A TEXT BOOK OF

BIOTECHNOLOGY

CLASS - XII

CENTRAL BOARD OF SECONDARY EDUCATION
PREET VIHAR, DELHI - 110092

Biotechnology : Class - XII

June - 2011 : Second Edition
No. of Copies : 1,50,000

Paper used : 70 gsm CBSE watermark Maplitho Paper

Price : Rs.

PUBLISHED BY : The Secretary, Central Board of Secondary Education, Shiksha Kendra, 2, Community Centre, Preet Vihar, Delhi - 110301

DESIGN & LAYOUT : Multi Graphics, 5745/81, Reghar Pura, Karol Bagh, New Delhi - 110005.

PRINTED BY :

First Edition

First Edition: March 2003

ADVISORY COMMITTEE

- Mr Ashok Ganguly, Chairman, CBSE
- Mr G. Balasubramanian, Director (Academic), CBSE
- Prof. K. Kannan, Convenor, CBSE Committee of Courses in Biotechnology

EDITORS

- Prof. K. Kannan, Dean, School of Biotechnology, Guru Gobind Singh lndraprastha University, Delhi
- Dr K. Nirmala, Reader, Department of Biochemistry, Daulat Ram College, University of Delhi, Delhi

CONTRIBUTORS

- Prof. K. Kannan, Dean, School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi (Convenor, CBSE Committee of Courses in Biotechnology)
- Dr Vijay Kumar, Senior Research Scientist, International Centre of Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi
- Dr V.L. Kumar, Additional Professor, Department of Pharmacology, All India Institute of Medical Science, New Delhi.
- Dr M.V. Rajam, Reader, Department of Genetics, University of Delhi, South Campus, New Delhi
- Dr Amulya K Panda, Scientist, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi
- Dr J.S. Virdi, Reader, Department of Microbiology, University of Delhi, South Campus, New Delhi
- Dr S. Ramachandran, Scientist, Institute of Genomics & Integrative Biology, Mall Road, Delhi
- Dr Rajesh S. Gokhale, Scientist, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi
- Dr Pushkar Sharma, Scientist, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi
- Dr K.K. Aggarwal, Lecturer, School of Biotechnology. Guru Gobind Singh Indraprastha University, Delhi
- Dr Ashwani Pareek, Lecturer, School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi
- Dr Nimisha Sharma, Lecturer, School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi
- Dr Promila Gupta, Lecturer, School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi
- Mr Suresh Kumar, Lecturer, School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi

DESIGN, LAYOUT & ILLUSTRATIONS :

Dr Ashwani Pareek, Mr. Ramesh Chand Sharma, Ms. Tannistha Nandi, School of Biotechnology, Guru Gobind Singh Indraprastha University, Delhi

PROJECT CO-ORDINATOR :

Mr R.P. Sharma, Education Officer (Science), CBSE

PUBLISHED BY:

Mr B.R. Sharma, Secretary, Central Board of Secondary Education, 'Shiksha Kendra', 2, Community Centre, Preet Vihar, Delhi - 110 092

Second Edition

Second Edition: June 2011

PUBLISHED BY

Secretary, Central Board of Secondary Education, 'Shiksha Kendra', 2, Community Centre, Preet Vihar, Delhi - 110 092

Foreword To The First Edition

The Central Board of Secondary Education has been constantly striving for a qualitative improvement of education at school level for the past many years. It has to its credit many educational innovations. Introduction of new contemporary areas to school curriculum and updation of its syllabi in different subjects from time to time in the form of 'Frontline Curriculum' is one of the examples of many such innovations. The subject of Biotechnology has been introduced as an elective subject at senior secondary level from the academic session 2002-2003 as a follow up of this approach.

The Board has already published its own textbook as well as laboratory manual in the subject for Class XI during the preceding academic session. The present document aims at meeting the requirements of different content areas included in Class XII syllabus. Some of the major areas discussed in this publication include Protein Structure and Engineering, Recombinant DNA Technology, Genomics and Bioinformatics, Microbial Culture, Plant Cell Culture, Animal Cell Culture and their applications. Every care has been taken to lay due emphasis on applications of basic concepts in the subject to daily life situations, particularly in relation to agriculture, medicine, industry and environment. The complete set of instructional package also includes another document on laboratory manual for Class XII.

Eminent subject experts from Guru Gobind Singh Indraprastha University, University of Delhi, International Centre for Genetic Engineering and Biotechnology, All India Institute of Medical Sciences, National Institute of Immunology and Institute of Genomics and Integrative Biology have contributed to the writing of the original manuscript. The writing group, under the leadership of Prof. K. Kannan, Dean, School of Biotechnology, GGSIP University has worked overtime to develop the material in stipulated time. I would like to record my deep appreciation and sincere thanks to all the members of the writing group for their invaluable contribution in development of this material. I shall also thankfully acknowledge the untiring efforts put in by Prof. K. Kannan and Dr. K. Nirmala in editing the whole material and making it print worthy.

Shri G. Balasubramanian, Director (Academic), CBSE, deserves special appreciation for his innovative ideas and intellectual perception of the whole project. Deep appreciation is also due to Shri R.P. Sharma, Education Officer (Science) for his total involvement and commitment in coordinating different activities related to this publication.

Suggestions from users for further improvement of the document will be highly appreciated.

Ashok Ganguly

Chairman, CBSE

Foreword To The Second Edition

The Central Board of Secondary Education, in its intent to facilitate growth through education at school level endeavours to update curriculum in various emerging areas of study. In this light Biotechnology was introduced as an elective subject at senior secondary school level in the academic session 2002- 2003.

Text book is an extension of the curriculum initiatives. It serves as a ready-to-access resource material, both for the teachers and the students.

This edition of Biotechnology text book for class XII has been revised in order to keep pace with the galloping advancement in technology on one front and increasing role of Biotechnology to manipulate and promote a productive and healthy life.

The demand from the teachers and the students through the website of the Board provided important feedback and need to revise the text book. In the present edition, though the scheme of chapters remains the same, there are inclusions that have a certain edge over the previous work for better conceptual connections, some of these are in terms of diagrammatic presentations and explanations.

The team of experts under the convenor Prof. M. V. Rajam, Department of Genetics, South Campus, University Of Delhi, invested their expertise, experience and willingness to revise the book. The effort of each member: Prof M.V. Rajam, Prof. J.S. Virdi, Dr. K. Nirmala and Dr. S. Ramachandran is appreciatively acknowledged by the Board. The work once taken up was efficiently owned up and zestfully completed.

The efforts of Sh. P. V. Sai Ranga Rao, former Education Officer (Science) for the inception of the whole idea and providing the necessary impetus and Dr. Rashmi Sethi, Education Office (Science) to coordinate and give final shape to the manuscript to its worthwhile completion, deserve a special mention as a team work.

The Board invites observations and suggestions from the readers to further upgrade the content and presentation of the text book in future as well.

VINEET JOSHI

CHAIRMAN

Preface To The First Edition

A modern Biotechnologist does simple mathematical operations such as addition, subtraction, multiplication or division of bio-molecules in a cell to produce products of commercial value. One adds, subtracts, multiplies or divides a gene, which in turns changes the levels of a protein, which may in turn change the level of carbohydrates and lipids. A classical biotechnologist first used natural bio-factories (cells) to make products such as Curd, Ethanol, Penicillin, Insulin etc. Modern biotechnology on the contrary is making cells-do jobs which they have never done before. An instance being bacteria producing human insulin. Should biotechnology teaching-learning be incorported at the senior secondary school level? Recall the impact created by the automobile industry in the 20th century. Till about 1930, most of the day to day used products were derived from plants, animals or microbes. Every new technological development brings about a change in life style. The advent of automobile industry brought in a sea of change in our life styles leading to the discovery of better and cheaper fuel such as petrol from oil fields. We saw the growth of the petroleum industry. This consequently produced several byproducts from oil. These petroleum-derived byproducts being cheaper, led to the development of technologies from these byproducts as raw materials during the period 1930's to 1960's. Chemical Engineers and Technologists created a whole range of synthetic products for daily use from petroleum byproducts. By the end of 20th century, 95% of the products of daily use were derived from petroleum byproducts and hardly 1 % from plants, animals or microbes. The reason was that the raw materials derived from nature were not economically viable to make products. Further no efforts were made to develop cheaper technologies from natural raw materials.

With the sudden increase in oil prices from 1973 onwards, natural products became cheaper as raw materials than oil based byproducts. Further, it was realized that oil is an unsustainable source in the long run whereas plants, animals and microbes (PAM) are everlasting. Thus in the scheme of sustainable development, plants, animals and microbes acquired special importance as well as helped in better environment management. USA has taken a policy decision that 20% of the products of daily use will be made from raw materials derived from PAM by 2020 and 50% by 2050. Indian economy heavily relies on oil and we incur billions of dollars in foreign exchange. Our nation has been bestowed with a rich biodiversity waiting to be explored for sustainable economic prosperity. Biotechnology holds the greatest promise for sustainable economic growth and make India a developed nation.

A Textbook of Biotechnology for class XII has been written with the hope that we will create a new breed of Biotechnologist, who is sensitive to the issue of economic slavery. The Textbook has been divided into two units and is a continuation of class XI Textbook.

We sincerely hope that the units covered in this book will enthuse the young minds who read them to think differently about Biology and the role it will play in fostering a new Biotechnology Culture.

Any textbook requires meticulous and painstaking efforts by all its contributors. We are grateful to all the contributors who have cheerfully revised and re-revised their sections to keep the subject level interesting and readable at the class XII level. In addition, several other have made contributions in reading and recommending changes wherever necessary and these include students from the School of Biotechnology, GGSIPU. Special thanks are due to Anjana, Neeraj, Bhavana, Shruti, Ritu, Nitika, Amit, Yasha, Archana, Aditya and Deepak Grover. We would like to acknowledge the assistance of Mr Sanjeev Kumar and Mr Amreesh Kumar in preparation of the book.

The Convener and the Resource team are ever so grateful to Mr. Ashok Ganguly, Chairman CBSE, Mr. G Balalsubramanian, Director Academic and Mr. R P Sharma, Education Officer (Science), CBSE whose complete faith in the team has had an undoubted influence on much of the content. Without the help of these people, the book could not have seen the light of the day. I can only hope that they will all look upon the results of the influence and endeavor with pleasure. Wherever possible, credit has been given and if I have failed to acknowledge the source of an idea or a technique, it is because the source for some reason is unknown to me. In the rapidly expanding world of biotechnology, it is often impossible to know where the idea really originated.

We hope that all the budding biotechnologists will fulfill our dream and convert India into a sustainable developed nation.

Prof. K. Kannan

Convenor

CBSE Committee of Courses in Biotechnology

Preface To The Second Edition

The text book of Biotechnology was first introduced to the Senior Secondary School students to provide an overview of many of the fundamentals of Biology and Biotechnology as well as the potential applications of biotechnology in industry, health-care and agriculture, and to make them understand the impact of biotechnology on human welfare. This text book of Class XII Biotechnology was first published in March 2003. It has been nearly eight years since the first edition of this book appeared and since then much has happened in biotechnology in these intervening years. In fact, the developments in this area have radically altered our concepts of health-care and agriculture and this has been the main reason to bring out the second edition of this book. A collaborative effort has been made to present as updated content along with visual presentation for a better comprehension by the readers. We have condensed, improved and elaborated here and there, improved or replaced the existing diagrams and photos and added some new pictures, and done overall improvement of the book. Even though the various units and chapters remain, more or less, the same as in the first edition.

In preparing this second edition, I am grateful to the authors, Prof. J. S. Virdi, Dr. K. Nirmala, Dr. Vijay Kumar and Dr. S. Ramachandran for their academic contributions and support. I would also like to thank all the contributing authors of the first edition of this book, and a special thanks to Prof. K. Kannan, the former Convenor of the CBSE Committee of Courses in Biotechnology for his initiative and active participation in the first edition of this book. My sincere thanks are due to the former Chairman (Ashok Ganguly), Director (Dr. G. Balasubramanian) and Education Officer (Mr. R. P. Sharma) and the present Chairman (Mr. Vineet Joshi), Director (Mrs. Chitralekha Gurumurthy) and Education Officers (Science), Mr. P.V. Sai Ranga Rao and Dr. Rashmi Sethi, CBSE, New Delhi for constant support to bring out this book. I thank the Multigraphics for their meticulous work in producing this book.

I am indebted to some Biotechnology school teachers, particularly Ms. Vaishali Aggarwal, SLS DAV Public School, Mausam Vihar, Delhi-51, Ms. Shalu Bajaj and Dr. Supriya Sharma, Manav Sthali School, New Rajendra Nagar, New Delhi, for reviewing the chapters and providing valuable suggestions, which were important for the improvement of the book. Special thanks to Ms. V. Vineela for her help in preparing some diagrams for a chapter on rDNA tecnology.

Last but not least, I am grateful to my wife Padma for her loving support.

Suggestions for the improvement of this book would be gratefully acknowledged.

Prof. M. V. Rajam

Convenor

CBSE Committee of Courses in Biotechnology

THE CONSTITUTION OF INDIA

PREAMBLE

WE, THE PEOPLE OF INDIA, having solemnly resolved to constitute India into a [1] **SOVEREIGN SOCIALIST SECULAR DEMOCRATIC REPUBLIC** and to secure to all its citizens :

JUSTICE, social, economic and political;

LIBERTY of thought, expression, belief, faith and worship;

EQUALITY of status and of opportunity; and to promote among them all

FRATERNITY assuring the dignity of the individual and the [2] [unity and integrity of the Nation];

IN OUR CONSTITUENT ASSEMBLY this twenty-sixth day of November, 1949, do **HEREBY TO OURSELVES THIS CONSTITUTION.**

1. Subs, by the Constitution (Forty-Second Amendment) Act. 1976, sec. 2, for "Sovereign Democratic Republic (w.e.f. 3.1.1977)
2. Subs, by the Constitution (Forty-Second Amendment) Act. 1976, sec. 2, for "unity of the Nation (w.e.f. 3.1.1977)

THE CONSTITUTION OF INDIA

Chapter IV A

Fundamental Duties

ARTICLE 51A

Fundamental Duties - It shall be the duty of every citizen of India-

(a) to abide by the Constitution and respect its ideals and institutions, the National Flag and the National Anthem;

(b) to cherish and follow the noble ideals which inspired our national struggle for freedom;

(c) to uphold and protect the sovereignty, unity and integrity of India;

(d) to defend the country and render national service when called upon to do so;

(e) To promote harmony and the spirit of common brotherhood amongst all the people of India transcending religious, linguistic and regional or sectional diversities; to renounce practices derogatory to the dignity of women;

(f) to value and preserve the rich heritage of our composite culture;

(g) to protect and improve the natural environment including forests, lakes, rivers, wild life and to have compassion for living creatures;

(h) to develop the scientific temper, humanism and the spirit of inquiry and reform;

(i) to safeguard public property and to abjure violence;

(j) to strive towards excellence in all spheres of individual and collective activity so that the nation constantly rises to higher levels of endeavour and achievement.

