B.Sc. (Hons.) Biotechnology
Core Course 13:
Basics of Bioinformatics and Biostatistics (BIOT 3013)

Unit 5:
Applications of bioinformatics in biotechnology

Dr. Satarudra Prakash Singh
Department of Biotechnology
Mahatma Gandhi Central University, Motihari

Introduction

- Bioinformatics is an interdisciplinary subject that develops databases and software tools for understanding biological data.

- It combines the knowledge from biology, computer science, information technology, mathematics and statistics to analyze and interpret biological data.

Applications

The key applications of bioinformatics include:

Biological databases,

Sequence alignment,

Gene and promoter prediction,

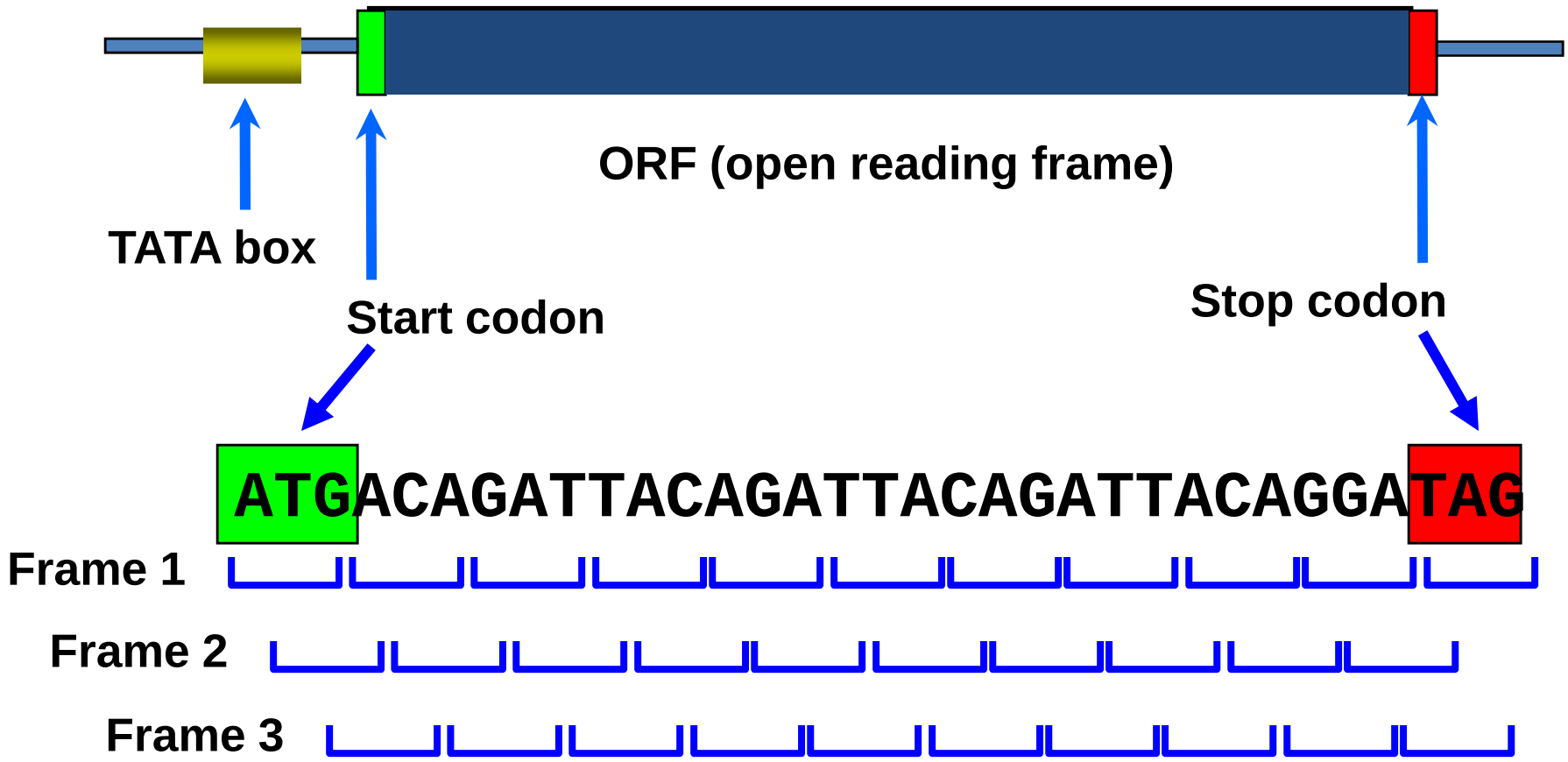
Molecular phylogenetic,

Genomics,

Primer designing,

Proteomics and drug designing

Prokaryotic gene annotation



Prokaryotic gene

- Advantages
  - Simple gene structure (no introns) with small genomes (0.5 to 10 million bp)
  - Genes are called Open Reading Frames (ORFs) with high coding density (>90%)
- Disadvantages
  - Some genes overlap (nested)
  - Some genes are quite short (<60 bp)

Gene finding approaches

- Rule-based method i.e. region of between start and stop codons (open reading frames with no. of codon >=50)e.g., GeneFinder
- Content-based method i.e. codon usage and promoter sites including GC content and TATA box.
- Similarity-based method i.e. finding orthologs through BLAST
- Pattern-based method i.e. machine-learning e.g., Artificial neural network (Grail, GrailEXP)

Example ORF

```
 5'                                                  3'
  atgcccaagctgaatagcgtagaggggttttcatcatttgaggacgatgtataa

1 atg ccc aag ctg aat agc gta gag ggg ttt tca tca ttt gag gac gat gta taa
   M   P   K   L   N   S   V   E   G   F   S   S   F   E   D   D   V   *
