DNA combination and recombination state

JELENKA SAVKOVIC STEVANOVIC
Faulty of Technology and Metallurgy, Belgrade University, Karnegijeva 4, 11000 Belgrade, SERBIA
savkovic@tmf.bg.ac.rs

Abstract: - DNA sequencing strands information were defined. The synthesis rates of purine and pyrimidine oxyribonucleotides and deoxyribonucleotides characterized DNA features. The most important regulator of de novo purine biosynthesis is the intracellular concentration of the phosphoribosyl-pyrophosphate-phosphoribosylamine. Mammalian liver is a major site of purine nucleotide synthesis and provides purines in the form of bases or nucleosides to be salvaged and utilized by those tissues incapable of synthesizing purines de novo. In this paper DNA recombination was studied. The information value of the fragment in genome was derived.

Key-Words: - Recombination, enzyme, catabolism, nucleotide, DNA features, genomic library.

1 Introduction
The structure and function of the purines and pyrimidines and their nucleosides and nucleotides were studied in numerous literature. Synthetic analogs of naturally occurring nucleotides find application in cancer chemotherapy as enzyme inhibitors and can replace the naturally occurring nucleotides in nucleic acids. Therapeutic attempts to inhibit the growth of cancer cells or certain viruses have often employed administration of analogs of bases, nucleosides, or nucleotides that inhibit the synthesis of either DNA or RNA. Allopurinol, a purine analog, is widely used in the treatment of gout.

Biomedical important it neither nucleotides nor their parent purine and pyrimidine bases in the diet are incorporated into human tissue nucleic acids or into purine or pyrimidine coenzymes. Even when a diet rich in nucleic proteins is ingested, human subjects form the constituents of tissue nucleic acids from amphibolic intermediates. This de novo synthesis permits purine and pyrimidine analogs with potential as anticancer drugs to be incorporated into DNA. The rates of synthesis of purine and pyrimidine oxy- and deoxyribonucleotides are subject to precise regulation[1]-[3]. Mechanisms have evolved to ensure production of these compounds in quantities and at times appropriate to meet varying physiologic demand[4]-[7].

In addition to de novo synthesis, these include “salvage” pathways for reutilization of purine or pyrimidine bases released by degradation of nucleic acids in vivo. Human diseases that involve abnormalities in purine or pyrimidine metabolism include gout, Lesch-Nyhan syndrome, Reye’s syndrome, adenosine deaminase deficiency, and purine nucleoside phosphorylase deficiency.

In this paper regulatory mechanisms of purine and pyrimidine biosynthesis were studied.

2 Synthetic nucleotide
Synthetic analogs of nucleobases, nucleosides, and nucleotides are widely used in the medical sciences and clinical medicine. In the past, most of these uses have depended upon the role of nucleotides as components of nucleic acids for cellular growth and division. For a cell to divide, its nucleic acids must be replicated. This requires that the precursors of nucleic acids – the normal purine and pyrimidine deoxy- be readily available. One of the most important components of the oncologist’s pharmacopeia is the group of synthetic analogs of purine and pyrimidine nucleobases and nucleosides.

The pharmacologic approach has been to use an analog in which either the heterocyclic ring structure of the sugar moiety has been altered in such a way as to induce toxic effects when the analog becomes incorporated into various cellular constituents. Many of these effects results from inhibition by the drug of specific enzyme activities necessary for nucleic acids synthesis or from the incorporation of metabolites of the drug into the nucleic acids where they alter the base pairing essential to accurate transfer of information. Thus, it can be assumed:

\[ E + S \Leftrightarrow \ \overset{+}{k_+}\ \overset{+}{k_-}\ \overset{+}{P} + \overset{+}{E} \]

(1)
The reduction of ribonucleoside diphosphates to deoxyribonucleotide diphosphates is subject to complex regulation.

3 Catabolism of Purines

In humans the ultimate catabolite (end product) of purines is uric acid. On the basis of observations in humans with inherited enzyme deficiencies, it appears that over 99% of the uric acid is derived from substrates of purine nucleoside phosphorylase, a component of the purine salvage pathway. The purine products of purine nucleoside phosphorylase, guanine and hypoxanthine, are converted to uric acid by way of xanthine in reaction catalysed by the enzymes guanase and xanthine oxidase, respectively, of liver, small intestine, and kidney.