भारत का संविधान

उद्देशिका

हम, भारत के लोग, भारत को एक '[सम्पूर्ण प्रभुत्व-संपन्न समाजवादी पंथनिरपेक्ष लोकतंत्रात्मक गणराज्य] बनाने के लिए, तथा उसके समस्त नागरिकों को:

सामाजिक, आर्थिक और राजनैतिक न्याय,
विचार, अभिव्यक्ति, विश्वास, धर्म
और उपासना की स्वतंत्रता,
प्रतिष्ठा और अवसर की समता

प्राप्त कराने के लिए, तथा उन सब में, व्यक्ति की गरिमा और [2] [राष्ट्र की एकता और अखण्डता] सुनिश्चित करने वाली बंधुता बढ़ाने के लिए दृढ़संकल्प होकर अपनी इस संविधान सभा में आज तारीख 26 नवम्बर, 1949 ई॰ को एतद्द्वारा इस संविधान को अंगीकृत, अधिनियमित और आत्मार्पित करते हैं।

1. संविधान (बयालीसवां संशोधन) अधिनियम, 1976 की धारा 2 द्वारा (3.1.1977) से "प्रभुत्व-संपन्न लोकतंत्रात्मक गणराज्य" के स्थान पर प्रतिस्थापित।
2. संविधान (बयालीसवां संशोधन) अधिनियम, 1976 की धारा 2 द्वारा (3.1.1977 से), "राष्ट्र की एकता" के स्थान पर प्रतिस्थापित।

भाग 4 क

मूल कर्त्तव्य

51 क. मूल कर्त्तव्य – भारत के प्रत्येक नागरिक का यह कर्त्तव्य होगा कि वह –

(क) संविधान का पालन करे और उसके आदर्शों, संस्थाओं, राष्ट्रध्वज और राष्ट्रगान का आदर करे;

(ख) स्वतंत्रता के लिए हमारे राष्ट्रीय आंदोलन को प्रेरित करने वाले उच्च आदर्शों को हृदय में संजोए रखे और उनका पालन करे;

(ग) भारत की प्रभुता, एकता और अखंडता की रक्षा करे और उसे अक्षुण्ण रखे;

(घ) देश की रक्षा करे और आह्वान किए जाने पर राष्ट्र की सेवा करे;

(ङ) भारत के सभी लोगों में समरसता और समान भ्रातृत्व की भावना का निर्माण करे जो धर्म, भाषा और प्रदेश या वर्ग पर आधारित सभी भेदभाव से परे हों, ऐसी प्रथाओं का त्याग करे जो स्त्रियों के सम्मान के विरुद्ध हैं;

(च) हमारी सामाजिक संस्कृति की गौरवशाली परंपरा का महत्त्व समझे और उसका परीक्षण करे;

(छ) प्राकृतिक पर्यावरण की जिसके अंतर्गत वन, झील, नदी, और वन्य जीव हैं, रक्षा करे और उसका संवर्धन करे तथा प्राणिमात्र के प्रति दयाभाव रखे;

(ज) वैज्ञानिक दृष्टिकोण, मानववाद और ज्ञानार्जन तथा सुधार की भावना का विकास करे;

(झ) सार्वजनिक संपत्ति को सुरक्षित रखे और हिंसा से दूर रहे;

(ञ) व्यक्तिगत और सामूहिक गतिविधियों के सभी क्षेत्रों में उत्कर्ष की ओर बढ़ने का सतत प्रयास करे जिससे राष्ट्र निरंतर बढ़ते हुए प्रयत्न और उपलब्धि की नई उंचाइयों को छू ले।

Contents

UNIT 5 : PROTEIN AND GENE MANIPULATION

Contents

UNIT 6 : CELL CULTURE AND GENETIC MANIPULATION


CBSE
CONTINUOUS AND COMPREHENSIVE EVALUATION
KNOW - AS YOU GROW

UNIT 5

PROTEIN AND GENE MANIPULATION

CHAPTER

1

RECOMBINANT DNA TECHNOLOGY

5.1.1. Introduction

Every day the newspapers tell you about the remarkable feats of science and technology in helping man combat disease and improve his environment. Many of these involve the use of a powerful technique called gene cloning or Recombinant DNA Technology (RDT). Using such technology bacteria in the past were engineered to produce human insulin a hormone which fights diabetes, yeast cells were made to produce Hepatitis B vaccine, plants such as cotton were made insect resistant (Bt-cotton) and even as you read this chapter projects are on to engineer bacteria to cleanup environmental waste such as polythene. Have you not wondered how Scientists are able to achieve all this? The present chapter will introduce you to the main techniques used in gene cloning along with some important applications for you to understand and marvel at the simplicity and power of this area of biotechnology.

The basic steps involved in RDT are illustrated schematically below **Fig. 1** :

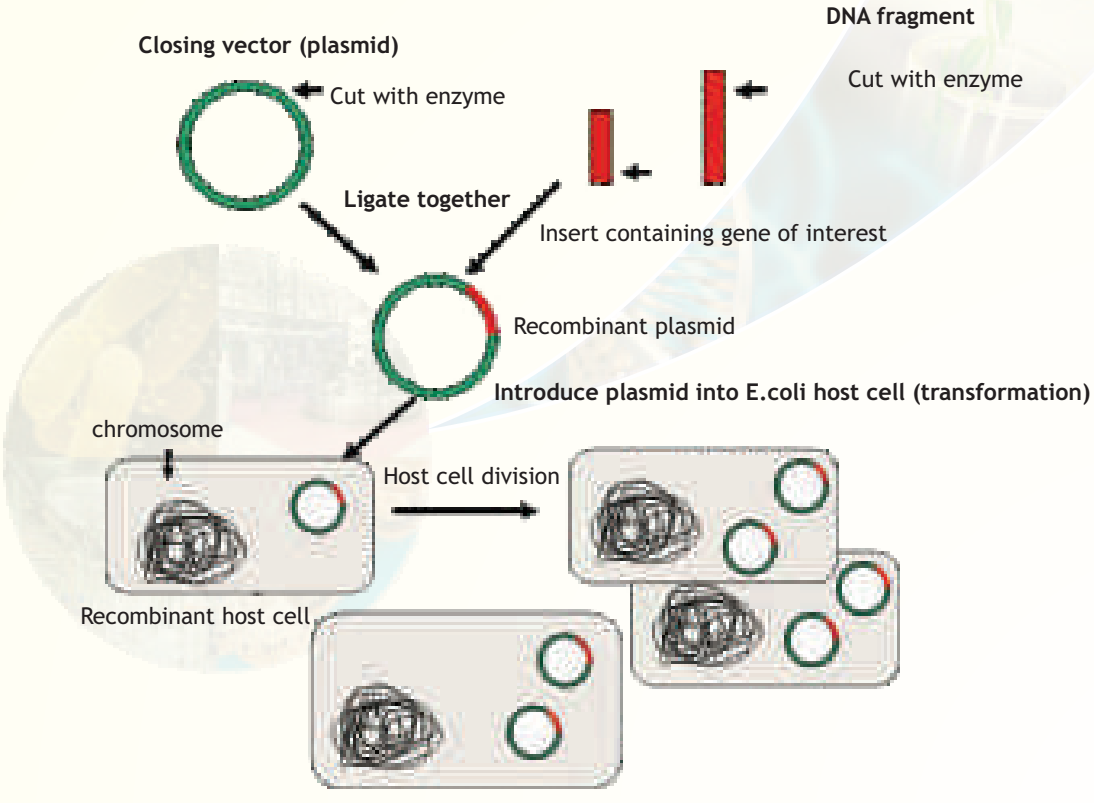


Fig. 1. Schematic representation of the basic steps in RDT.

The steps involved are:

1. Isolation of a DNA fragment containing a gene of interest that needs to be cloned (called as **insert**).
2. Generation of a recombinant DNA (rDNA) molecule by insertion of the DNA fragment into a carrier DNA molecule called **vector** (e.g. plasmid) that can self replicate within a host cell.
3. Transfer of the rDNA into an *E. coli* host cell (process called **transformation**).
4. Selection of only those host cells carrying the rDNA and allowing them to multiply thereby multiplying the rDNA molecules.

The whole process thus can generate either a large amount of rDNA (**gene cloning**) or a large amount of protein expressed by the insert. The first rDNA molecules to be generated using these procedures were established by the combined efforts in 1973 by the molecular biologists Paul Berg, Herbert Boyer, Annie Chang and Stanley Cohen.

5.1.2. Tools of rDNA technology

In order to generate recombinant DNA molecules, we not only require the vector and insert DNA but also a method to precisely cut these DNA molecules and then join them together **(ligation)**. Several molecular tools are required to perform the various steps and these are described in the following sections.

Restriction Enzymes: The Molecular Scissors

The foundations of rDNA technology were laid by the discovery of restriction enzymes. These enzymes exist in many bacteria where they function as a part of a defence mechanism called the Restriction-Modification System. This System consists of two components:

1. A restriction enzyme that selectively recognises a specific DNA sequence and digests any DNA fragment containing that sequence. The term restriction is derived from the ability of these enzymes to restrict the propagation of foreign DNA (*e.g.* Viral/phage DNA) in a bacterium.
2. A modification enzyme that adds a methyl group to one or two bases within the sequence recognised by the enzyme. Once a base is modified by methylation, the sequence cannot be digested. It is thus obvious that the Restriction-Modification enzyme system within a given bacterium protects its DNA from digestion by methylation but can digest foreign DNA which is not protected by similar methylation.

Different species of bacteria contain their own sets of restriction endonucleases and corresponding methylases. Three main classes of restriction endonucleases- type I, type II and type III are present, of which, only type II restriction enzymes are used in rDNA technology as they

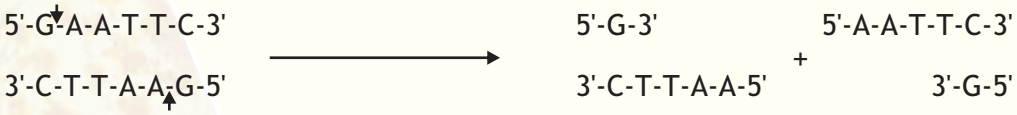
recognise and cut DNA within a specifc sequence typically consisting of 4-8 bp. This sequence is referred to as a restriction site and is generally palindromic, which means that the sequence in both DNA strands at this site read same 5' to 3' direction. Type II restriction enzymes are named after the bacterial species they have been isolated from. For example a commonly used restriction enzyme *Eco*RI isolated from the bacterial species *E. coli* is named so with the first three italicised alphabets referring to the genus (E) and species (co), the capital R referring to the strain (RY 13) and the number designated with the roman numeral (I) indicating that it was the first enzyme to be isolated from this strain of bacteria. Restriction enzymes were first discovered and studied by the molecular biologists W. Arber, H. Smith and D. Nathans for which they were awarded the Nobel Prize in 1978. Today more than a thousand restriction enzymes are available for genetic engineers to use. Some commonly used restriction enzymes (type II) along with their source and restriction- modification sequences have been listed in **Table1** below.

Table 1. Type II restriction enzymes, their sources, recognition and cleavage sites.

Restriction enzyme	Microbial source	Recognition sequence
Alu I	*Arthrobacter luteus*	5'A-G↓-C-T 3' 3'T-C-G-↑A 5'
*Bam*HI	*Bacillus amyloliquefaciens*	5'G↓-G-A-T-C-C 3' 3'C-C-T-A-G-↑G 5'
*Eco*RI	*Escherichia coli*	5'G↓-A-A-T-T-C 3' 3'C-T-T-A-A-↑G 5'
*Eco*RII	*Escherichia coli*	5'↓C-C-T-G-G 3' 3'G-G-A-C-↑C 5'
*Hae*III	*Haemophilus aegyptus*	5'G-G↓-C-C 3' 3'C-C-↑G-G 5'
*Hind*III	*Haemophilus influenza*	5'A↓-A-G-C-T-T 3' 3'T-T-C-G-↑A-A 5'
*Pst*I	*Providencia stuartii*	5'C-T-G-C↓-A-G 3' 3'G-↑A-C-G-T-C 5'
*Sal*I	*Streptomyces albus*	5'G↓-T-C-G-A-C 3' 3'C-A-G-C-↑T-G 5'

The exact kind of cleavage produced by a restriction enzyme is important in the design of a gene cloning experiment. Some cleave both strands of DNA through the centre resulting in a blunt or flush end. These are also known as symmetrical cuts. From **Table I** it is obvious that the enzyme *AluI* cuts symmetrically.

However *Eco*RI cuts in a way producing protruding and recessed ends known as sticky or cohesive ends because these ends can base pair and stick the DNA molecule back together again. Such cuts are termed staggered.



Note that *Eco*RI generates 5' overhangs at the cut site (and 3' recessed ends). Hence if two different DNA fragments containing *Eco*RI recognition sites are cleaved and mixed the sticky ends can bind and generate a hybrid or recombinant DNA **(Fig.2)**

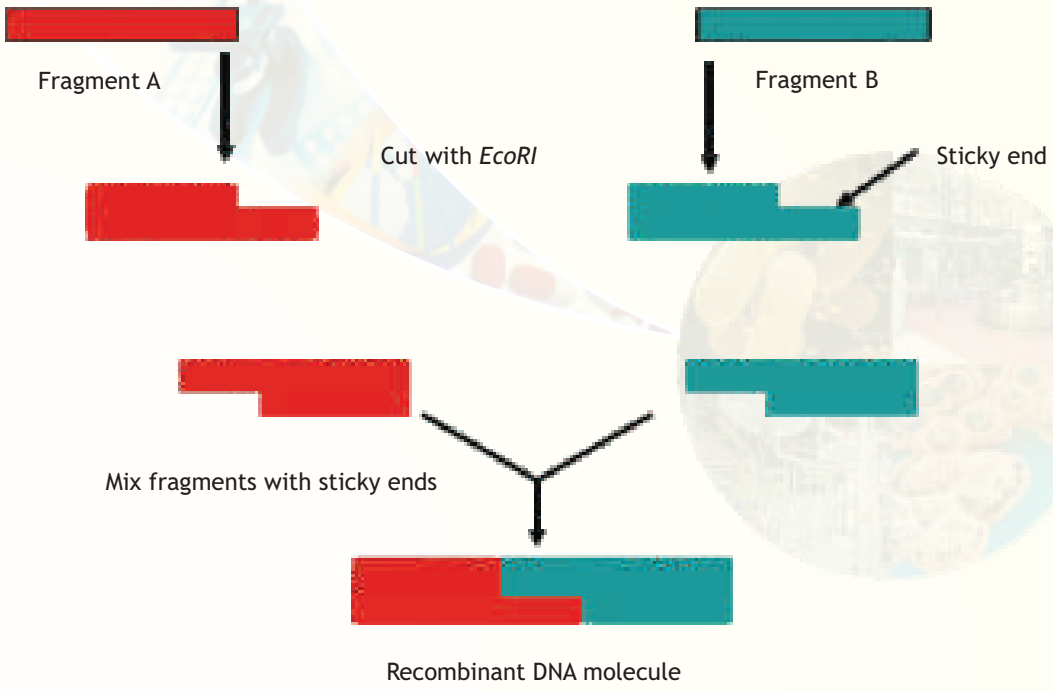


Fig. 2. Construction of rDNA using fragments from different sources.

Restriction Fragment Length Polymorphism (RFLP)

The DNA isolated from an individual organism has a unique sequence and even the members within a species differ in some parts of their sequence. The restriction sites would also vary and hence if DNA from a given individual was subjected to digestion with a restriction enzyme the fragments generated would vary when compared with another individual's DNA similarily digested. This variation in size (length) of the restriction enzyme generated fragments among individuals within a given species is termed RFLP. A schematic representation of how RFLPs are generated is given below **(Fig. 3)**. A major application of this technique is DNA fingerprinting analysis which is explained below.

