DNA and Enzyme Precautions on Mutations

Final Report

Abraham Ramirez Sierra
Carlos Vega
Viraj Khatri
Zion Zapien

Doctor Vasileios Christopoulos
Department of Bioengineering
University of California Riverside

Abstract

The goal of this study and computation is to understand how the cell is able to capture and read the information of the nucleotide sequence and be able to analyze whether or not there is a mutation or if there is the correct sequence. This happens when there is a nucleotide change in the structure of the DNA causing errors in the reading and completely changes the chain of nucleotides changing the sequence of the amino acids. Mutations have different effects, that is it can be harmful, neutral and or a helpful mutation for the organism. We will look at how different enzymes and proteins surround DNA polymerase and understand the mechanical and chemical functions that occur to find and correct the mutation. We will also look at factors that affect fidelity of the cell. In addition to that we computate a model on MATLAB to show nucleotide density throughout the sequence and will also run a code that demonstrates how DNA and the enzymes act as a feedback to mistakes found in its DNA .

Introduction

Many cells undergo the process of mitosis in which cell division produces new cells for growth, repair and the replacement of older cells. Various cell types have various reproduction rates of new cells which can be continuous in nature. In the process of mitosis, every element of the cell is copied down to the deoxyribonucleic acid (DNA). Daughter cells are produced, which are exact copies of the parent cell, encompassing the entire DNA library that the parent cell fosters. However, this continuous and constant process throughout an organism does result in errors. These errors are known as mutations which can occur when an incorrect nucleotide is placed where the intended nucleotide should be situated. Mutations can result in permanent changes within the organism that have several effects ranging from nonexistent to exceedingly harmful. These harmful effects ultimately lead to cell deformation and tumor growth. Mutations can occur in several various ways such as by substitution, insertion, or deletion. However, these mutations are not always of permanent consequence as cells have developed ways to locate, modify, and repair these mutations that frequently occur within the human body.

DNA replication is a monumentally important process within the human body that involves the copying of a cell's DNA to produce two identical copies of this DNA. DNA replication can also be semiconservative meaning that each copy of DNA consists of an original strand and a synthesized strand. These differing strands have different classifications with original strands of DNA being known as parent or template strands and the synthesized strands being classified as daughter strands. The process of DNA replication begins within a cell when an enzyme called helicase unwinds and unzips the DNA molecule. More specifically, the helicase binds to the DNA molecule and unwinds the helix bond by breaking the hydrogen bonds between the nitrogenous bases. Another element in DNA replication is DNA polymerase III which synthesizes DNA, proofreads its errors, locates the error, and corrects it. DNA polymerase also carries out transcription along with accessory proteins known as transcription factors which bind

to DNA sequences. These DNA sequences are known as enhancer and promoter sequences. During the process of DNA replication, DNA polymerase adds nucleotides and is able to detect the mutation rate within an organism. Proofreading enzymes can also recognize incorrect mutations however, it is not possible to remove all mutations. As a result of this, enzymes and proteins come together to form a complex transcription initiation complex. The process of DNA replication has much significance because every part is responsible for the creation and maintenance of DNA in a living organism. Elements of DNA replication such as DNA polymerase are integral because they ultimately maintain the integrity and accuracy of the DNA code. These enzymes within DNA replication also reduce potentially harmful errors in the daughter strand while it is being synthesized. Ultimately, DNA replication is significant because it fosters the perpetuation of the structure that is DNA which is essential to all life.