Xanthine oxidase is an important site for pharmacologic intervention in patients with hyperuricemia and gout. In lower primates and other mammals, but not in humans, uricase hydrolys uric acid to allantoin, a highly water soluble end product. Ambhibians, birds, and reptiles, which lack uricase, excrete uric acid and guanine as the end products of both purine metabolism and nitrogen (protein) metabolism.

4 Catabolism of Pyrimidines

The catabolism of pyrimidines, which occurs mainly in the liver, produces highly soluble end products. This contrasts with the production of the sparingly soluble uric acid and sodium urate by purine catabolism. The release of respiratory CO₂ from the ureido carbon of the pyrimidine nucleus represents a major pathway for the catabolism of uracil, cytosine, and thymine. β-alanine and β-aminoisobutyrate are the major end products of cytosine, uracil, and thymine catabolism, respectively.

Thymine is the precursor β-aminoisobutyrate in laboratory animals and in humans. The excretion of β-aminoisobutyrate is increased in leukemia as well as after the body has been subjected to x-irradiation. This is undoubtedly a reflection of increased destruction of cells and their DNA. A familial occurrence of an abnormally high excretion of β-aminoisobutyrate has also been observed in otherwise normal individuals. This genetic trait is traceable to recessively expressed gene. Higgs excretors results only when the trait is homozygous. Approximately 25% of tested . Although little is known about the mechanism whereby β-aminoisobutyrate is degraded in humans, an enzyme that catalyzes the reversible transamination reaction has been identified in pig kidney. The β-aminoisobutyrate is converted to methylmalonic semialdehyde and thence to propionate, which in turn proceeds to succinate.

The initial steps in degradation of pyrimidine nucleotides, including the removal of the sugar phosphate moiety by hydrolysis of the N-glucosidic bond are similar to reversing the latter part of the synthetic pathway. For pseudouridine, which is formed in t-RNA by a rearrangement reaction, there is no mechanism to catalyze the hydrolysis or phosphorylsis of this unusual nucleoside to its respective pyrimidine base, uracil. Consequently, pseudouridine is excreted unchanged in the urine of normal persons. Some individuals with urate overexcretion greater than 600mg/urate acid per 24 hours can be categorized as having secondary hyperuricemia. They have other disease processes such as cancer or psoriasis that lead enhanced tissue turnover.

Finally, there are persons with identifiable enzyme defects, including abnormalities of PRPP synthetase, the HGPRTase (hypoxanthine-guanine phosphoribosyltransferase ) deficiencies, both the complete Lesch-Nyhan syndrome and incomplete deficiencies and glucose-6-phosphatase deficiency von Gierke’s disease. There exists also a group of patients exhibiting idiopathic overproduction hyperuricemia, which will certainly be regarded as a heterogeneous group of disease ons the molecular bases for their metabolic defects are recognized.

5 DNA features

DNA is a complex biopolymer that is organized as a double helix. The fundamental organizational element is the sequence of purine (adenine [a] or guanine [G] and pyrimidine (cytosine [C] or thymine [T]) bases. These bases are attached to the C-1’ position of the sugar deoxyribose, and the bases are linked together through joining of the sugar moieties at their 3’and 5’positions via a phosphate groups from the backbone of the double helix.

Adenine and thymine always pair, by hydrogen bonding, as do guanine and cytosine. These base pairs are said to be complementary, and the guanine content of a fragment of double stranded DNA will always equal its cytosine content, likewise the thymine and adenine contents are equal. Interaction can be reduced by heating the DNA to denaturate it. The laws of base pairing predict that 2 complementary DNA strands will reanneal exactly in register upon renaturaton,as happen when the temperature of the solution is slowly reduced to normal. Indeed, the degree of base-pair matching or mismatching can be estimated from the temperature required for denaturation – renaturation. Segments of DNA with high degrees of base – pair matching require more energy input heat to accomplished denaturation, or, to put it another way, a closely matched segment will withstand more heat before the strand separate. This reaction is used to determine whether there are significant differences between 2 DNA sequences, and it underlines
the concept of hybridization, which is fundamental to the processes which be described hear.