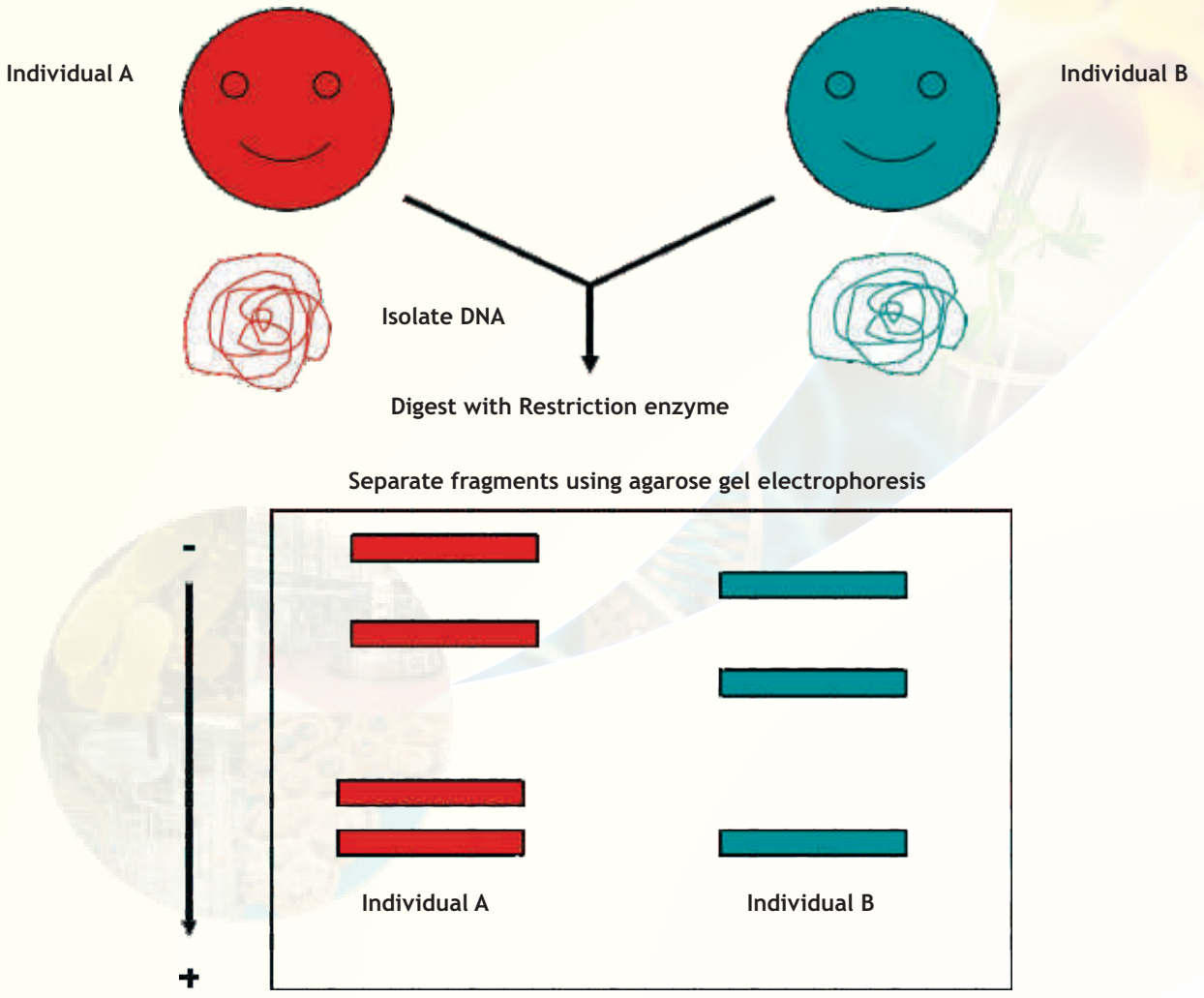


Fig. 3. RFLP technique.

Individuals except identical twins vary in their RFLP pattern as indicated schematically in the agarose gel electrophoresis. Hence the term DNA fingerprint is used and this is the basis of a major technique used in forensic science to identify and relate individuals.

Other enzymes used in cloning

In addition to restriction enzymes, there are several other enzymes that play an important role in rDNA technology. Two of these are DNA ligase and alkaline phosphatase.

DNA ligase: This enzyme forms phosphodiester bonds between adjacent nucleotides and covalently links two fragments of DNA . The reaction requires one of the fragments to have a 5' phosphate residue and the other a 3' hydroxyl group. In a previous section it was indicated how two fragments cut with *Eco*RI could stick together; DNA ligase seals this by forming a covalent bond. DNA ligase isolated from the bacteriophage T4 is frequently used to ligate different DNA fragments in order to generate rDNA molecules.

Alkaline phosphatase: Ligation requires the presence of a 5'phosphate group. If some of the fragments are treated with alkaline phosphatase to remove their phosphate groups then these cannot ligate within themselves and are forced to ligate with other fragments containing 5'phosphate groups. Hence this is a useful strategy to prevent self-ligation which would otherwise lead to wasteful ligation of fragments treated with restriction enzymes. An insert is ligated to the vector in generating rDNA as the vector is prevented from self-ligation by treating it with alkaline phosphatase. Alkaline phosphatase used for this purpose is purified from bacteria or calf intestines.

Vectors: Vehicles for cloning

Another major component of a gene cloning experiment is a vector such as a plasmid. A vector serves as a vehicle to carry a foreign DNA sequence into a host cell. A vector must possess certain features:

1. It must contain an origin of replication (ori) so that it is independently able to replicate within the host. This implies that any foreign insert it carries is automatically replicated.
2. It should incorporate a selectable marker, a gene whose product can identify the host cells containing the vector. Selectable markers include genes conferring antibiotic resistance, enzymes such as β-galactosidase which can turn substrates blue in the vicinity of the host cell colony and gene expressing Green Fluorescent Protein (GFP) which cause host cells containing the vector to fluoresce when viewed under UV light.

3. The vector must also have one unique restriction enzyme recognition site which can be used for cutting and introducing an insert. Most of the commonly used cloning vectors have more than one restriction site, they contain a Multiple Cloning Site (MCS) or polylinker. The MCS provides flexibility in the choice and use of restriction enzymes.

4. Another desirable feature of a cloning vector is that it should be small in size thereby facilitating entry/transfer into a host cell.

   A number of vectors have been developed incorporating these features but only plasmids and bacteriophage based vectors will be discussed.

Plasmids

Plasmids are extrachromosomal, self-replicating, usually circular, double-stranded DNA molecules found naturally in many bacteria and also in some yeasts. Although plasmids are not essential for normal cell growth and division, they often confer useful properties to the host such as resistance to antibiotics that can be a selective advantage under certain conditions. The plasmid molecules can be present as 1 or 2 copies or in multiple copies (500-700) inside the host organism. These naturally occurring plasmids have been modified to serve as vectors in the laboratory and are by far the most widely used, versatile and easily manipulated vectors.

One of the earliest plasmid vectors to be constructed was pBR322 **(Fig. 4a)**. This plasmid contains two different antibiotic resistance genes and recognition sites for several restriction enzymes. A popular series of plasmid cloning vectors is the pUC family **(Fig. 4b)**. These vectors contain a region of the *lacZ* gene that codes for the enzyme β-galactosidase. This region also contains a polylinker and thus insertion of a foreign DNA into any of the restriction sites will result in an altered non-functional enzyme. During screening of recombinant plasmid containing host cells the absence of β-galactosidase activity is indicative of plasmids containing the insert.

The plasmid vectors described above can replicate only in *E. coli*. Many of the vectors used in eucaryotic cells are constructed such that they can exist both in the eukaryotic cell and *E. coli*. Such vectors are known as shuttle vectors. These vectors contain two types of origin of replication and selectable marker genes, one set which functions in the eukaryotic cells (*e.g.* yeast) and another in *E. coli*. An example of a shuttle vector is the yeast plasmid Yep **(Fig. 4c)**. In the case of plants, a naturally occurring plasmid of the bacterium *Agrobacterium tumefaciens* called Ti plasmid has been suitably modified to function as a vector.

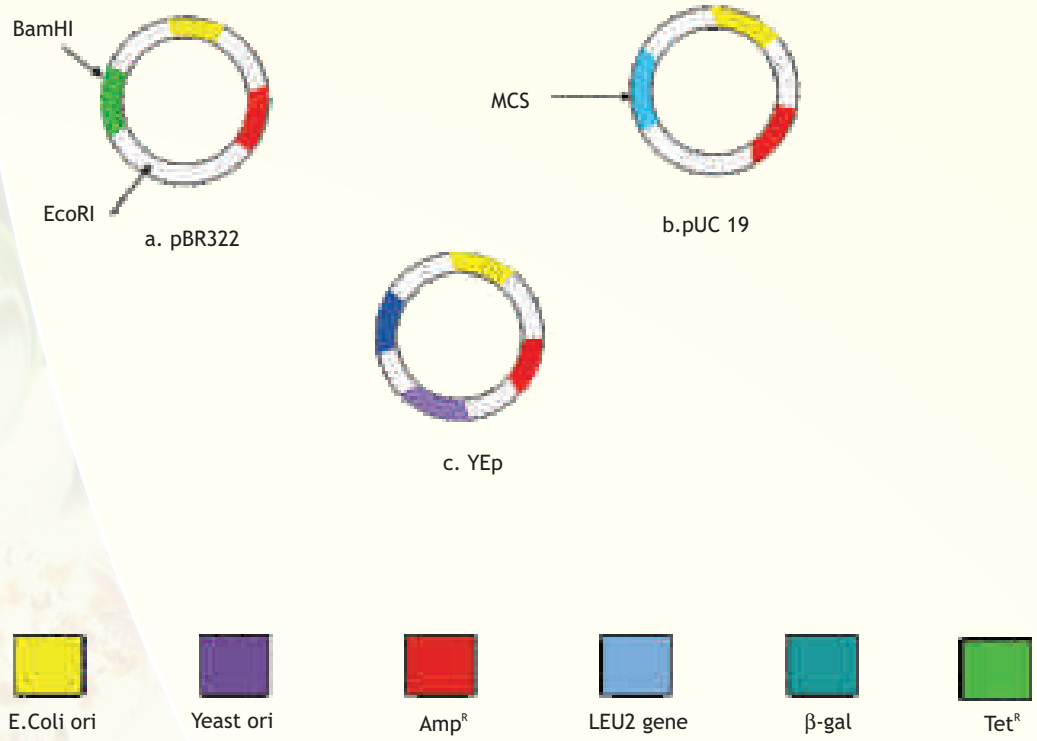


Fig. 4. Schematic representation of some plasmids (not drawn to scale).

The *LEU2* gene (see Yep) codes for an enzyme which is needed for the synthesis of the amino acid leucine. Yeast cells having this plasmid can grow on a medium lacking leucine and hence can be selected over cells not containing the plasmid.

The discussion so far has focussed on cloning a DNA fragment to sufficiently amplify it. However the goal of the cloning experiment may be to produce a foreign protein in the host. In such cases the fragment containing the gene expressing the foreign protein is incorporated into the vector along with signals necessary for transcription and translation in the given host. Vectors which are suitable for expressing foreign protein are called expression vectors and one such vector is the pUC 19.

Vectors based on bacteriophages

Bacteriophages are viruses that infect bacterial cells by injecting their DNA into them and consequently take over the machinery of the bacterial cells to multiply themselves. The injected DNA hence is selectively replicated and expressed in the host bacterial cell resulting in a number of phages which eventually extrude out of the cell ***(lytic pathway)*** and infect neighbouring cells. This ability to transfer DNA from the phage genome to specific bacterial hosts during the process of viral infection gave scientists the idea that specifically designed phage based vectors would be useful tools for gene cloning experiments. Two phages that have been extensively modified for the development of cloning vectors are lambda (λ) and M13 phages.

Bacteriophage lambda has a double stranded, linear DNA genome containing 48,514 bp, in which 12 bases on each end are unpaired but complementary. These ends therefore are sticky and are called cohesive or cos sites and are important for packaging DNA into phage heads. An important feature of the lambda genome is that a large fragment in the central region of its genome is not essential for lytic infection of *E. coli* cells. Therefore, vectors have been designed such that this region can be replaced by foreign insert DNA. These phage based vectors allow cloning of DNA fragments up to 23 kb in size.

M13 is a filamentous phage which infects *E. coli* having a pilus (protrusion) which is selectively present in cells containing a F plasmid (called F+ cells). The genome of the M13 phage is a single stranded, circular DNA of 6407 bp. Foreign DNA can be inserted into it without disrupting any of the essential genes. In the life cycle of the phage following infection of the host *E.coli* cell the single stranded DNA is converted to a double-stranded molecule which is referred to as the Replicative Form (RF). The RF replicates until there are about 100 copies in the cell. At this point DNA replication becomes asymmetric and single stranded copies of the genome are produced and extruded from the cell packaged with protein as M13 particles **(Fig. 5)**. The major advantages of developing vectors based on M13 are that its genome is less than 10 kb in size; the RF can be purified and manipulated exactly like a plasmid. In addition, genes cloned into M13 based vectors can be obtained in the form of single stranded DNA. Single stranded forms of cloned DNA are useful for use in various techniques including DNA sequencing and site-directed mutagenesis, techniques which will be discussed in a latter section.

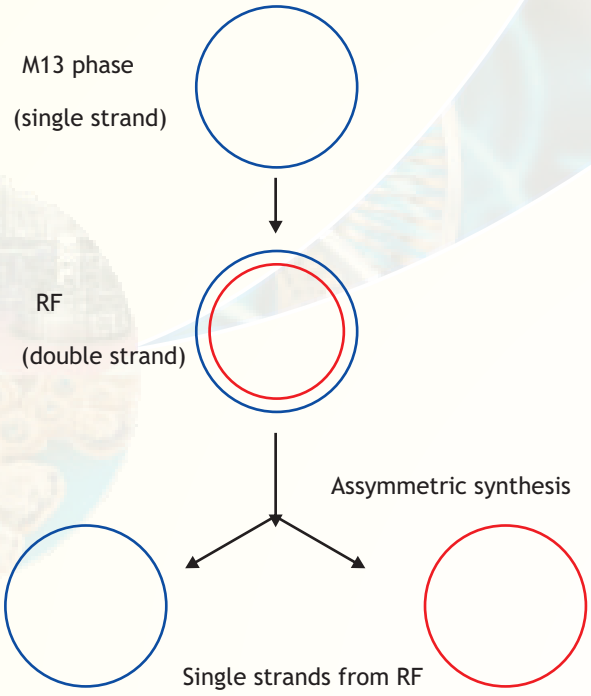


Fig. 5. Life cycle of M13 phage.

Cosmids

Cosmids have been constructed by combining certain features of plasmid and the 'cos' sites of phage lambda. The simplest cosmid vector contains a plasmid, origin of replication, a selectable marker, suitable restriction enzyme sites and the lambda cos site. Cosmids can be used to clone DNA fragments up to 45 kbp in size.

YAC vectors

YACs or Yeast Artificial Chromosomes are used as vectors to clone DNA fragments of more than 1 Mb in size. Therefore, they are useful in cloning larger DNA fragments as required in mapping genomes such as in the Human Genome Project. These vectors contain a *teleomeric* sequence, the centromere and an autonomously replicating sequence, features required to replicate linear chromosomes in yeast cells. These vectors also contain suitable restriction sites to clone foreign DNA as well as genes to be used as selectable markers.

BAC vectors

BACs or Bacterial Artificial Chromosomes are vectors based on the natural, extrachromosomal plasmid from *E. coli* - the fertility or F plasmid. A BAC vector contains genes for replication and maintenance of the F plasmid, a selectable marker and cloning sites. These vectors can accommodate inserts up to 500 kb and are used in genome sequencing projects. **Table 2** lists the common cloning vectors with the size of insert that can be cloned into them.

Table 2. Common cloning vectors.

Vector Type	Insert size (kb)
Plasmid	0.5-8
Bacteriophage lambda	9-23
Cosmid	30-40
BAC	50-500
YAC	250-1000

Animal and Plant viral vectors

You have already learnt how bacteriophages have been used to derive suitable vectors for gene cloning experiments in *E. coli*. Similarly, viruses that infect plant and animal cells have also been manipulated to introduce foreign genes into plant and animal cells. The natural ability of viruses

to adsorb to cells, introduce their DNA and replicate, have made them ideal vehicles to transfer foreign DNA into eukaryotic cells in culture. A vector based on Simian Virus 40 (SV40) was used in the first cloning experiment involving mammalian cells. A number of vectors based on other type of viruses like Adenoviruses and Papillomavirus have been used to clone genes in mammals. At present, retroviral vectors are popular for cloning genes in mammalian cells. In case of plants, viruses like Cauliflower Mosaic Virus, Tobacco Mosaic Virus and Gemini viruses have been used with limited success.