2  tgc cca agc tga ata gcg tag agg ggt ttt cat cat ttg agg acg atg tat
    C   P   S   *   I   A   *   R   G   F   H   H   L   R   T   M   Y
3   gcc caa gct gaa tag cgt aga ggg gtt ttc atc att tga gga cga tgt ata
     A   Q   A   E   *   R   R   G   V   F   I   I   *   G   R   C   I
```

Combined Methods

- GRAIL (http://compbio.ornl.gov/Grail-1.3/)
- FGENEH (http://www.bioscience.org/urllists/genefind.htm)
- HMMgene (http://www.cbs.dtu.dk/services/HMMgene/)
- GENSCAN(http://genes.mit.edu/GENSCAN.html)
- GenomeScan (http://genes.mit.edu/genomescan.html)
- Twinscan (http://ardor.wustl.edu/query.html)

Drug target identifications

- Bioinformatics is playing an important role in drug discovery and drug development.
- Currently, existing drugs in the market have about 500 proteins target.
- With an improved understanding of pathophysiology and advancement in computational tools, we can identify and validate new drug targets.

Drug target identifications

- These targets have more specific medicines that act on the cause, not merely the symptoms of the disease that led to lesser side effects .

- Bioinformatics tools are also effective in prediction, analysis and interpretation of clinical and preclinical findings.

Computer aided drug designing

- Computer can be used to identify and structurally modify a natural product and/ or design a synthetic compound with the desired properties.
- The analysis including similarity searching, clustering, QSAR modeling, virtual screening (docking)etc can be used to assess the therapeutic effects of drugs.

Protein-Ligand interactions

- Every biological reaction is started by protein-ligand interaction.
- Ligand binding plays an important role in regulation of biological function.
- Ligand binding may leads to the conformational changes in proteins and thus function.