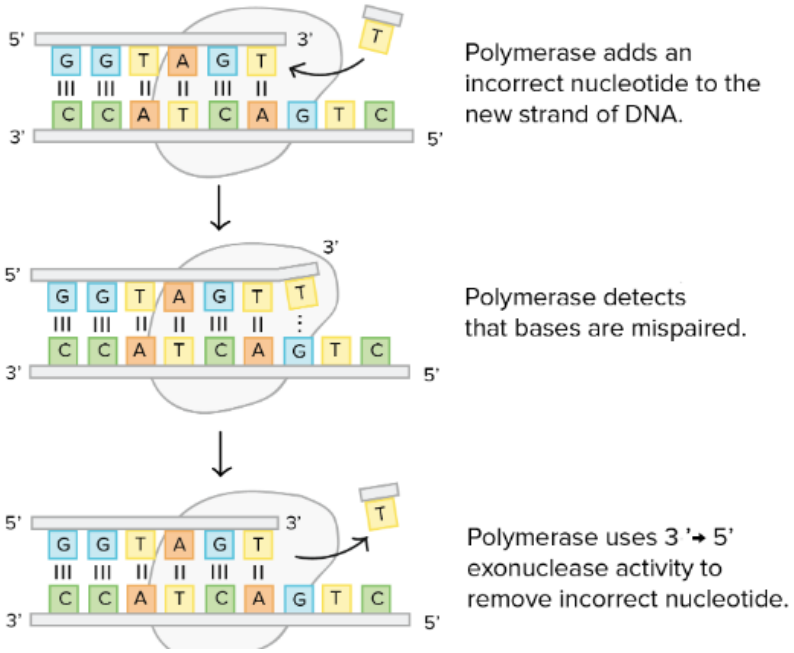


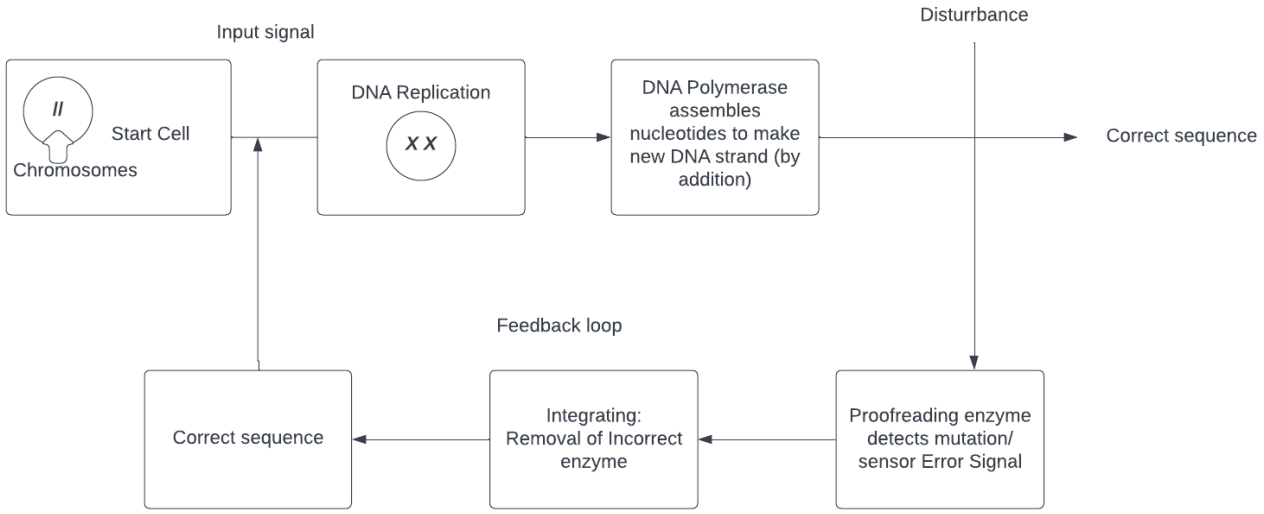
Figure 1: The process of proofreading undergone by DNA polymerase.