Host Cells

The tools described in the previous sections will result in the generation of recombinant DNA molecules in the laboratory. Eventually the propagation of these DNA molecules must occur inside a living system or host. Many types of host cells including *E. coli*, yeast, animal and plant cells are available for gene cloning and the type of host cell used depends on the aim of the cloning experiment. *E. coli* has become the most widely used organism in rDNA technology because its genetic make-up has been intensively studied, it is easy to handle and grow, can accept a range of vectors and has been extensively studied for safety. Another major advantage of using *E. coli* as host cells is that under optimal conditions the cells divide every 20 minutes making it possible to clone large amounts of foreign DNA and if the appropriate signals are incorporated into the vector large amounts of recombinant proteins are available for therapeutics and other uses.

For the expression of eukaryotic proteins, eukaryotic cells are often preferred because to be functionally active, proteins require proper folding and post translational modifications such as glycosylation which is not possible in prokaryotic *(E. coli)* cells. Even cloned eukaryotic genes containing introns cannot be processed in *E. coli* thereby necessitating the use of only eukaryotic host cells. Yeast cells have been used extensively for functional expression of eukaryotic genes because of several features. Yeasts are the simplest eukaryotic organisms (unicellular) and like *E. coli* have been extensively characterised genetically, easy to grow and manipulate and large amounts of cloned genes or recombinant proteins can be obtained from yeast cultures grown in fermentors (large culture vessels). Plant and animal cells may also be used as hosts in rDNA experiments and cells can be grown in tissue culture or can be induced and manipulated to form whole organisms (creation of transgenic animals and plants).

5.1.3. Making rDNA

The first step in the construction of a recombinant DNA molecule is to isolate the vector and the fragment containing the gene to be cloned. The vector and target DNA fragment are separately digested with the same restriction enzyme such as *Eco*RI which generates sticky ends. The vector is then treated with alkaline phosphatase enzyme so that later in the ligation step the vector does not self ligate. The cut vector and DNA fragment are mixed in a suitable ratio and then ligated with the enzyme DNA ligase to yield a recombinant vector containing insert. This procedure is

schematically explained in **Fig. 6**.

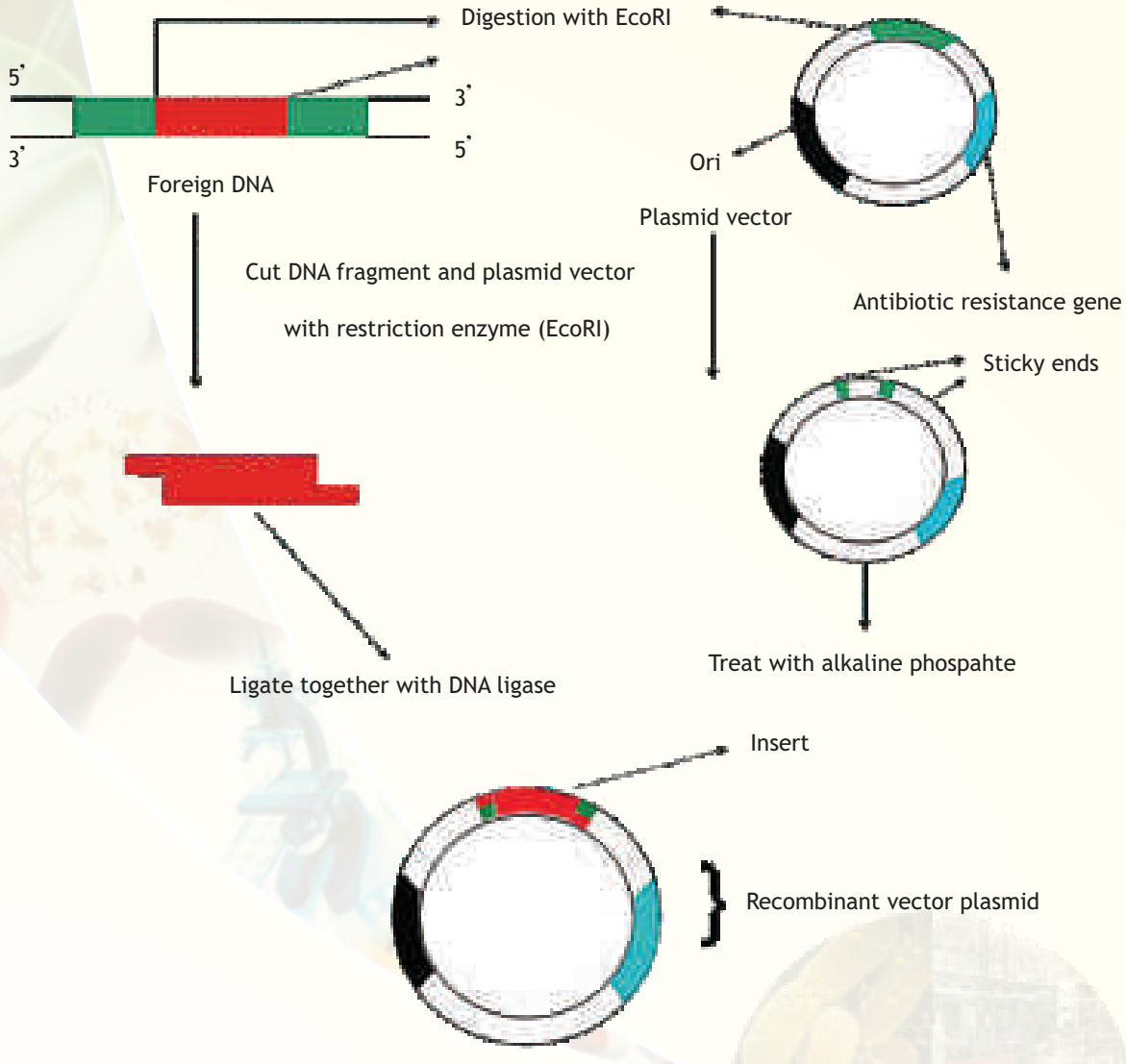


Fig. 6. Making recombinant plasmid

5.1.4. Introduction of rDNA into host cells

The next step after a recombinant molecule has been generated is to introduce it into a suitable host. There are many methods to introduce recombinant vectors and these are dependant on several factors such as the vector type and host cell. Some commonly used procedures are discussed below.

Transformation: In rDNA technology, the most common method to introduce rDNA into living cells is called transformation. In this procedure, bacterial cells take up DNA from the surrounding environment. Many host cell organisms such as *E. coli*, yeast and mammalian cells do not readily take up foreign DNA and have to be chemically treated to become competent to do so. In 1970, Mandel and Higa found that *E. coli* cells become markedly competent to take up external DNA when suspended briefly in cold calcium chloride solution.

Transfection: Another method to transfer rDNA into host cells involves mixing the foreign DNA with charged substances like calcium phosphate, cationic liposomes or DEAE dextran and overlaying on recipient host cells. Host cells take up the DNA in a process called transfection.

Electroporation: An electric current is used to create transient microscopic pores in the recipient host cell membrane allowing rDNA to enter.

Microinjection: Exogenous DNA can also be introduced directly into animal and plant cells without the use of eukaryotic vectors. In the procedure of microinjection, foreign DNA is directly injected into recipient cells using a fine microsyringe under a phase contrast microscope to aid vision.

Biolistics: A remarkable method that has been developed to introduce foreign DNA into mainly plant cells is by using a gene or particle gun. Microscopic particles of gold or tungsten are coated with the DNA of interest and bombarded onto cells with a device much like a particle gun. Hence the term biolistics is used.

Another method of introducing foreign genes is by the natural genetic engineer *Agrobacterium tumefaciens*. This method and its principle will be covered in the chapter of plant cell culture in Unit VI.

5.1.5. Identification of Recombinants

Once a recombinant DNA molecule has been introduced into appropriate host cells, it becomes imperative to select only those cells which have the rDNA from those of the original host cells which have not taken up the DNA. All procedures described in the previous section are only minimally efficient (about 1%) and hence after such an experiment majority of the cells do not have the foreign DNA. However the use of selectable marker genes which are an integral part of any cloning vector makes the selection of transformed cells quite easy. Generally, the selection methods are based on the expression or non-expression of certain traits such as antibiotic resistance, expression of an enzyme such as β-galactosidase or protein such as GFP (Green Fluorescent Protein) and dependence or independence of a nutritional requirement such as the amino acid leucine. For example if the host *E. coli* cells have taken up the plasmid pBR322 then these cells will grow in media containing the antibiotics ampicillin or tetracycline whereas normal *E. coli* cells will be killed by the antibiotics. Thus only transformed cells, however few,

will be selected for growth and division.

The simplest method for selecting the transformants relies on the presence of antibiotic resistance genes on the plasmid or phage based vectors as already discussed. It is possible however, that the transformants have the vector without the foreign DNA. This is because the procedure for making a recombinant vector is efficient but not 100% foolproof! If on the other hand a vector has two antibiotic resistant genes, e.g. pBR322, and the insert is contained in the tetracycline resistant gene, then the ampicillin resistant gene will be normally expressed allowing the transformed cells to grow on an ampicillin containing medium but due to a phenomenon called insertional inactivation (insert in tetracycline gene) the cells will be tetracycline sensitive. How does a scientist select for a sensitive or negative trait following a transformation experiment? A procedure called **replica plating** is used. As schematically explained in **Fig. 7** following an experiment of transforming *E. coli* cells with recombinant pBR322 plasmid, the host cells are first plated on solid media (agarose containing) with the antibiotic ampicillin (assume that the insert has been ligated within the tetracycline resistance gene). Colonies from every single cell plate having the plasmid will develop overnight. The role of the recombinant plasmid is to help the cell to multiply in the presence of antibiotic, which it would otherwise not be able to do. In order to select those colonies alone which are tetracycline sensitive and therefore, are relevant to the experiment as they have the insert, a procedure called replica plating is used. A petri plate containing solid media with antibiotic tetracycline is kept carefully under asceptic conditions(Laminar flow hood) to which a circular piece of velvet or velvet paper is aligned and pressed onto the colony containing ampicillin plate (master plate). With the same alignment it is pressed onto the tetracycline plate. Overnight only colonies not containing the insert will grow while due to insertional inactivation no colonies will grow which have the insert. The colonies which have the insert can easily be scored off by comparing the two plates.

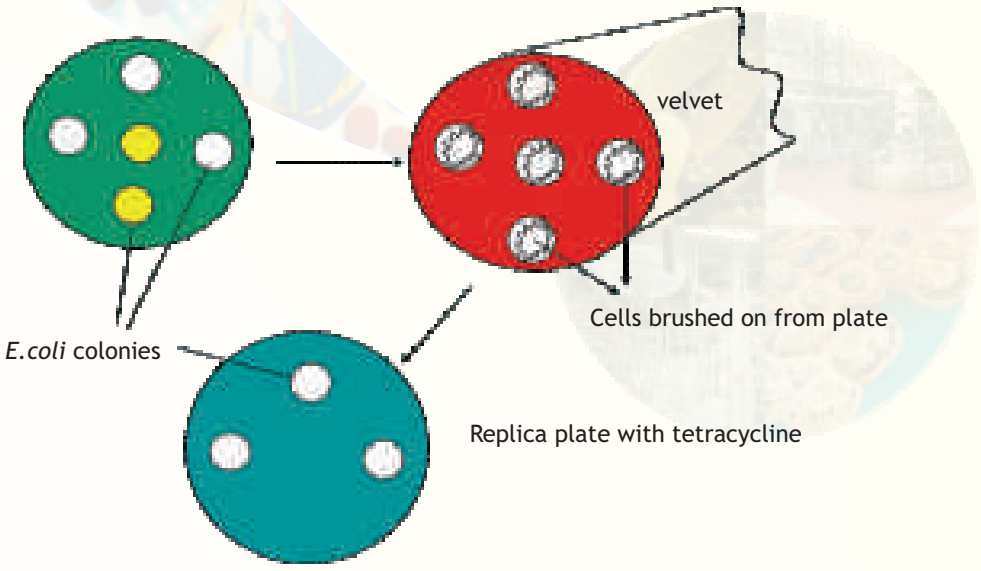


Fig. 7. Replica plating (note only colonies marked yellow have insert).

Another powerful method of screening for the presence of recombinant plasmids is referred to as blue- white selection. This method is based upon the insertional inactivation of the *lac Z* gene present on the vector (*e.g.* pUC 19). This gene expresses the enzyme β-galactosidase whose activity can cleave a colourless substrate called X-Gal into a blue coloured product. If the *lac Z* gene is inactivated due to the presence of the insert then the enzyme is not expressed. Hence if after a transformation experiment the *E. coli* host cells are plated on an ampicillin and X-Gal containing solid media plate then colonies which appear blue are those which have transformed cells (antibiotic resistant) but do not have the insert (express active enzyme). Colonies which appear white are both ampicillin resistant and have the insert recombinant DNA and thus are the cells to be used for future experiments.

The above described methods are used for selection of *E. coli* recombinants. There are several methods used for other host cell recombinants but the principles remain the same. Furthermore techniques for actually detecting recombinant proteins from colonies also have to be used where relevant. Where amplification of the insert DNA is the primary objective, plasmids are isolated from the host cells after growing the latter in large amounts and using the same restriction enzyme the insert is cut from the plasmid and recovered after electrophoresis.

5.1.6. Polymerase Chain Reaction (PCR)

The polymerase chain reaction or PCR as it is commonly known, was invented by Kerry Mullis in 1985. It results in the selective amplification of a specific region of a DNA molecule and so can also be used to generate a DNA fragment for cloning. The basic principle underlying this technique is that when a double-stranded DNA molecule is heated to a high temperature, the two DNA strands separate giving rise to single stranded molecules which can be made to hybridise with small oligonucleotide primers (single-stranded) by bringing down the temperature. If to this an enzyme called DNA polymerase and nucleotide triphosphates are added, much like what happens during replication, i.e primer extension occurs. This procedure is repeated several times, see **Fig. 8** which ultimately results in amplification of the DNA stretch between the two primers (one on each strand of the DNA). The basic requirements of a PCR reaction are:

1. DNA template to be amplified.
2. Primers which are oligonucleotides, usually 10-18 nucleotides long that hybridise to the target DNA region one to each strand of the DNA. Two primers of such a sequence are required so that they can hybridise as indicated in the **Fig. 8**.
3. DNA polymerase which is stable at temperatures above 80°C . Taq polymerase which has been isolated from a thermostable bacterial species is used.
4. Deoxynucleotide triphosphates and buffer.

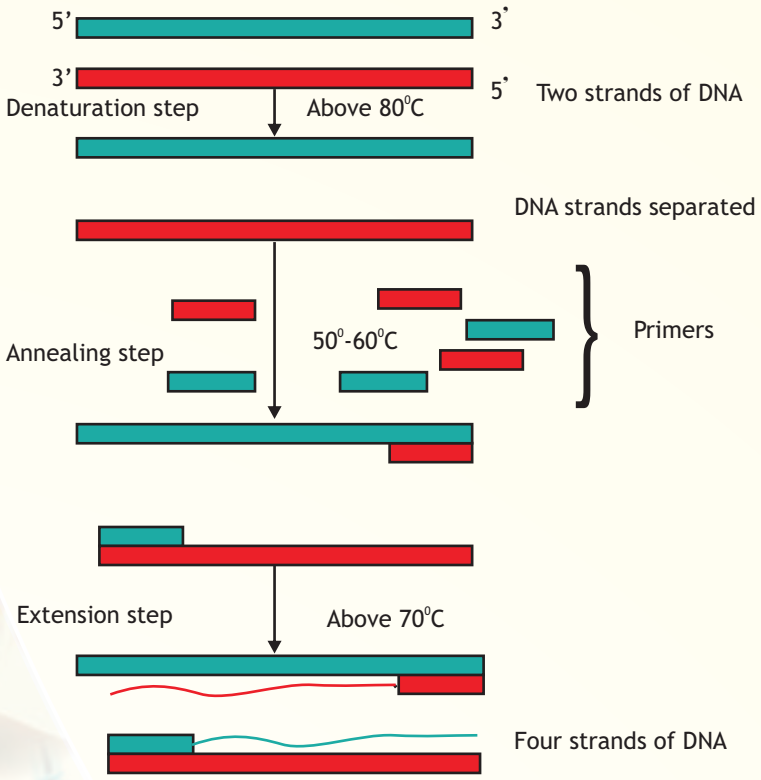


Fig. 8. PCR technique

A single PCR amplification cycle involves three basic steps of denaturation, annealing and extension **(Fig. 8)**. In the denaturing step, the target DNA is heated to a high temperature, above 80°C which results in DNA strand separation. Each single strand then anneals with a primer at a lower temperature between 50-60°C in such a way that extension can occur from it in a 5' - 3' direction (a requirement of the DNA polymerase). The final step is extension, wherein the enzyme Taq polymerase extends each primer using dNTPs and the DNA strand as template. The temperature for extension is around 70°C. The procedure is repeated and each set of steps is considered as one cycle (i.e denaturation, annealing and extension). At the end of one cycle two DNA molecules have become four and this geometric progression occurs with each cycle. Hence it can be computed that at the end of n cycles the number of DNA molecules is 2^n which is enormously amplified target DNA.