Molecular docking

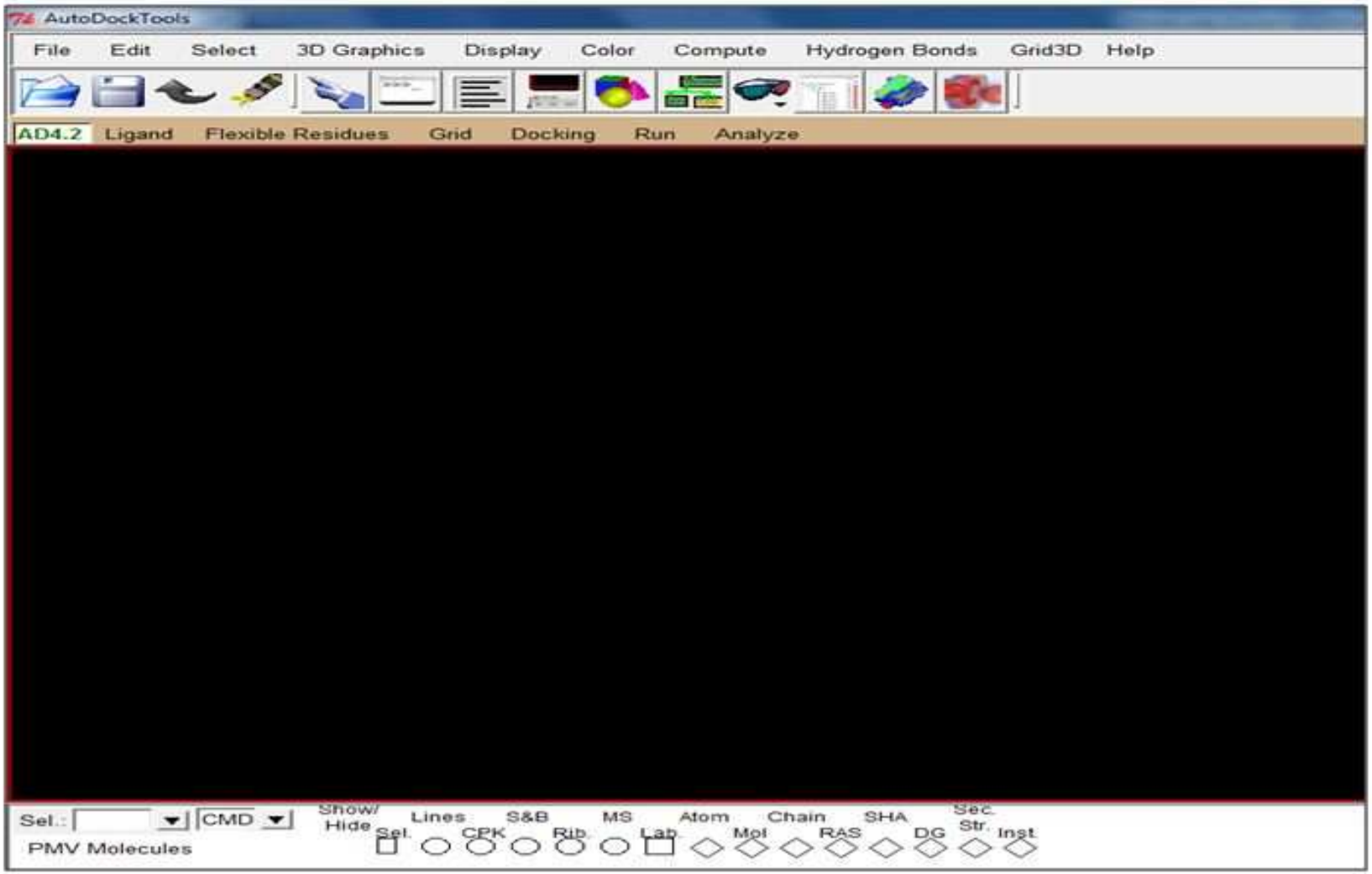
- Molecular docking predicts the preferred orientation of protein when bound to the ligand and form a stable complex.
- It can be referred as lock (protein) and key (ligand) model.
- Types of Docking: Rigid and flexible docking.
- In a rigid molecular docking, the molecules are referred as rigid objects and they cannot change their shape during interaction.

AutoDock

- In a flexible docking , the molecules are referred as flexible objects and can change their shapes according to the shape of ligand.

- AutoDock tool can be used to predict the behavior of the small molecules and helps user to perform the docking of ligands to a set of grids (target protein).