Main Body

The mutation is specifically a change in nucleic acid. Mistakes in the cell's DNA that lead to abnormal protein production are called mutations. Any change that occurs within the RNA or DNA can have a mutation. They originate at the DNA level and sow their effects at the protein level. Mutations occur in all animals including humans and also in plants, fungi, protists, bacteria, and viruses. Many mutations are neutral mutations. But they can be harmful and neutral. There is no will of which mutations can occur; they are very random. Both external and internal factors can cause these mutations. Mutations include substitution which means the wrong base is matched, an insertion which means an extra-base or bases are added and finally

there is deletion where a base is removed. Insertion and deletion are where there is a high risk of mutation. There is a shift in reading. The addition of the single base can shift the reading frame and therefore lead to many amino acid changes. Chromosomal mutations are made up of highly organized DNA and protein, they have a lot of genes. Inversion is where chromosomes are flipped. Deletion is where there is a lost chromosome. Duplication is where there is a repeat and in translocation there where one chromosome is aberrantly attached to another chromosome. An entire protein will get altered because of mutation such as deletion or insertion. Nucleotides are read by the group of three codons and let's say if the second nucleotide then the resulting codon will change entirely and read a different set of proteins. When it needs to Stop, there is a STOP codon which is an indication for the ribosome to terminate; because of mutation, it is entirely different, causing a premature stop codon because of nucleotide removal causing changes of the amino acids in proteins. Therefore the earlier a nucleotide is removed from the chain (triplet) the higher the probability of a mutation. As there is a shift in the reading frame where the right condones shift to the left and vice versa depending on if it is addition or subtraction. This means that every single amino acid downstream from that has a high likelihood of being correct and that protein will have very different properties from the desired one.

Block Diagram



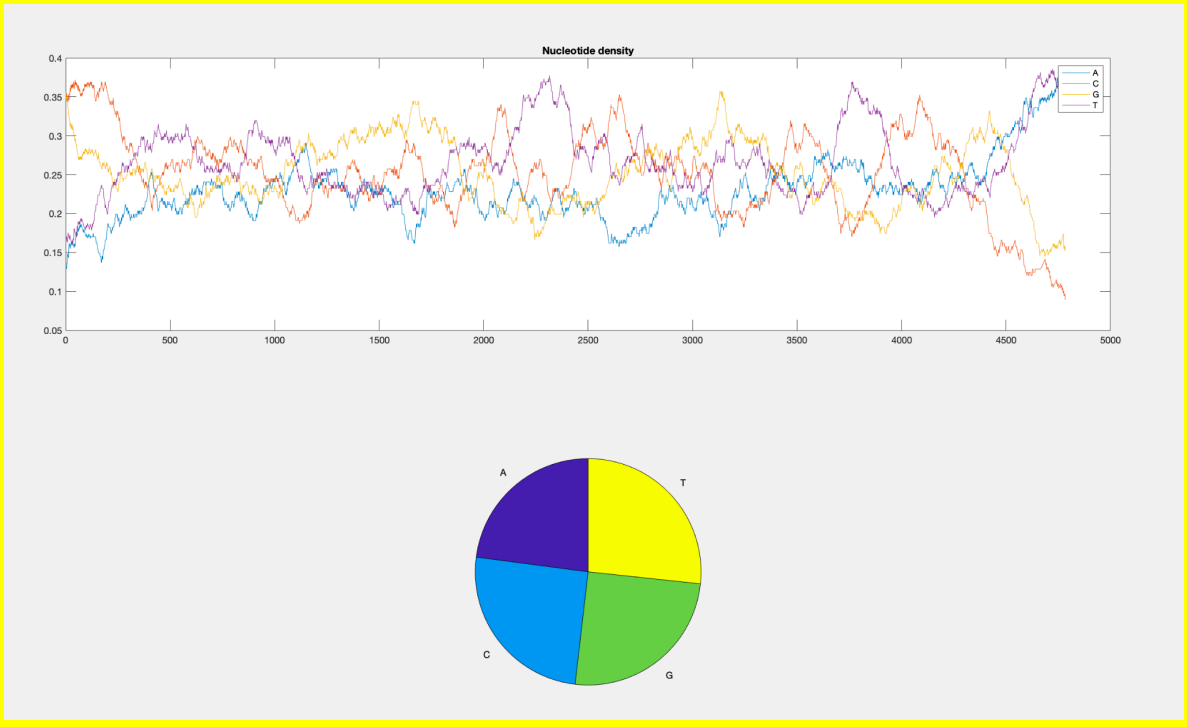


Figure 2: The code ran showed the number for each nucleotide. It also demonstrates the nucleotide density of each nucleotide as we go down the DNA strand.