The invention of the PCR technique has revolutionised every aspect of modern biology. To detect pathogens, microbiologists in the past used techniques based on culturing and detecting antibodies against enzymes or proteins specific to the pathogen. Apart from taking time many of these procedures were not specific. PCR based diagnosis is faster, safer and more specific because it does not use live pathogens; instead DNA from the infected tissue is isolated and the PCR technique is carried out using primers having specific complementary sequences to the pathogen DNA. PCR is also a valuable tool in forensic science as large amounts of DNA can be

amplified from the small amounts present at the crime site, for DNA fingerprinting analysis. In recent years, PCR has also found use in detecting specific microorganisms from environmental samples of soil, sediments and water. It is interesting that archaeologists are using combinations of PCR and fingerprinting analysis to relate and establish ancient Egyptian dynasties from samples obtained from mummies.

5.1.7. Hybridisation Techniques

Once a specific DNA sequence has been isolated by cloning, it can be used as a probe to detect the presence and amount of complementary DNA sequences present in isolated DNA, for example from different species. A question can be asked such as, Is this particular gene or DNA fragment present in a mammalian cell also present in an insect cell?

DNA probes are relatively small single stranded sequences of DNA that recognise and bind to complementary sequences. Recall that two strands of DNA are held together by base complementarity *i.e.*, base A on one DNA strand hydrogen bonds with base T on the complementary strand and likewise base C hydrogen bonds with base G. A DNA probe which is single stranded will bind to a complementary sequence with the same base pairing rules (hybridisation) and if the probe can be tagged with a flourescent or radioactive label the complementary sequence can be located either in a cell nucleus or on a gel chromatogram. This principle is the basis of all hybridisation techniques, see **Fig. 9.**

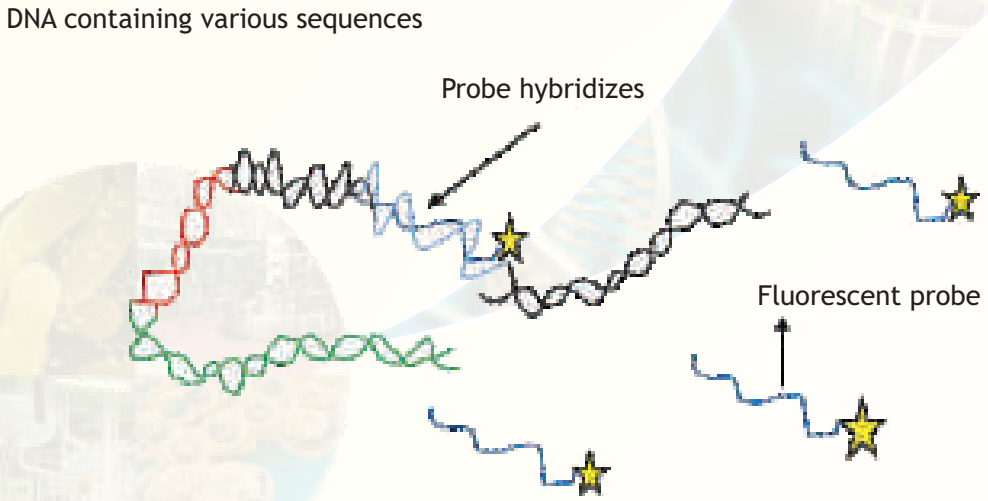


Fig. 9. Principle of hybridisation technique.

If the probe is tagged with a fluorescent label, under UV light its location can be easily seen as it will fluoresce. On the other hand if the probe is tagged with a radioactive label a technique called autoradiography is used wherein the gel is placed on a photographic film and the probe location is indicated by white spots on the developed film.

Southern Hybridisation Technique

This technique of identifying and locating specific sequences in DNA gels using probes was invented in 1975 by Edward Southern and is named Southern Hybridisation technique in his honor. This is an essential technique in all rDNA experiments and its aim is to identify a specific DNA sequence in a heterogenous population of DNA molecules. As discussed in the previous section the principle of the technique is based on the ability of a probe to seek out and bind to its complementary sequence.

The procedure involves isolation and digestion of total genomic DNA with one or more restriction enzymes. The DNA fragments thus generated are separated in agarose gels using the technique of electrophoresis, about which you have learnt from Class XI Biotechnology textbook. Following separation of the DNA fragments due to size differences, they are transferred from the gel to a nylon or nitrocellulose membrane in a technique called blotting, see **Fig. 10**. In the blotting procedure the DNA fragments are forced from the gel onto the membrane by capillary action. The membranes are baked briefly to fix the DNA fragments to the membrane so that they do not diffuse during the next step of hybridisation. Note that the membrane will have the same pattern of DNA bands as the original agarose gel. The membrane is then treated with the single stranded labelled probe for an appropriate period after which the membrane is washed and either photographed under UV light (if probe label is fluorescent) or overlaid with a photographic film (if probe is radioactive). The location of the probe is determined leading to the identification of a gene or specific DNA fragment obtained from that given genomic DNA. A similar principle is used for determining RNA locations and this technique is called Northern hybridisation.

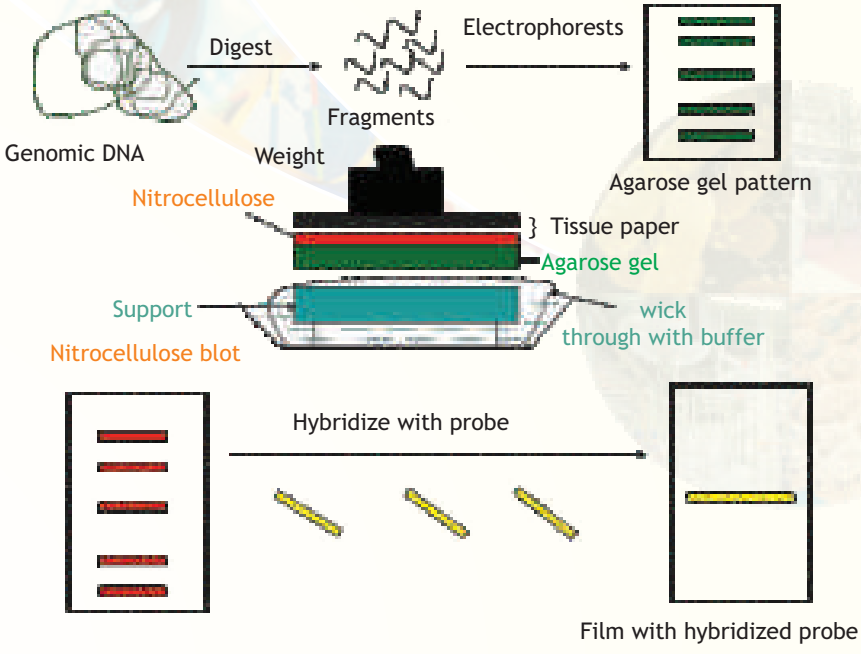


Fig. 10. Southern hybridisation technique.

5.1.9. DNA Library

In order to obtain a DNA fragment carrying a useful target gene it is essential to make a DNA library containing all possible DNA fragments from the genome of a given species or cell type. The target gene can then be easily identified from the library and then used for cloning. Remember per cell only two copies (diploid) of a target gene are present and therefore too little is there for detection and use. A DNA library not only contains all possible fragments of DNA from a given cell or organism but also large amounts of the same as a resource. Two types of DNA libraries can be constructed - a genomic DNA library and a cDNA library. A genomic DNA library has all possible DNA sequences in large amounts from the given cell type. A cDNA library on the other hand has only expressed gene sequences such as protein encoding genes. It is obvious that a genomic DNA library is larger than a cDNA library. Whether the fragments are genomic or cDNA (protein encoding) these are inserted into vectors such as plasmids and then introduced into *E. coli* hosts. The cells are diluted and plated on agarose plates so that distinct colonies are formed starting from single cells. Each colony would therefore have an amplified amount of a given fragment. However unlike an *E.coli* regular library which has a classification system, a DNA library has to be screened for finding a fragment or gene of interest which is a tedious and time taking process. A schematic diagram of the steps in constructing a DNA library is indicated in **Fig. 11**.

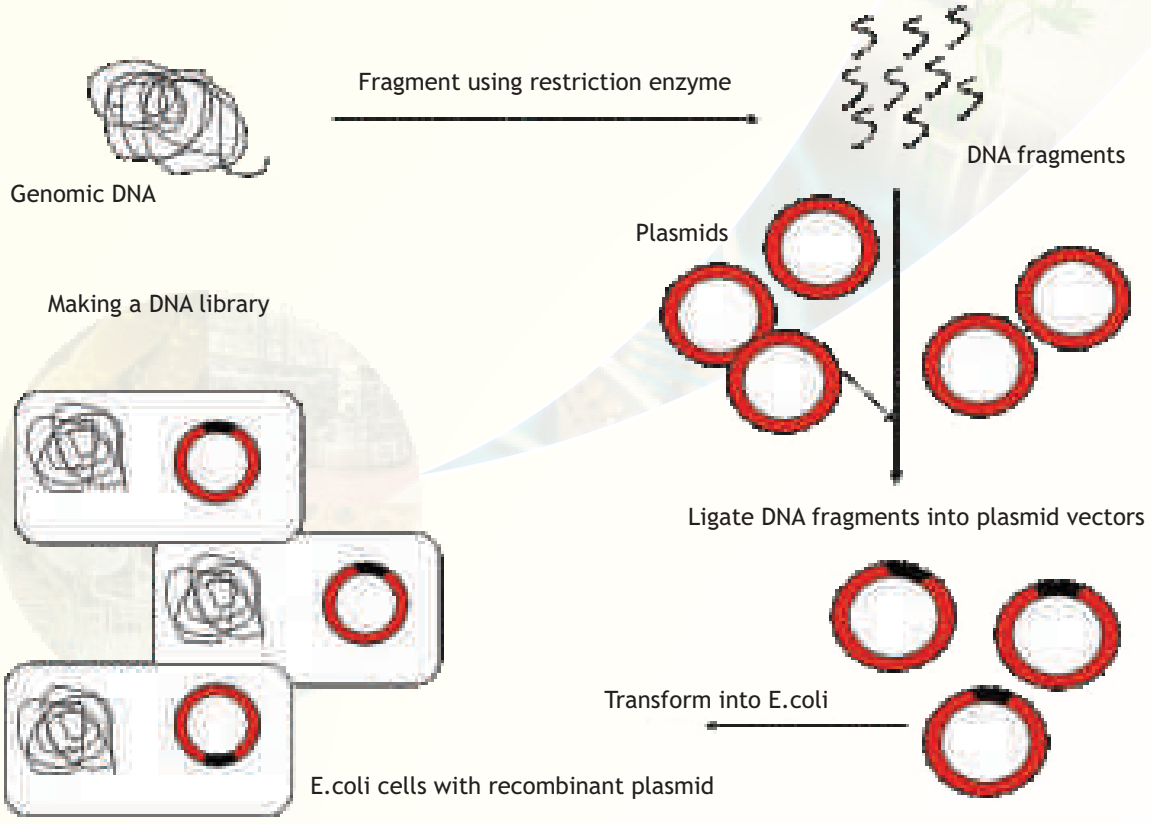


Fig. 11. Making a DNA library.

As shown in **Fig. 11**, to prepare a genomic library, the total genomic DNA is isolated from a tissue or organism and then fragmented using a restriction enzyme such as *Eco*RI. An appropriate vector such as pBR322 (plasmid based) is also digested with the same enzyme and is then treated with the enzyme alkaline phosphatase which as indicated elsewhere removes the 5'phosphate group to prevent the plasmid from self ligation. The DNA fragments and cut vector are mixed and then treated with the enzyme DNA ligase. Each vector molecule will contain a different fragment of DNA and these are introduced into *E. coli* host cells by a technique called transformation. More details about this technique will be discussed in the following section.

Although a genomic library represents the entire genome of an organism, it may not be useful in case of eukaryotic organisms. Genomic DNA of eucaryotes contains non-coding DNA like introns, control regions and repetitive sequences unrecognised by prokaryotic systems such as in *E. coli*. Therefore another library called cDNA library is preferred especially if the aim of the cloning experiment is to express eukaryotic proteins in an *E. coli* host. This library has two major advantages over a genomic library. Firstly, it represents only those genes that are being expressed by a particular cell or under specific conditions. Eucaryotes, which include multicellular organisms such as animals and plants have differentiated cells which means that different cells e.g. liver and brain cells express some common proteins and also other proteins which are not common. Secondly, since the source material in constructing such libraries is mRNA, these molecules lack introns and hence would represent only the coding sequences of the genome. However mRNA molecules are highly unstable as they are easily degraded by RNAses and hence these molecules are faithfully copied into the more stable DNA (now called cDNA) before cloning. The construction of a cDNA library begins with the isolation of mRNA from a given cell type or tissue which are copied into cDNA using a special enzyme called **reverse transcriptase.** The procedure results in double-stranded cDNA which can be incorporated into vectors such as pBR322.

5.1.9. DNA Sequencing

In this era of genomics wherein whole genomes of species are being sequenced and compared to get a vision into the fundamental nature of DNA, the blueprint of life, the ease with which DNA is sequenced has played a major role. In the seventies two major methods were developed to sequence DNA:

1. The dideoxynucleotide chain termination method invented by Fred Sanger (the same scientist who invented a protein sequencing) and Andrew Coulson
2. Chemical degradation method invented by Walter Gilbert and Allan Maxam (also called the Maxam and Gilbert method).

   Of the two methods the first method is more popularily used and hence this technique will be discussed.

Dideoxynucleotide Chain Termination method

In your class XI textbook you would have read about replication of DNA and also about the various enzymes and substrates required. DNA polymerases, the major enzymes required in replication, have certain properties:

- A single stranded DNA template is required for them to act upon.
- A new strand cannot be initiated; only primers can be extended using the single strand DNA template as a guide.
- Extension or DNA synthesis occurs in a 5'→ 3' direction which requires that a new nucleotide is added to the 3' hydroxyl group of the chain.
- Deoxynucleotide 5' triphosphates are the normal substrates.
- However if the 3' hydroxyl group of a deoxynucleotide triphosphate is absent as in the 2',3' dideoxynucleotide triphosphate **(Fig. 12)** it would result in, that nucleotide being incorporated into the growing chain but subsequently the chain cannot further be extended (no 3' hydroxyl) and this is the main principle of the Sanger Dideoxynucleotide Chain Termination method.

