations and primates.

The preclinical data for the treatment of retinoblastoma with Cyclosporin is given by Finger PT *et al.*, [15]. As a practice, retinoblastoma is cured using technique such as chemotherapy and radiotherapy. The drugs inhibiting the activity of Rb are used for chemotherapy in treating retinoblastoma as well as retinoblastoma associated malignancies. The preclinical data of cyclosporine suggest that it is having significant role in treating retinoblastoma malignancies. Whereas, the docking report of actinomycin-D with Rb shows far better binding capacity of actinomycin-D than cyclosporine. This is evident by the low e-total value of actinomycin-D; therefore the drug is suggested as a potential drug for future in treatment of retinoblastoma and associated retinoblastoma malignancies.

Actinomycin-D has been used successfully for the treatment of liver associated cancers, but the drug is now not used to treat retinoblastoma till date. No previous docking data for Actinomycin-D with retinoblastoma protein has been found before.

Therefore the mined data was characterized and organized in a manner that could be inferred easily for future utilization and preclinical trials can be carried out with actinomycin-D.

References

1. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100: 57–70. [Crossref]
2. Vo N, Goodman RH (2001) CREB-binding protein and p300 in transcriptional regulation. *J Biol Chem* 276: 13505-13508. [Crossref]
3. Yang XJ, Ogryzko VV, Nishikawa J, Howard BH, Nakatani Y (1996): A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* 382: 319–324
4. Yuan W, Condorelli G, Caruso M, Felsani A, Giordano A (1996) Human p300 protein is a coactivator for the transcription factor MyoD. J Biol Chem 271: 9009–9013. [Crossref]
5. Schiltz RL, Nakatani Y (2000) The PCAF acetylase complex as a potential tumor suppressor. *Biochim Biophys Acta* 1470: M37-53. [Crossref]
6. Eckner R, Yao TP, Oldread E, Livingston DM (1996) Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cell differentiation. *Genes Dev* 10: 2478-2490. [Crossref]
7. Puri PL, Sartorelli V, Yang XJ, Hamamori Y, Ogryzko VV, et al., (1997) Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Mol Cell* 1: 35-45. [Crossref]
8. McKinsey TA, Zhang CL, Olson EN (2001) Control of muscle development by dueling HATs and HDACs. *Curr Opin Genet Dev* 11: 497-504. [Crossref]
9. Kouzarides T (2000) Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* 19: 1176-1179. [Crossref]
10. Chan HM, Krstic-Demonacos M, Smith L, Demonacos C, La Thangue NB (2001) Acetylation control of the retinoblastoma tumour-suppressor protein. *Nat Cell Biol* 3: 667-674. [Crossref]
11. MacCarthy A, Birch JM, Draper GJ, Hungerford JL, Kingston JE, et al., (2009) Retinoblastoma in Great Britain 1963-2002. *Br J Ophthalmol* 93: 33-37. [Crossref]
12. Lee MH, Williams BO, Mulligan G, Mukai S, Bronson RT, *et al.,* (1996): Targeted disruption of p107: functional overlap between p107 and Rb. *Genes Dev* 10: 1621–1632.
13. Buchkovich K, Duffy LA, Harlow E (1989) the retinoblastoma protein is phosphorylated during specific phases of the cell cycle. *Cell* 58: 1097-1105. [Crossref]
14. Chen PL, Scully P, Shew JY, Wang JY, Lee WH (1989) Phosphorylation of the retinoblastoma gene product is modulated during the cell cycle and cellular differentiation. *Cell* 58: 1193-1198. [Crossref]
15. Lee EY, Chang CY, Hu N, Wang YC, Lai CC, et al., (1992) Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. *Nature* 359: 288-294. [Crossref]
16. Finger PT , Nadal-Ginard B (1999) Preclinical data for the treatment of retinoblastoma with Cyclosporin crystalline structure. NCBI. PubMed. ZX: 10986