A: 1097

C: 1207

G: 1204

T: 1277

The data up above shows the entire DNA sequence that was run from the code using the Bioinformatics tool on MatLab. The first part of the graph shows the nucleotide sequence and the densities of the certain nucleotide that is shown as we move down the DNA strand. It could be one of the four Adenine, Thymine, Guanine and Cytosine. The gene that this incorporates is for the Tay-sachs which is a disease that is inherited from both of the parents. This disease is one that is caused by frameshift mutation and this creates an entirely new different set of proteins after mRNA passes it though protein synthesis. The nucleotides that are more prone to mutation are the Cytosine and the Guanine and for this case if the mutation occurs this leads to a frame shift that makes the ‘unnecessary’ proteins for the body. Over time this will lead to neurological problems in the brain due to neurons dying from not receiving the correct proteins.

For the second part of the Matlab code

```
%For a Frameshift strand

Frameshift_strand =

   'CGTATATCTATCCTATGCCCCTGAC'

protein =

   'Arg Ile Ser Ile Leu Ser Pro'


%For a non-shift strand

nonshift_strand =

   'CGTATATCCTATGCCCCTGAC'

protein =

   'Arg Ile Ser Tyr Gly Pro Asp'
```

Above we see that the frameshift mutation can be damaging due to the protein being completely different just from moving the rest of the strand down after the ninth nucleotide by four base pairs.

```
%For complimenting coding and template strands

Template_strand = 'GCATATAGGATACGGGGACTG';
Coding_strand =   'CGTATATCCTATGCCCCTGAC';

outputstring1 =
   'GCATATAGGATACGGGGACTG'

outputstring2 =
   'CGTATATCCTATGCCCCTGAC'


%For coding and template strand with a base pair that does not match

Template_strand = 'GCATATAGGATACGGGGATTG';
```

Coding_strand = 'CGTATATCCTATGCCCCTGAC';

outputstring1 =

'GCATATAGGATACGGGGAXTG'

outputstring2 =

'CGTATATCCTATGCCCCTXAC'

Above we see how the DNA polymerase will recognize when base pairs of the coding and template strands do not match, using a feedback loop shown in the code below, to mark the mistakes with X's. If they are complementary to each other, then it is left alone.

Matlabcode

```
clearvars

close all

clc


humanHEXA = getgenbank('NM_000520')


basecount(humanHEXA)

ntdensity(humanHEXA)

basecount(humanHEXA,'chart','pie');

amino_acid_sequence=nt2aa(humanHEXA)


LocusName: 'NM_000520'
      LocusSequenceLength: '4785'
     LocusNumberofStrands: ''
           LocusTopology: 'linear'
        LocusMoleculeType: 'mRNA'
```

LocusGenBankDivision: 'PRI'
LocusModificationDate: '07-JUN-2022'
Definition: 'Homo sapiens hexosaminidase subunit alpha (HEXA), transcript variant 2, mRNA.'