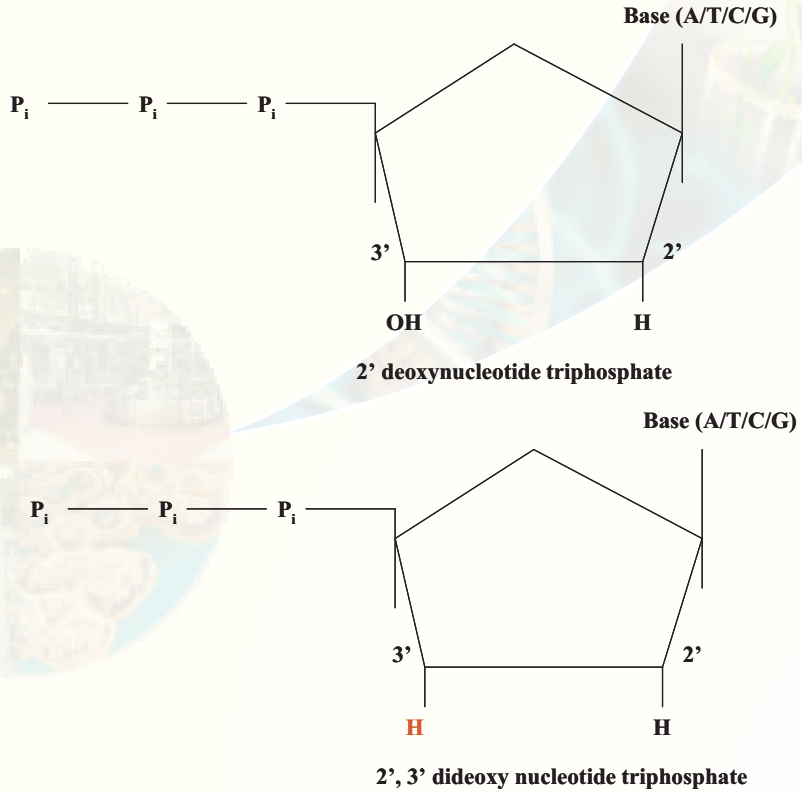


Fig. 12. Deoxy and dideoxy nucleotides.

During a primer extension process, whenever a dideoxynucleotide is incorporated, the DNA polymerase enzyme cannot further carry out its reaction and the new DNA strand is thereby terminated **(Fig. 13)**

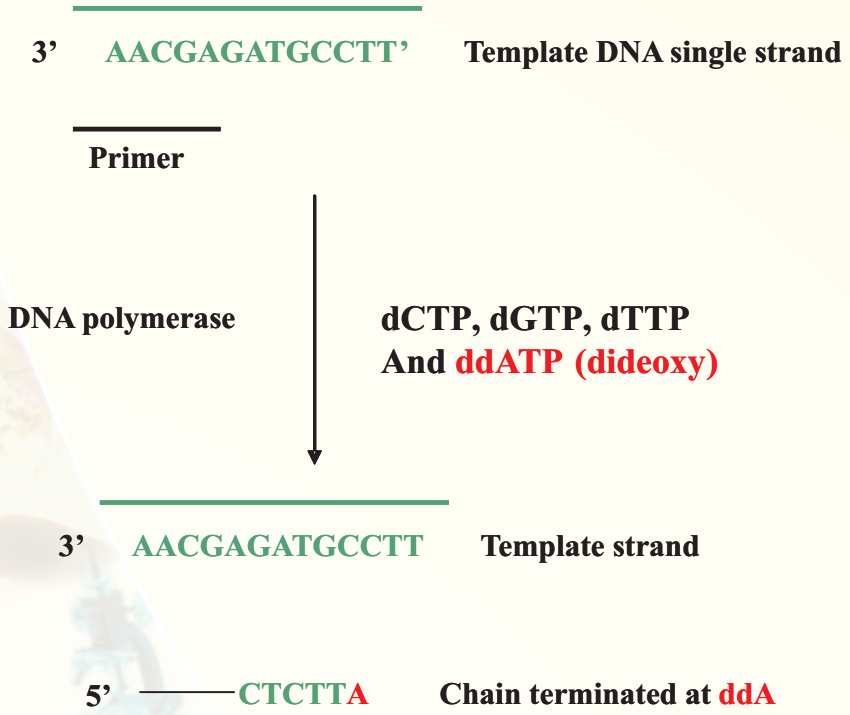


Fig. 13. Principle of dideoxynucleotide chain termination method.

Typically the sequencing technique is carried out in four test tubes. Each test tube carries the following common reagents, single stranded DNA template, deoxynucleotide triphosphates, primer and DNA polymerase. In addition to it a small amount of the four dideoxynucleotide triphosphates i.e ddATP, ddCTP, ddGTP and ddTTP is added separately into the four test tubes. The purpose of adding only a small amount of the dideoxy derivative is that only one dideoxy derivative is incorporated into each extending chain causing termination. Hence in a test tube containing ddATP all chains will end at ddA but at different positions of T present in the template **(Fig. 14)**. The same is applicable to the other test tubes containing ddCTP, ddGTP and ddTTP. It is to be noted that the dideoxynucleotide is added wherever the complementary base is present on the template.

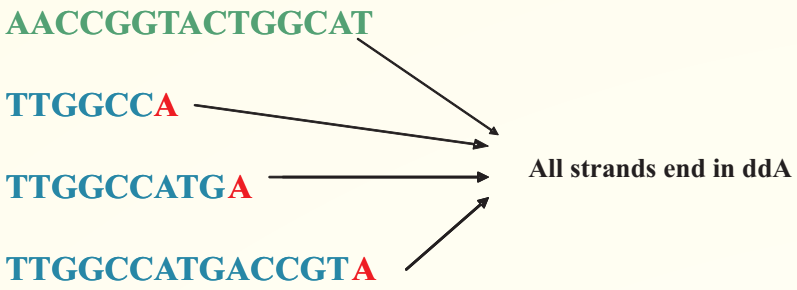


Fig. 14. Principle of chain termination at a particular dideoxynucleotide.

The various strands prematurely terminated at the particular ddNTP in a given tube are subjected to electrophoresis in special gels wherein bands can be resolved (separated) even if they differ by one nucleotide. The strands migrate in the gel with the shorter fragments moving faster towards the anode. The primers used in each tube can be made radioactive and hence the position of the separated strands in the gel chromatogram can be easily visualised using autoradiography. **Fig. 15** is a schematic representation of a typical gel and its read sequence. Colours for different dideoxy nucleotides are for clarity.

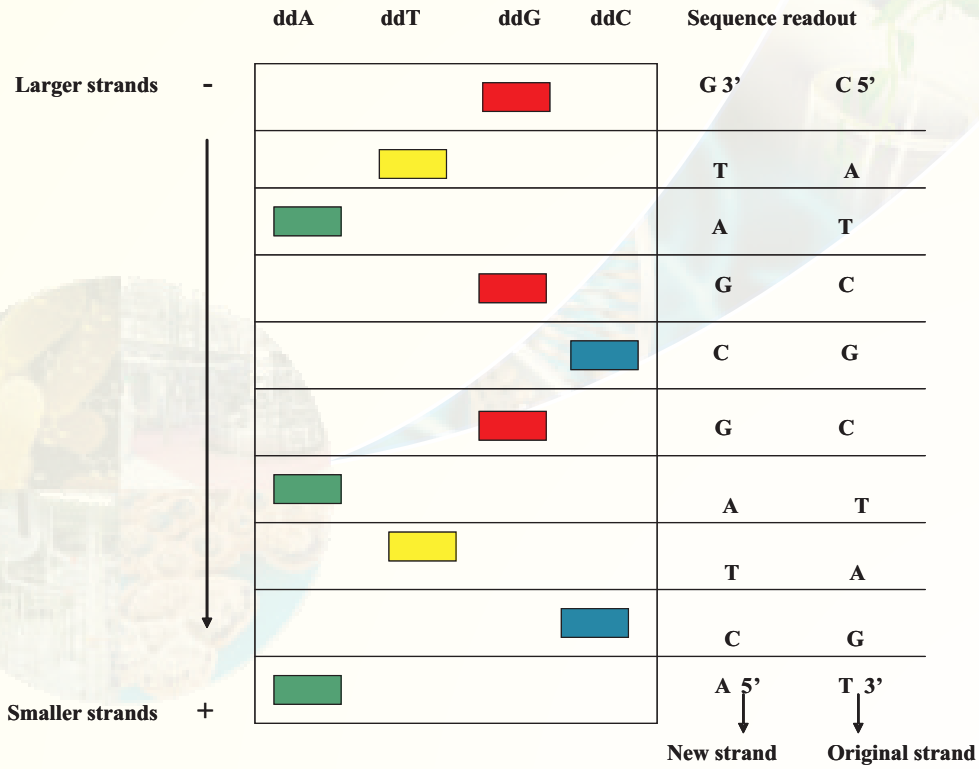


Fig. 15. Reading a sequencing gel pattern (lines drawn for clarity in reading).

Nowadays DNA sequencing technologies have become automated. To avoid using radioisotopes and their consequent danger, dideoxynucleotides are conjugated with florescent molecules which on excitation give a different colour each. Hence each band on the gel (read from anode to cathode) indicates the particular base as its terminal dideoxy nucleotide fluoresces with a given colour. This avoids the use of a four lane gel, a single lane gel electrophoresis is instead conducted and the gels are then laser scanned and the data fed into a computer. The computer is programmed to display the gel scan and the base readout as shown in **Fig. 16.**

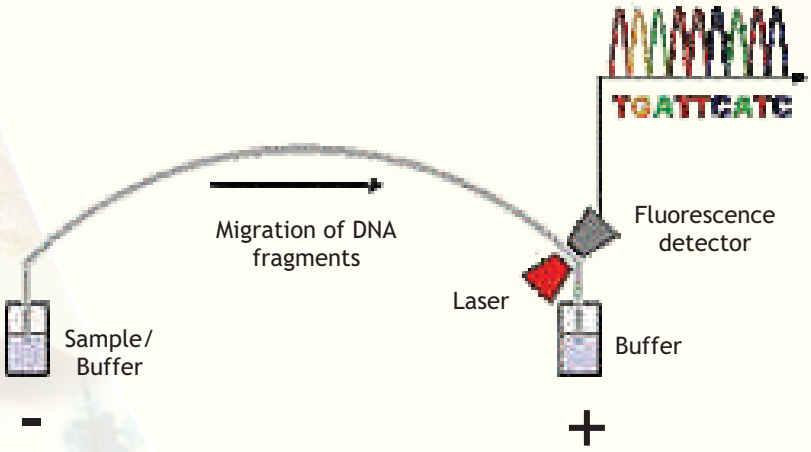


Fig. 16. Single Tube Sequencing experiment.

5.1.10. Site-directed Mutagenesis

Mutation is an alteration in any of the base of a DNA sequence sometime's leading to a defective protein or prematurely terminated non-functional protein. Mutations are spontaneous in nature although rare (*e.g* sickle cell haemoglobin). In the technique of site-directed mutagenesis a Biotechnologist is able to create mutation selectively, rather than that which occurs randomly in nature. Using this technique amino acids can be substituted in the expressed proteins making them more stable or functionally better. Furthermore the role of specific amino acids in proteins has led to a better understanding of protein structure and function.

The principle of site-directed mutagenesis as schematically explained in **Fig. 17** involves cloning the target gene into an M13 vector wherein it is presented as a single stranded part of the phage genome. A small oligonucleotide is added containing a complementary sequence to the gene but with one or more altered nucleotides. This allows the oligonucleotide to bind to a complementary portion in the target gene. This then acts like a primer *in vitro* to synthesise a double stranded replicative form. Note that the duplex RF form has one strand with the original target gene sequence, wild type and the other strand with the altered nucleotide(s). The duplex DNA

molecule is then introduced into bacterial cells by transformation. Subsequent replication inside bacterial cells will produce either wild type or mutant gene containing plasmids. If appropriate expression signals are present altered protein can be expressed and studied. In the unit on Protein engineering a site-directed mutagenesis experiment has been described wherein a stable form of the proteolytic enzyme subtilisin has been generated which is used in laundry detergent.

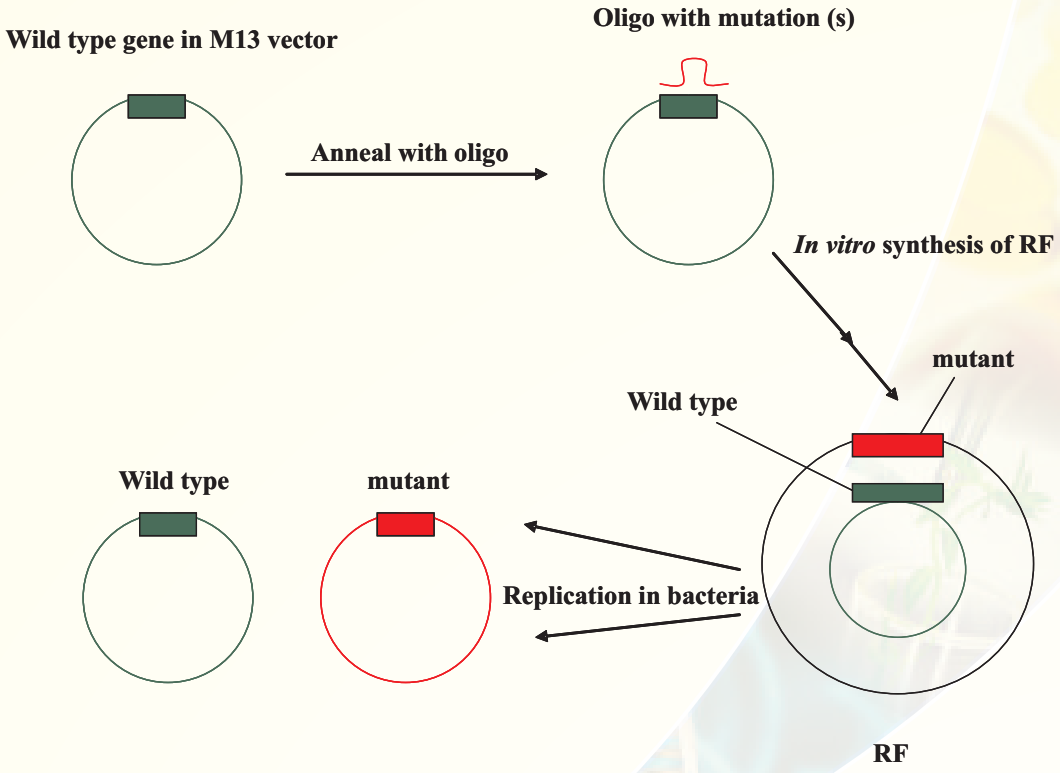


Fig. 17. Site-directed mutagenesis

Review Questions

1. Define the following:

   Plasmid — Restriction site
   Transformation — Mutation
   Transfection

2. What are restriction enzymes and why are they so important in rDNA technology?
3. Enlist the various steps involved in a rDNA experiment.
4. What are the essential features of a vector?
5. What does PCR stand for? Name the different steps in a PCR reaction.
6. Explain 'Insertional Inactivation'.
7. What are the disadvantages of using E. coli for production of eukaryotic proteins?
8. Distinguish between:
   i) Blunt ends vs sticky ends
   ii) YAC vs BAC
   iii) Genomic library vs cDNA library
   iv) Microinjection vs Electroporation
9. Write a short note on RFLP and indicate one of its important applications.
10. Why are ddNTPs used in sequencing? Briefly indicate the principle of DNA sequencing using these.
11. Indicate one application of site-directed mutagenesis.

References

1. Molecular Biotechnology by S.B. Primrose, Panima Publishing Corporation, New Delhi (1999), 2nd Edition.
2. Molecular Biotechnology by Glick and Pasterneck, Panima Publishing Corporation, New Delhi (1999), 2nd Edition.
3. An Introduction to Molecular Biotechnology Edited by Michael Wink, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany (2006).
4. Principles of Gene Manipulation and Genomics by S. B. Primrose and R. M. Twyman, Blackwell Publishing, Carlton, Victoria, Australia (2006), 7th Edition.
5. From Genes to Clones by Ernst-L. Winnacker, Panima Publishing Corporation, New Delhi (2003).
6. Gene Cloning and DNA Analysis: An Introduction by T. A. Brown, Blackwell Sciences Ltd., Oxford (2001), 4th Edition.