AutoDock tool



Pharmacogenomics

- Pharmacogenomics is the study of how an individual's genetic make up affects the response of drugs.
- Clinicians have to use trial and error method to find out the best drug for treating a particular patient because the same clinical symptoms can show a different range of responses to the same drug.
- In the future, doctors will be able to analyze a patient's genetic make up and prescribe the best available drug therapy and dosage.

Comparative genomic

- Analyzing and comparing the genetic material of different biological species is an important method for studying the functions of genes, the mechanisms of inherited diseases and evolution of species .

- Bioinformatics can be used to make comparisons of biochemical functions of genes in different organisms.

PCR primer designing

- Selecting appropriate primers is possibly the single most important factor affecting the polymerase chain reaction (PCR).
- A primer can be defined as short nucleic acid sequences that can act as a initial point for DNA synthesis.
- Specific amplification of the desired target requires that primers do not have matches to other targets and not allow undesired amplification.

Specific PCR primer designing

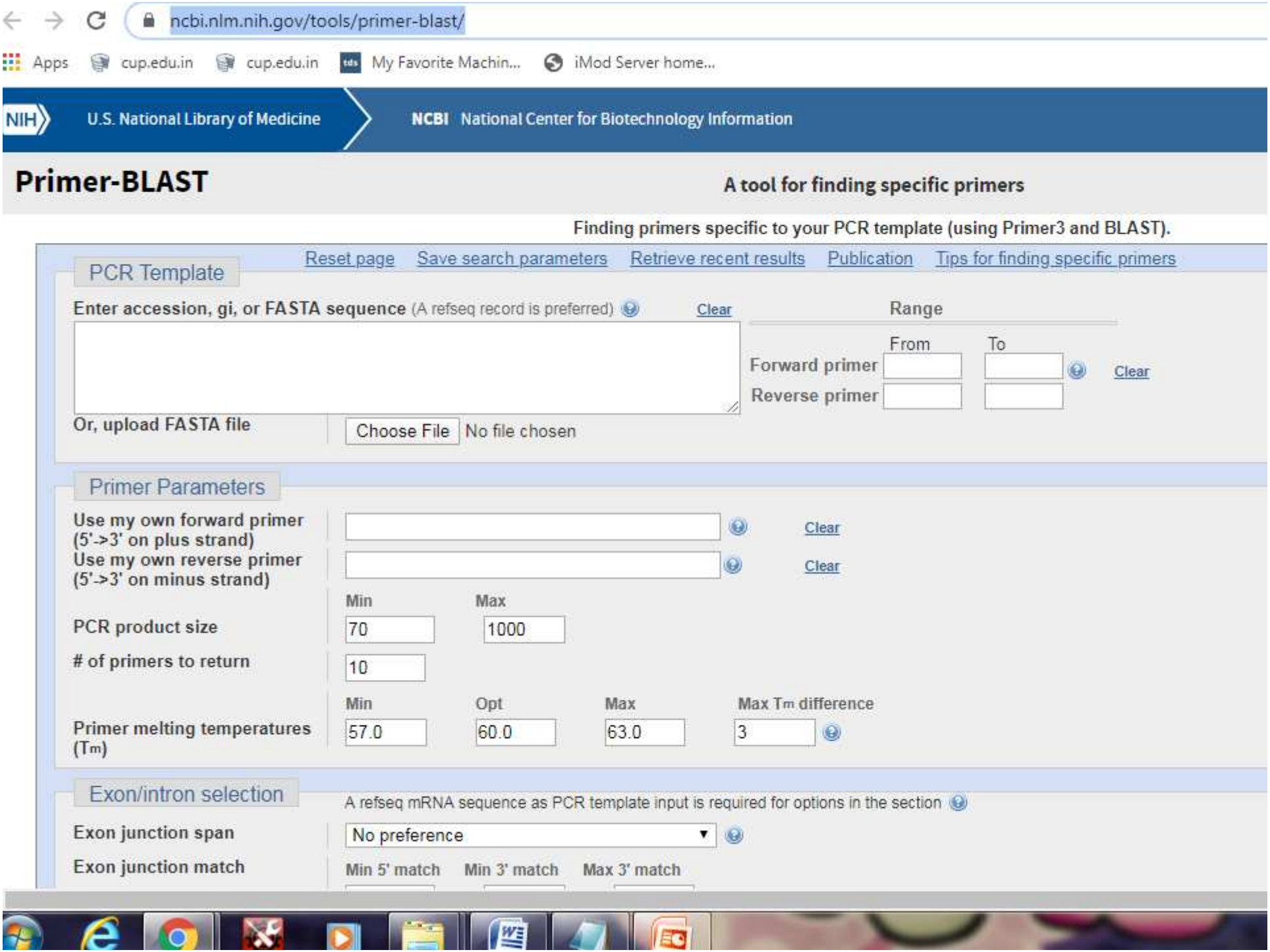
The process involves two steps:

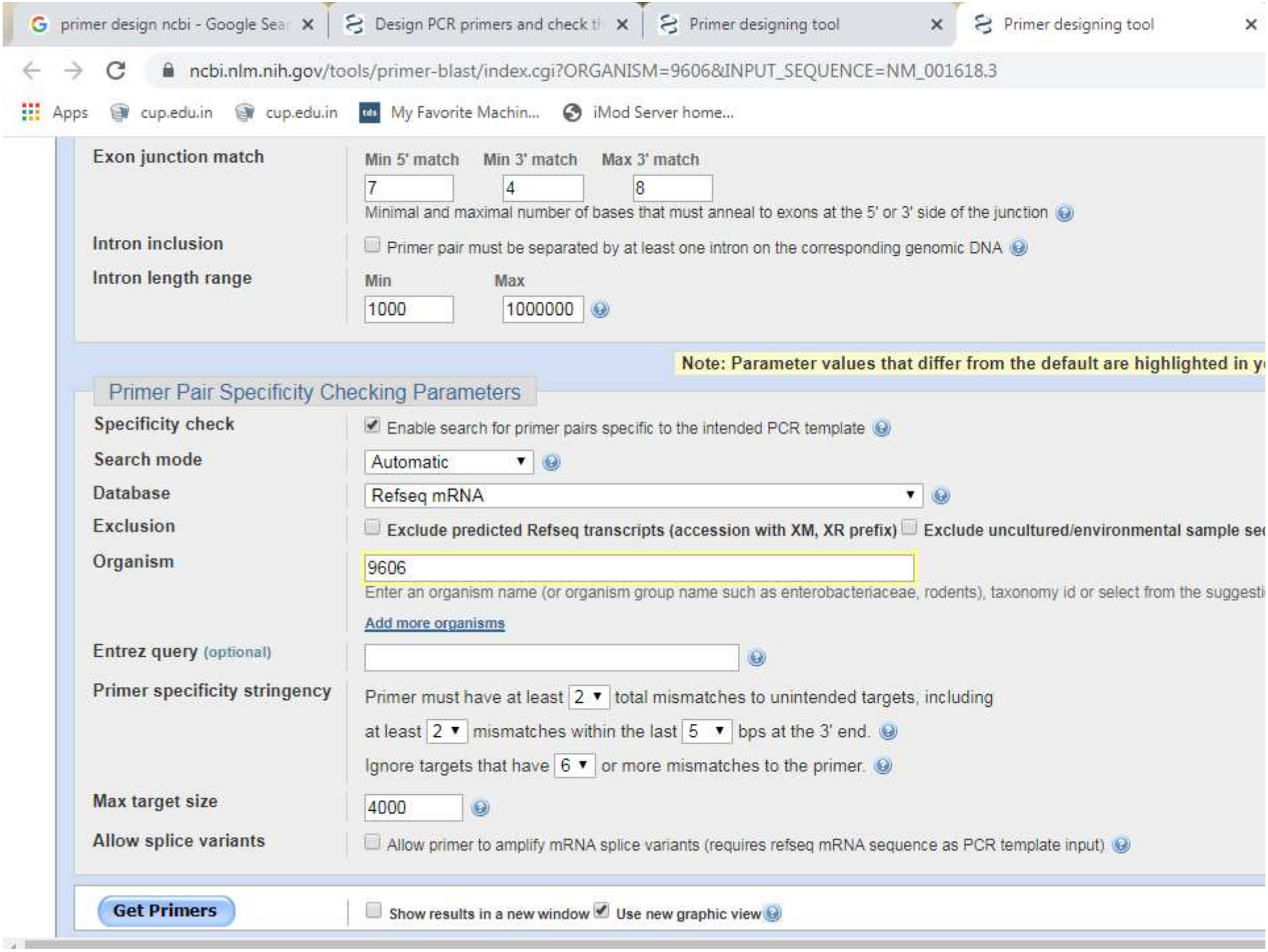
i) the primers flanking regions of interest are generated either manually or using software tools.