There are about $3 \times 10^9$ base pairs (bp) in each human haploid genome. If an average gene length is $3 \times 10^3$ bp (three kilobases, kb), the genome could consist of $10^9$ genes, assuming that there is no overlap and that transcription proceeds in only one direction.

The double – helical DNA is packaged into a more compact structure by a number of proteins, most notably the basic proteins called histones. The DNA presents within nucleus of a cell, if simply extended, would be about a meter long. The chromosomal proteins compact this long length of DNA so that it can be packaged into a nucleus with a volume of a few cubic microns.

Gene organization is very interesting. In general, prokaryotic genes consist of a small regulatory region (100-500 bp) and a large protein-coding segment (500-10000 bp). Several genes are often controlled by a single regulatory unit. Most mammalian genes are more complicated, in that the coding region are interrupted by noncoding regions that that are eliminated when the primary RNA transcript is processed into mature messenger RNA (mRNA). The coding regions ( those regions that appear in the mature RNA species) are called exons, and noncoding regions, which interpose or intervene between the exons, are called introns. Introns always removed from precursors RNA before transport into the cytoplasm occurs. The process by which introns are removed from precursor RNA and by which exons are ligated together is called RNA splicing. Incorrect processing of the primary transcript into the mature mRNA can results in disease in humans. Regulatory regions for specific eukaryotic genes are usually located in the DNA that flanks the transcription initiation site at its 5’. Occasionally, such sequences are found within the gene itself or in the region that flanks the 3’ end of the gene. In mammalian cells, each gene has its own regulatory region. Many eukaryotic genes, and some viruses that replicate in mammalian cells, have special regions, called enhancers, that increase the rate of transcription. Some genes also have DNA sequences, known as silencers, that diminish transcription. Mammalian genes are obviously complicated, multicomponent structure.

In gene transcription, information generally flows from DNA to mRNA to protein. This is rigidly controlled process involving a number of complex steps, each of which no doubt is regulated by one or more enzymes or factors, faulty function at any of these steps can cause disease.

Concept use in recombinant DNA technology consists from isolation and manipulation of DNA, including end-to-end joining of sequences from very different sources to make chimeric molecules (eg. molecules containing both human and bacterial DNA sequences in a sequence independent fashion).  

### 6 Concept use in recombinant DNA

Concept use in recombinant DNA technology consists from isolation and manipulation of DNA, including end-to-end joining of sequences from very different sources to make chimeric molecules (eg. molecules containing both human and bacterial DNA sequences in a sequence independent fashion). This involves several unique techniques and reagents.

#### 6.1 Restriction enzymes

Certain endonucleases, enzymes that cut DNA as specific DNA sequences within the molecule, as opposed to exonucleases, which digest from the ends of DNA molecules, are a key tool in recombinant DNA research. These enzymes were originally called restriction enzymes because their presence in a given bacterium restricted the growth of certain bacterial viruses called bacteriophages. Restriction enzymes cut DNA into short pieces in a sequence specific manner, in contrast to most other enzymatic, chemical, or physical methods, which break DNA randomly. These defensive enzymes, over 200 have been discovered, protect the host bacterial DNA from DNA from foreign organisms, primarily infective phages. However, they are only present in cells that also have a companion enzyme that methylates the host DNA, rendering it an unsuitable substrate for digestion by the restriction enzyme. Thus, site-specific DNA methylases and restriction enzymes always exist in pairs in a bacterium.

Restriction enzymes are named after the bacterium from which they are isolated and EcoRI from *Escherichia coli*, Bam HI from *Bacillus amyloliquefaciens*. If the nucleotides are distributed randomly in a given DNA molecule, one can calculate how frequently a given enzyme would cut a length of DNA. For each position in the DNA molecule there are 4 possibilities A,C,G,T, therefore, a restriction enzyme that recognizes a 4 bp sequence will cut, on average, once every 256 bp (4^4), whereas another enzyme that recognizes a 6–bp sequence will cut once every 4^6 (4096) bp. A given piece of DNA will have a characteristic linear array of sites for the various enzymes, hence, a restriction map can be constructed. When DNA is digested with a given enzyme, the ends of all the fragments will have the same DNA sequence. The fragments produced can be isolated by electrophoresis on agarose or polyacrylamide, this is an essential step in cloning and a major use of these enzymes.