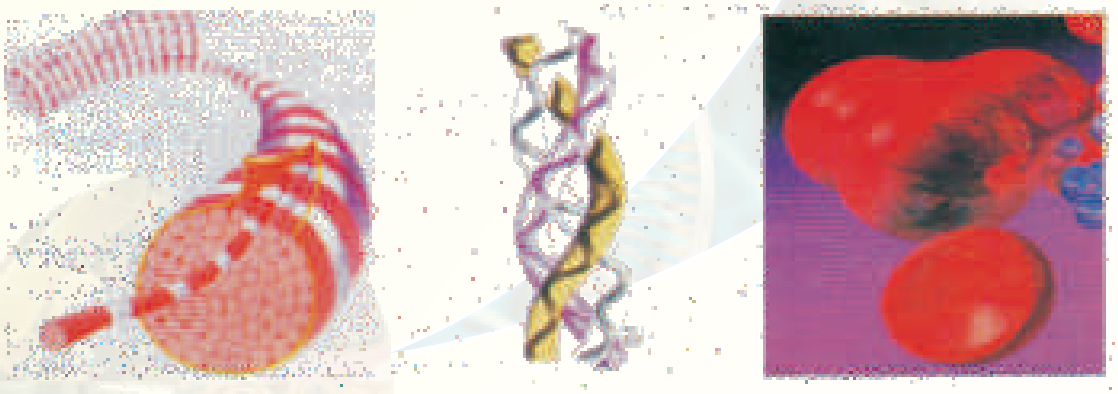
CHAPTER

2

PROTEIN STRUCTURE AND ENGINEERING

5.2.1. Introduction to the World of Proteins

The shape, structure and function of the human body is one of the Nature's marvels. The fertilisation of an egg by a sperm to the growth of a whole human body involves numerous steps of growth and differentiation. When we breathe, we feel a sense of oxygen flowing through our lungs and racing in our blood vessels, to be delivered to all our tissues. While we flex our muscles, we can feel them first tightening and then relaxing. The molecules involved are proteins, haemoglobin which transports oxygen, collagen which provides the strength to our bones and extracellular tissue and actin, myosin and several others which help in muscle contraction **(Fig. 1)**. It is noteworthy that among the biomolecules you have studied, proteins have the maximum diversity in function. The key to this enormous diversity is the unique structure of proteins. Although all proteins are made up of 20 different amino acids the sizes and sequence combinations and variations of each protein leads to millions of unique 3-D structures and thereby functions. Scientists have been striving to relate protein structure with function and hence the first step would be to determine 3-D structure of a protein.



Myosin and actin | Collagen | RBC with Haemoglobin

Fig. 1. Proteins having multiple roles

Even more amazing is the structure and data processing abilities of the human brain. We tend to marvel at the incredible speed and processing functions of the super computers little realising the creativity of the human being who invented them. The speed and correlation of sensory stimulation is unique to the human brain which can grasp the diversity of sensory inputs and convert them to learning and memory for later application. What are these proteins which enable

these functions and why are some brain related diseases like Alzeimers occurring, in which certain proteins show abnormal structure and behaviour?

A number of human diseases are due to the deficiency or abnormal structure of proteins. The lack of a particular subunit, alpha or beta of the oxygen carrying protein haemoglobin results in Thalassaemia, a devastating disease in which an infant cannot grow without repeated transfusions. If the beta chain is present but with a substituent in one of the amino acid residues another debilitating condition called Sickle Cell Anaemia results which is endemic to certain parts of Africa. The absence of an enzyme- Adenosine deaminase results in the birth of a severely immunocompromised baby who cannot last infancy (SCID). More recently it has been discovered that certain "rogue proteins" whose structure has been altered can result in diseases such as the Mad cow disease wherein the disease itself appears to be propagated by infectious proteins called **"prions"**. Clearly proteins need to be understood in detailed terms.

The completion of the Human Genome Sequence has revealed about 35,000 genes. However the actual number of proteins encoded by these genes may be many more due to posttranscriptional modifications. Different cells have specialised proteins for their unique functions in addition to the housekeeping proteins required for metabolism and generation of ATP. Sometimes these proteins are secreted to the outside like the proteolytic enzymes from the pancreas or hormones from ductless glands like the pituitary. We are yet to identify all the proteins required for a body to function and this presents a challenge to the future biotechnologists. One of the outcomes is the merging field of protein structure and function- proteomics. This chapter will enable you to understand various features of the area of proteomics- 3-D structure, functions and applications of protein products, some generated by biotechnological processes.

5.2.2. 3-D Shape of Proteins

The morphology, function and activity of a cell are all dependant on the proteins expressed. Proteins perform a variety of roles. The three dimensional properties of proteins have an important bearing on their function. The first step in determining the structure of a protein is to isolate it in a pure form from its cellular location (note bacterial, plant or animal cell). The purified protein is then crystallised so that using a technique called X-ray crystallography its three dimensional structure can be deduced. Nowadays another powerful technique called Nuclear Magnetic Resonance (NMR) has been developed which can deduce protein structures in solution and hence crystallisation is not required. However the protein in either technique has to be purified and some general procedures used to purify proteins will be discussed in subsequent sections of this chapter. In general when we refer to the structure of a protein this involves two aspects- the chemical structure which is the amino acid sequence of the polypeptide and its folding in space which is referred to its 3-D structure.

One of the major breakthroughs in protein sequence determination was achieved in the middle of the last century by Dr. Frederick Sanger who developed the first sequencing reagent FDNB (fluoro dinitro benzene) and a general strategy for sequencing. By using these methods he was able to sequence the important hormone insulin which is required by diabetics and more importantly he demonstrated for the first time that proteins were linear polymers of amino acids. For this work he was awarded the Nobel Prize and it will be interesting for you to know that several years later he was awarded a second Nobel Prize for developing a sequencing technique for DNA which has been described in the Recombinant DNA technology chapter previously. Another protein chemist, Pehr Edman in 1950 developed another sequencing reagent and procedure which is used in modern day sequenators as the procedure has been automated. These methods have been discussed in the XIth class' textbook of Biotechnology in Unit II, Chapter 2. Notably using the sequence of insulin established by Sanger a biotechnology company called Eli Lilli was able to develop recombinant human insulin which is the major source for insulin administration to diabetics worldwide.

With the availability of pure proteins, scientists like Linus Pauling, G.N.Ramachandran, Max Perutz and John Kendrew to name a few started developing techniques to study the 3-D shapes of proteins using high resolution X-rays. They laid the foundation for deducing protein structure by enunciating the basic rules which govern protein folding and the forces which cause the folding and stabilise them. Hence from these studies the concepts of planarity of the peptide bond, secondary structures such as alpha helix and beta pleats were developed. These concepts were introduced in the Class XI Biotechnology textbook.

Let us reiterate some important points regarding protein structure from the Class XI Biotechnology textbook. Protein structure has been divided into four hierarchial levels to understand their organisation:

The linear order or sequence of covalently linked amino acid sequence is defined as **primary structure**. Depending on the nature and arrangement of the amino acids present different parts of the polypeptide chain form **secondary structures** like alpha helices and beta pleats. The **tertiary structure** organisation of these secondary structural elements occurs when these get compacted with each other to form compact spherical or globular units which are also thermodynamically stable conformations of these molecules in aqueous solutions (note cytoplasm is mainly water). In compaction several non-covalent interactions occur between the amino acid side chains. The **quarternary structure** is the association of two or more independent proteins/polypeptides via non-covalent forces to give a multimeric protein **(Fig. 2)**. The individual peptide units of this protein are referred to as subunits and they may be identical or different from one another.

The dominant forces which cause linear protein chains to undergo folding in space lies to a large extent in the chemistry of the amino acid residues they contain. Amino acids are broadly divided

into three main groups- **polar or hydrophilic** (eg. serine, glutamine), **charged** (eg. aspartate, arginine) and **hydrophobic** (eg. tryptophan, valine). Hence based on these features amino acid side chains can interact in space by a variety of non-covalent forces which is the basis of forming and stabilising protein structures in space. Let us examine some of the major non-covalent forces found in proteins.

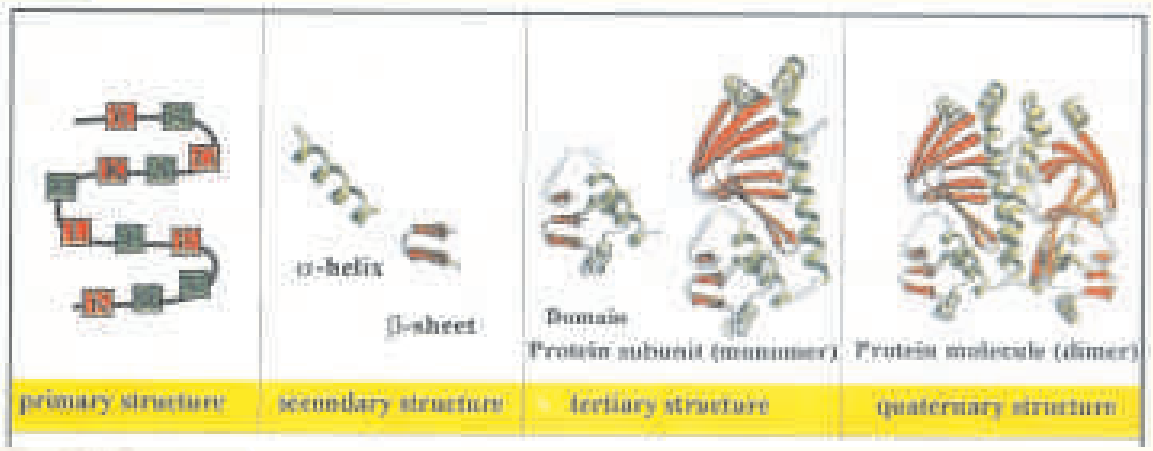


Fig. 2. Hierarchical organization in protein structure

Non-covalent bonds

The non-covalent interactions involved in organising the structure of protein molecules can be broadly divided into four categories:

- Ionic bonds
- Hydrogen bonds
- Van der Waals forces
- Hydrophobic interactions

Ionic bonds

These involve interactions between the oppositely charged groups of a molecule. For example the positively charged amino acid side chains of lysine and arginine can form salt bridges with the negatively charged side chains of aspartate and glutamate. These ionic interactions are also known as **salt bridges** because these are dominant bonds found in salts like sodium chloride wherein the positively charged sodium ion interacts with the negatively charged chloride ion. However, although ionic bonds have similar strengths to covalent bonds in vacuo, the bond strength of ionic bonds is vastly reduced in water due to the insulating qualities (dielectric strength) of water. Ionic bonds are highly sensitive to pH and salt concentration.

Hydrogen bonds

Hydrogen bonds are formed by "sharing" of a hydrogen atom between two electronegative atoms such as Nitrogen and Oxygen. In this case strongly polarised bonds between hydrogen and a small, very electronegative atom (N,O or F) allow a strong dipole-dipole bond to be formed with another small very electronegative element (N, O or F, **Fig. 3)**. Importantly, the very small sizes of these elements also allow them to approach each other so closely that a partial covalent bond is also formed (*e.g.*O-H---N). It is to be noted that the partial covalent character means that these bonds (H-bonds) are directional and strongest when the nuclei of all three involved atoms are in a linear arrangement.

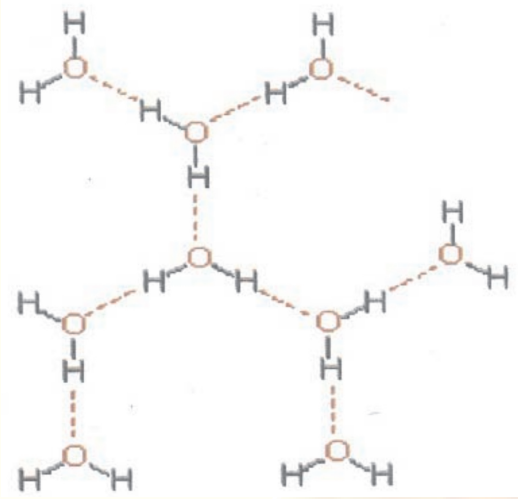


Fig. 3. Hydrogen bonding network in water

Van der Waals forces

These forces are weak attractions (or repulsions) which occur between atoms at close range. The Van der Waals types of forces are essentially contact forces, proportional to the surface areas in contact. These forces are of little significance at a distance due to the rapid 1/r6 (r is the inter-atomic distance) fall off. Even though weak, these bonds can be important in macromolecules because the large surface areas involved can result in reasonably large total forces.

Hydrophobic interactions

Hydrophobic interactions can be best explained by taking an example of oil in water. The oil tends to separate out fairly quickly, not because the oil molecules "want to get together", but because the water forces them out. The hydrophobic interaction is thus a manifestation of hydrogen bonding network in water. In water, each molecule is potentially bonded to four other molecules through H-bonds **(Fig. 3)**.

If a non-polar molecule, which cannot participate in hydrogen bonding, or in electrostatic interactions with water molecules, is added into water , a number of hydrogen bonds will be broken and not replaced. Since hydrogen bonds are favourable interactions, there will be an energy cost to putting non-polar molecules into water. Water therefore forces these molecules out of solution to minimise the surface of contact and thus the number of hydrogen bonds which are broken. Such forces known as hydrophobic forces are among the most important in driving proteins to fold into compact structures (globular) in water. Also, these forces are responsible to make different proteins assemble together to form structures found in muscles, membranes and

other organs. In proteins therefore, hydrophobic regions are preferentially located away from the surface of the molecule and form the interior core of the protein.

5.2.3. Structure-Function relationship in Proteins

As you have now learnt about how various forces drive proteins to assume characteristic shapes, it is worthwhile to consider why shape is paramount to the function of a protein. We will look at two proteins- an enzyme, chymotrypsin and the oxygen carrying protein, haemoglobin, to emphasise the importance of protein structure in its function.

Chymotrypsin, a proteolytic enzyme

As injested food makes way into the duodenum from the stomach, the proteins encounter a fierce proteolytic duo- trypsin and chymotrypsin which precisely cut the linear chains into short peptides which later on are acted upon by peptidases to release amino acids. Chymotrypsin, which hydrolyses peptide bonds following bulky aromatic amino acid residues in polypeptides is actually synthesised in the pancreas and through the pancreatic duct released into the duodenum. Have you wondered why this enzyme being a powerful proteolytic enzyme does not end up cutting cellular proteins within the pancreas itself? Nature has ensured that chymotrypsin and other proteolytic enzymes are synthesised as inactive harmless precursors known as zymogens which are then activated when required only in the duodenum, their site of activity, a process called *in-situ* activation. This activation in molecular terms results in an alteration in its shape so that it may now be able to interact with its substrate. The inactive precursor enzyme is termed chymotrypsinogen and the fully active enzyme is called chymotrypsin. The enzyme chymotrypsin is made up of a linear chain of 245 amino acids interrupted into three peptides- A,B,C. The protein folds into a globular structure. In the 3-D structure of the enzyme three important amino acid residues, his57, asp102 and ser195 come close together in space **(Fig. 4)** which allows a "charge relay system" to operate as indicated in **Fig 5**. The negatively charged asp102 is able to hydrogen bond with the adjacent his57 partially borrowing the hydrogen ion from the latter. The his57 makes good its partial hydrogen ion loss to aspartate by attracting a hydrogen ion from the adjacent ser195 through the his57 residue much like a relay race where the baton is passed from one member to another, the difference here being that the baton is a charge.