DNA Polymerase Feedback Loop Code

```
% This code starts out with our normal HEXA allele coding strand accompanied with its complementary strand. This represents the input in our block diagram

Template_strand = 'GCATATAGGATACGGGGACTG';
Coding_strand =    'CGTATATCCTATGCCCCTGAC';
outputstring1 = '';
outputstring2 = '';
for i = 1 : length(Template_strand)

  ch1 = Template_strand(i);
  ch2 = Coding_strand(i);

%These if statements are the feedback system of the DNA polymerase to make sure the right pairs are paired and if not it will take it out

  matching = 0;
  if (ch1 == 'A') & (ch2 == 'T')  matching = 1;  end;
  if (ch1 == 'T') & (ch2 == 'A')  matching = 1;  end;
  if (ch1 == 'C') & (ch2 == 'G')  matching = 1;  end;
  if (ch1 == 'G') & (ch2 == 'C')  matching = 1;  end;

% If the pairs are a match then they are left as is

  if (matching == 1)

     outputstring1(i) = ch1;
     outputstring2(i) = ch2;

%If they are not matching then they will get cut out shown by these X's

  else

      outputstring1(i) = 'X';
      outputstring2(i) = 'X';

  end
end
```

```
%The output of the DNA polymerase control system shown in our block diagram is the
correct DNA strand after checking for mismatches

outputstring1
Outputstring2

%This second half of the code focuses on the frame-shift mutation that occurs when a
chunk of nucleotides are inserted into the DNA strand causing the sequence to code for
the wrong proteins and eventually lead to harmful effects as shown in patients with Tay
Sachs disease.

Frameshift_strand = outputstring2;
insertionposition = 9;
insertionnucleotide = 'TATC';

%This chunk of nucleotides are inserted into the sequence of DNA after the 9th position

Frameshift_strand(insertionposition+length(insertionnucleotide):end+length(insertionnu
cleotide)) = Frameshift_strand(insertionposition:end);
Frameshift_strand(insertionposition:insertionposition+length(insertionnucleotide)-1) =
insertionnucleotide;
nCodon = floor(length(Frameshift_strand)/3);
protein = '';
stopcodon = 0;
for ii = 1 : nCodon
  codon = Frameshift_strand(3*ii-2:3*ii);
  switch (codon)
     case 'CGT'
        amino = 'Arg';
     case 'ATA'
        amino = 'Ile';
     case 'TCC'
        amino = 'Ser';
     case 'TAT'
        amino = 'Tyr';
     case 'GCC'
        amino = 'Gly';
     case 'CCT'
        amino = 'Pro';
     case 'GAC'
        amino = 'Asp';
     case 'TCT'
        amino = 'Ser';
```

```
    case 'ATC'
       amino = 'Ile';
    case 'CTA'
       amino = 'Leu';
    case 'TGC'
       amino = 'Ser';
    case 'CCC'
       amino = 'Pro';
    case 'TGA'
       stopcodon = 1;
    otherwise
       amino = '';
  end
  if (stopcodon == 0)
    if (length(amino) == 0)
       ;
    else
       if (length(protein) == 0)
          protein = amino;
       else
          protein(end+2:end+4) = amino;
       end
    end
  end
end
Frameshift_strand
Protein
```

Conclusion

We were able to obtain a successful code that allowed us to see how we can interpret a strand of DNA and see the nucleotide density as the code runs through the strand.This was done to show how the data of a strand is read and what is displayed if the case of a wrong nucleotide is inserted. Our second code allowed us to run a part of the strand and run a simulation that allowed us to see how DNA polymerase is used as a feedback mechanism and is also used to correct the problem. In the end we also were able to understand more fully how important this mechanism and feedback system is to the human body or any living organism.

References

Mutation Research / Fundamental and Molecular Mechanisms of Mutagenesis Special Issue: DNA Repair and Genetic Instability University of Toledo College of Medicine and Life Sciences, Department of Biochemistry & Cancer Biology, 3000 Transverse Dr., Toledo, OH 43614, USA

Shao et al., Coupling of DNA Replication and Negative Feedback Controls Gene Expression for Cell-Fate Decisions,iScience, Volume 6, 2018, Pages 1-12

Clancy, Suzanne. "Scitable by Nature Edition." *Genetic Mutation*, 2008, https://www.nature.com/scitable/topicpage/genetic-mutation-441/#:~:text=While%20a%20mutation%20is%20defined,is%20common%20in%20the%20population.

Khan Academy. (2022). *DNA proofreading and Repair (article)*. Khan Academy. Retrieved June 7, 2022, from https://www.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-replication/a/dna-proofreading-and-repair

Liguori M, Tagarelli G, Romeo N, Bagala A, Spadafora P. Identification of a patient affected by "juvenile chronic" Tay Sachs disease in South Italy. Neurol Sci.

Walker JR, Corpina RA, Goldberg J. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature. (2001);412:607–614.