A number of other enzymes that act on DNA and RNA are an important part recombinant DNA technology.
7 Cloning
A clone is a large population of identical molecules, bacteria or cells that arise from a common ancestor. Cloning allows for the production of a large number of identical DNA molecules, which can then be characterized or used for other purposes. This technique is based on the fact that chimeric or hybrid DNA molecules can be constructed in cloning vector, typically bacterial plasmids, phages, or cosmids, which continue to replicate in a host cell under their own control systems. In this way, the chimeric DNA is amplified.

Bacterial plasmid are small, circular duplex DNA molecules whose natural function is to confer antibiotic resistance to the host cell. Plasmids have several properties that make them extremely useful as cloning vectors. They exist as a single or multiple copies within the bacterium and replicate independently from the bacterial DNA. The complete DNA sequence of many plasmids is known, hence, the precise location of restriction enzyme cleavage sites for inserting the DNA is available. Plasmids are smaller than the host chromosome and are therefore easily separated from the latter, and the desired DNA is readily removed by cutting the plasmid with the enzymes specific for the restriction site into which the original place of DNA was inserted.

Phages usually have linear DNA molecules into which foreign DNA can be inserted at several restriction enzyme sites. The chimeric DNA is collected after the phage proceeds through its lytic cycle and produces mature, infective phage particles. A major advantage of phage vectors is that while plasmids accept DNA pieces about 6-10 kb long, phages can accept DNA fragments 10-20 kb long, a limitation imposed by the amount of DNA that can be packed into the phage head.

Even larger fragments of DNA can be cloned in cosmids, which combine the best features of plasmids and phages. Cosmids are plasmids that contain the DNA sequences, so-called cos sites, required for packing lambda DNA into the phage particle. These vectors grow in the plasmid form in bacteria, but since much of the unnecessary lambda DNA has been removed, more chimeric can be packaged into the particle head. It is not unusual for cosmids to carry inserts of chimeric DNA that are 35-50 kb long.

8 Genomic library construction
The combination of restriction enzymes and various cloning vectors allows the entire genome of an organism to be packed into a vector. A collection of these different recombinant clones is called library. A genomic library is prepared from the total DNA of a cell line or tissue. A cDNA library represents the population of mRNAs in a tissue. Genomic libraries are prepared by performing partial digestion of total DNA with a restriction enzyme that cuts DNA frequently. Phage vectors are preferred for these libraries because they accept large pieces of DNA up to 20 kb. The goal is to achieve a complete library. The number of fragments required to attain this objective is inversely related to fragment size and directly related to genome size.

The composition of some complete genomic libraries are shown in the Table 1.

Table 1. The composition of the complete genomic libraries

<table>
<thead>
<tr>
<th>Source</th>
<th>Complete genomic library</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>1500 fragments</td>
</tr>
<tr>
<td>Yeast</td>
<td>4500 fragments</td>
</tr>
<tr>
<td>Drosophila</td>
<td>50000 fragments</td>
</tr>
<tr>
<td>Mammals</td>
<td>800000 fragments</td>
</tr>
</tbody>
</table>

The number of random fragments, unique clones that a library should have to ensure that any single gene is represented is inversely related to the average fragment size used to construct the library and directly related to the number of genes in the organism. The numbers given above represent the number of fragments, independent clones, necessary to achieve 99% probability of finding of a given DNA sequence in a recombinant DNA library with an average insert of 5x10^4 nucleotides. The differences represent the variations in genomic complexity between the creatures.

The number of clones necessary \( N \),

\[
N = \frac{\ln(1 - P)}{\ln(1 - f)}
\]

where \( P \) is the probability desired and \( f \) is the function of the total genome in a single clone.