ii) then they are searched against an appropriate nucleotide sequence database using BLAST tool to examine the potential targets.

Properties of Primers

- Length of 18-24 bases.
- 40-60% G/C content.
- Start and end with 1-2 G/C pairs.
- Melting temperature (Tm) of 50-60°C.
- Primer pairs should have a Tm within 5°C of each other.
- Primer pairs should not have complementary regions.


ncbi.nlm.nih.gov/tools/primer-blast/
Apps
cup.edu.in
cup.edu.in
My Favorite Machin...
iMod Server home...
NIH
U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
Primer-BLAST
A tool for finding specific primers
Finding primers specific to your PCR template (using Primer3 and BLAST).
Reset page
Save search parameters
Retrieve recent results
Publication
Tips for finding specific primers
PCR Template
Enter accession, gi, or FASTA sequence (A refseq record is preferred)
Clear
Range
From
To
Forward primer
Reverse primer
Clear
Or, upload FASTA file
Choose File
No file chosen
Primer Parameters
Use my own forward primer (5'->3' on plus strand)
Clear
Use my own reverse primer (5'->3' on minus strand)
Clear
Min
Max
PCR product size
70
1000
# of primers to return
10
Min
Opt
Max
Max Tm difference
Primer melting temperatures (Tm)
57.0
60.0
63.0
3
Exon/intron selection
A refseq mRNA sequence as PCR template input is required for options in the section
Exon junction span
No preference
Exon junction match
Min 5' match
Min 3' match
Max 3' match


primer design ncbi - Google Sea
Design PCR primers and check t
Primer designing tool
Primer designing tool
ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?ORGANISM=9606&INPUT_SEQUENCE=NM_001618.3
Apps
cup.edu.in
cup.edu.in
My Favorite Machin...
iMod Server home...
Exon junction match
Min 5' match
Min 3' match
Max 3' match
7
4
8
Minimal and maximal number of bases that must anneal to exons at the 5' or 3' side of the junction
Intron inclusion
Primer pair must be separated by at least one intron on the corresponding genomic DNA
Intron length range
Min
Max
1000
1000000
Note: Parameter values that differ from the default are highlighted in y
Primer Pair Specificity Checking Parameters
Specificity check
Enable search for primer pairs specific to the intended PCR template
Search mode
Automatic
Database
Refseq mRNA
Exclusion
Exclude predicted Refseq transcripts (accession with XM, XR prefix)
Exclude uncultured/environmental sample se
Organism
9606
Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggesti
Add more organisms
Entrez query (optional)
Primer specificity stringency
Primer must have at least 2 total mismatches to unintended targets, including
at least 2 mismatches within the last 5 bps at the 3' end.
Ignore targets that have 6 or more mismatches to the primer.
Max target size
4000
Allow splice variants
Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)
Get Primers
Show results in a new window
Use new graphic view

References

- https://www.addgene.org/protocols/primer-design/
- https://www.ncbi.nlm.nih.gov/tools/primer-blast/

Thank you.

Email: sprakashsingh@mgcub.ac.in