In the case of mammalian genomic library given the presence of 3x10^9 nucleotides in the haploid genome, the equation is becoming:

\[
N = \frac{\ln(1 - 0.99)}{\ln[1 - (\frac{5 \times 10^3}{3 \times 10^9})]} = 1
\]

Information value will be defined [8]:

\[
I_c = -\sum_i P_i \log P_i
\]
where $P_i$ probability of finding $i$-th DNA sequence.

cDNA libraries are prepared by first isolating all the mRNAs in a tissue and then copying these molecules into double stranded DNA, using, sequentially, the enzymes reverse transcriptase and DNA polymerase. Plasmids are often the favored vectors for cDNA libraries because they are much more convenient to work with than are phages or cosmids, although various lambda phage vectors have special advantages for cDNA cloning.

A vector in which the protein coded by the gene introduced by recombinant DNA technology is actually synthesized is known as an expression vector. Such vectors are now commonly used to detect specific cDNA molecules in libraries and to produce proteins by genetic engineering techniques.

A variety of molecules can be used to “probe” libraries in search of a specific gene or cDNA molecule or to define and quantitate DNA or RNA separated by electrophoresis through various gels. Probes are generally pieces of DNA or RNA labeled with a $^32$P-containing nucleotide. The probe must recognize a complementary sequence to be effective. A cDNA synthesized from a specific mRNA can be used to screen either a cDNA library for a longer cDNA or a genomic library for a complementary sequence in the coding region of a gene. A popular technique for finding specific gene entails taking a short amino acid sequence and, using the codon usage for that species, making an oligonucleotide probe that will detect the corresponding DNA fragment in a genomic library. If the sequence match exactly, probes 15-20 nucleotides long will hybridize.

Visualization of a specific cDNA or RNA fragment among the many thousands of “contaminating” molecules requires the convergence of a number of techniques, which are collectively termed blot transfer.

Colony or plaque hybridization is the method by which specific clones are identified and purified. Bacteria are grown on colonies on an agar plate and overlaid with a nitrocellulose filter paper. Cells from each colony stick to the filter and are permanently fixed thereto by heat, which with sodium hydroxide treatment also lyses the cells and denatures the DNA so that it will hybridize with the probe.

9 DNA Sequencing

The segments of specific DNA molecules obtained by recombinant DNA technology can be analyzed for their nucleotide sequence. This method depends upon having a large number of identical DNA molecules. This requirement can be satisfied by cloning the fragment of interest, using the known techniques.

The isolation of a specific gene from an entire genome requires a technique that will discriminate one part in a million. The identification of a regulatory region that may be only 10bp in length requires a sensitivity of one part in $3 \times 10^8$, a disease such as sickle cell anemia is caused by a single base change, or one part in $3 \times 10^9$. Recombinant DNA technology is powerful enough to accomplish all of these things.

Gene mapping is technique which is used to localize specific genes to distinct chromosomes and, to define a map of the human genome. This is already yielding useful information in the definition of human disease. Somatic cells hybridization and in situ hybridization are 2 techniques used to accomplish this.

A practical goal of recombinant DNA research is the production of materials for biomedical application.

There is a normal variation of DNA sequence just as there is with more obvious aspects of human structure.

10 Conclusion

Carbamyl phosphate synthase is sensitive to feedback inhibition by both purine and pyrimidine nucleotides and activation by PRPP.

Variations of DNA sequence, polymorphisms, occur approximately once in every 500 nucleotides, or about $10^7$ times per genome. The nucleotides are distributed randomly in a given DNA molecule and DNA sequence information value was derived.

Notation

$f$-function of the total genome in a single clone

$N$-number of clones

$P$- probability of the sequence appear

Abbreviation

AMP- adenine monophosphate
ADP- adenine diphosphate
ATP- adenine triphosphate
CTP- cytosine triphosphate
DNA-deoxyribonucleic acid
DP-diphosphate
GDP- guanine diphosphate
GMP-guanine monophosphate
GTP- guanine triphosphate
HGPR Tase-hypoxanthine-guanine phosphoribosyl- transferase
IMP- monophosphate
IP- monononophosphate
NADPH-cofactor
OMP- orotidine monophosphate
PRPP- phosphoribosyl-pyrophosphate
RNA-ribonucleic acid
TP- triphosphat
UMP- uridine monophosphate  
UTP- uridine triphosphate  

References  