Normally the hydroxyl group of a serine residue is not acidic (pKa 12) and this is true for all other serine residues of chymotrypsin; only ser195 becomes acidic due to the unique constellation of the three amino acid residues because the protein has folded uniquely in space. You may be curious about the importance about an acidic serine residue. The negatively charged oxygen anion is able to make a nucleophilic attack on the carbonyl carbon of the peptide bond of its substrate, loosening it so that a water molecule can hydrolyse the bond **(Fig. 5)**. The specific site of chymotrypsin (recall that the enzyme is specific to aromatic residues) is a large space created

within the enzyme active site and lined by hydrophobic residues which therefore only allow bulky aromatic, hydrophobic amino acids to bind. This binding brings the susceptible peptide bond close to the attacking ser195 residue. In chymotrypsinogen, the substrate binding site is blocked and hence the enzyme is inactive. *In-situ* activation of trypsin involves a proteolytic cut in chymotrypsinogen which results in a conformational change, exposing the substrate binding pocket.

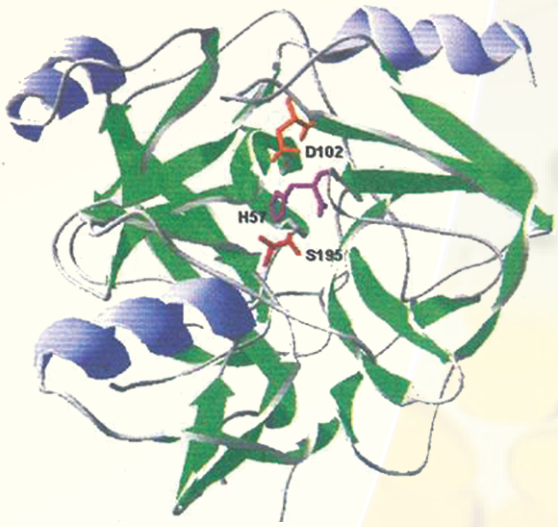


Fig. 4. Three dimensional structure of chymotrypsin

The interesting thing is that when nature has found a useful folding pattern which can cause hydrolysis of protein substrates, it repeats this in a variety of other enzymes. Trypsin, subtilisin (a proteolytic enzyme found in *B. subtilis*, a bacterium), thrombin (a proteolytic blood clotting factor) and the brain enzyme, acetyl choline esterase all have a reactive serine residue which is central to the catalytic mechanism.

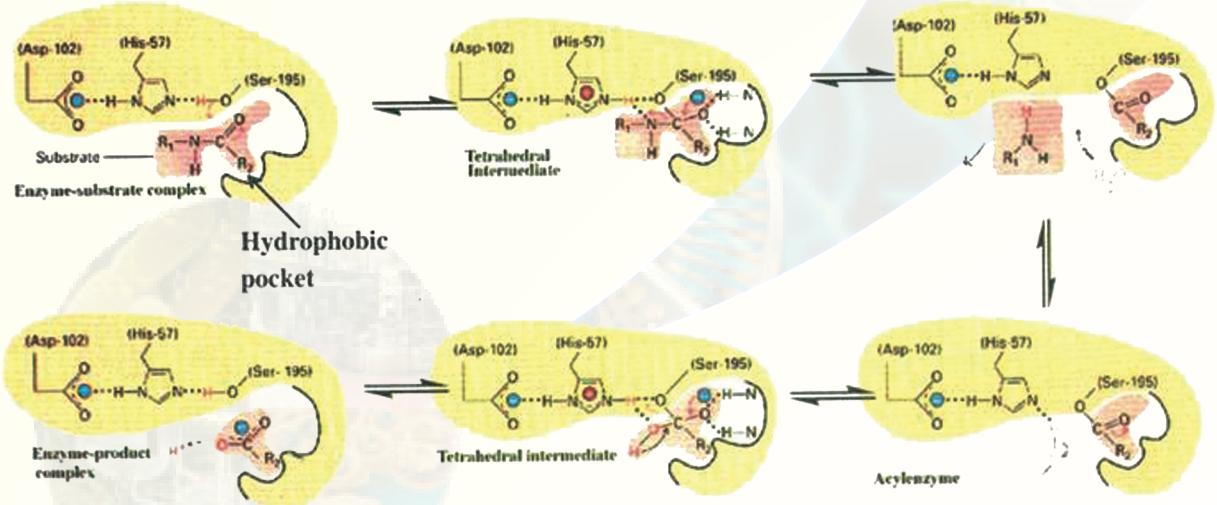


Fig. 5. Charge relay transfer in chymotrypsin. R_2 = aromatic amino acid; R_1 = any other amino acid.

Certain organophosphate compounds can selectively react with an acidic serine residue thereby knocking off enzyme activity. Nerve gas which was unfortunately used in the first world war had volatile serine alkylating compounds which inactivates the brain enzyme acetyl choline esterase

leading to death. Nowadays derivatives of organophosphates such as malathion and parathion which are not toxic to humans are used as mosquito repellants (Mortein, Good Knight) by effecting nerve transmission in insects.

Molecular Disease- Sickle cell anaemia

Sickle cell anaemia is a disease prevalent in parts of Africa and India where malaria is also endemic. The red cells of the patient have a pronounced morphological change and resemble the shape of a farmer's sickle and thus the name of the disease. Because these unusually shaped red cells have impaired oxygen carrying capacity and further get stuck in the small capillaries they lead to the anaemic conditions observed in patients. Interestingly such sickled RBCs resist malarial infection and hence offer some selection unfortunately for malaria to be co-prevalent with sickle cell anaemia. One of the first attempts to study the molecular basis of sickle cell anaemia was to compare the electrophoretic mobility of normal (Hb) and sickle cell haemoglobin (scHb). On finding that Hb moved faster than scHb, Linus Pauling predicted that the latter differed in a charged amino acid. This was confirmed by V. M. Ingram in 1957 who pioneered a useful technique called protein finger printing in the famous Laboratory of Molecular Biology (LMB) at Cambridge, UK. LMB has been the Mecca for protein sequencing, DNA sequencing, X-ray crystallography, deduction of the Double helix structure of DNA, Hybridoma technology and Nematode developmental studies. Established in 1952 under the leadership of Max Perutz (Received Nobel Prize for the structure of Haemoglobin) this institution has produced 9 Nobel Prize winners.

Protein Finger printing- Peptide Mapping

This technique involves the generation and 2-D analysis of peptides from a protein. Each protein has a unique peptide map (2-D analysis) and hence serves as a fingerprint for the protein. The steps involved in generating a peptide map/fingerprint are as follows **(Fig. 6)**.

1. Pure Hb and scHb are taken separately into test tubes.
2. The Hb and scHb are digested with the proteolytic enzyme trypsin which cleaves the protein after basic amino acid residues Arg and Lys.
3. Two separate strips of Whatman filter paper are spotted with Hb and scHb tryptic peptides and the peptides allowed to separate using the technique of paper electrophoresis at pH 2.0. Highly charged peptides will migrate more towards the anode/cathode.
4. The paper strips are dried, attached to larger squares of Whatman paper and chromatographed at right angles to the electrophoretic direction using a solvent system Butanol: Water:Acetic acid. In such a system peptides will separate based on their partition coefficient between the solvent and paper which is dependant on the relative hydrophobicity of the peptides. More hydrophobic peptides will move with the solvent to longer distances.

5. The chromatograms are dried and stained with a suitable visualisation reagent like Ninhydrin wherein peptide containing regions appear as orange yellow spots.

6. The peptide map for Hb and scHb are compared and it was found that one peptide was differently placed in the scHb map.

7. On examining this peptide and determining its amino acid sequence, Ingram found that it had a valine substitution for glutamic acid in the peptide.

   The single substitution of valine for glutamic acid (val/glu are at the 6th position of the haemoglobin beta chain) dramatically changes the structure of scHb making it form fibres within the RBC resulting in the deformation of the cell (sickling). Since the disease was due to a molecular alteration the term molecular disease was applied.

Peptide mapping became a useful technique to compare similar proteins from different sources. Slowly the information became too vast and computers were used to store this data into databases so that homology searches could be made. The protein fingerprinting data has been further augmented with new databases containing 2-D electrophoresis patterns of entire proteins from a given cell type, a technique developed by O'Farrel.

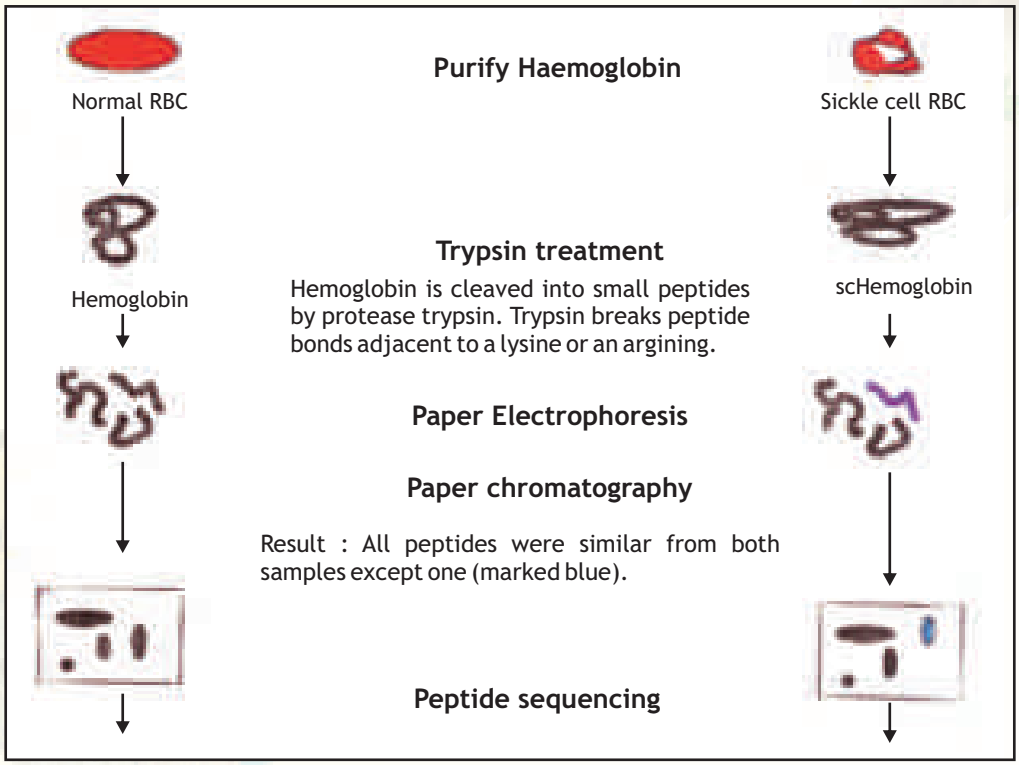


Fig. 6. Protein fingerprinting

2-D Gel Electrophoresis

Two different techniques are combined in this procedure- Isoelectro focussing (IEF) and SDS-PAGE

(Fig. 7).

In simple electrophoresis, the mobility of proteins is due to their charge, which is pH dependant. At its isoelectric pH (pI), a protein does not possess any charge and thus will not move in an applied electric field. This feature is exploited in the technique of IEF, which separates proteins on the basis of their different pI values. Usually IEF is performed in thin tube gels. A pH gradient is set up within the IEF gel by the inclusion of polymeric buffers known as ampholytes. These, like proteins have many positive and negative charges and hence varying pIs. Because of their smaller sizes they move rapidly in an electrophoretic run setting up pH gradients when they come to rest at specific distances from the anode/cathode when they have no net charge. A protein sample from a cell or any other source is then electrophoresed within these tubes wherein the different proteins separate and migrate to their pI zones. The tubes containing the separated proteins is then laid on a SDS-PAGE slab gel and electrophoresis continued at right angles to the IEF direction.

In SDS-PAGE proteins separate on the basis of their size and hence at the end of this electrophoretic run proteins are separated into 2-D patterns with high resolution as two properties of the proteins have been exploited in their separation- charge and size. Proteins in the gels are stained with silver stains or other highly sensitive dyes and can be scanned, and pictures stored into computer databases for analysis.

5.2.4. Purification of Proteins

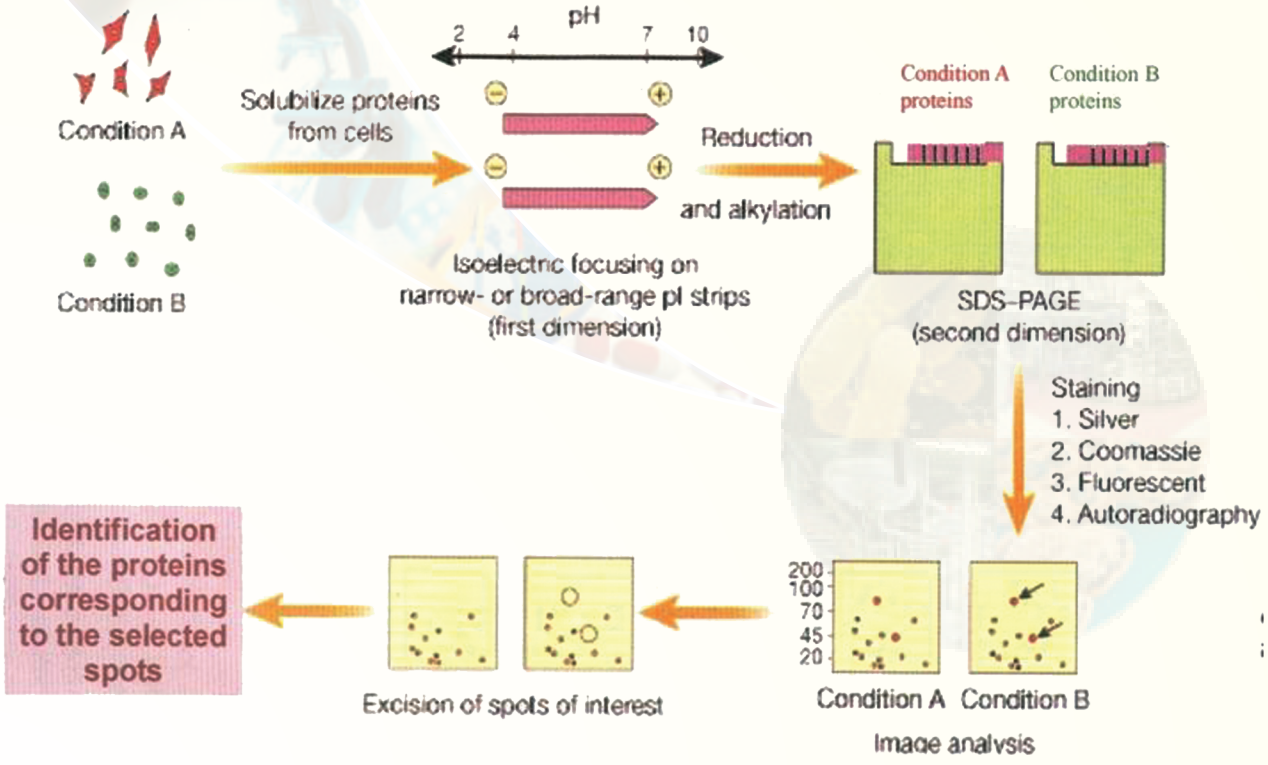


Fig. 7. Two dimentional gel electrophoresis

Isolation of a protein from a microbial culture, plant and animal sources involves various separation techniques (refer Textbook of Biotechnology for Class XI). These steps are collectively known as **downstream processing**. In spite of a large biodiversity of microbes we are restricted to certain bacteria/organisms which can be used as a source of protein as well as for introducing genes. These microorganisms are designated as "generally regarded as safe" (GRAS). **GRAS listed organisms are non-pathogenic, non-toxic and generally should not produce any antibiotics.** Similarily, plant tissue derived enzymes which have application in the food industry must be obtained from only non-toxic, edible plant species. One of the best known industrially useful enzymes is **papain** obtained from the latex of the green fruit and leaves of the Papaya tree. This enzyme finds application in meat tenderisation, clarification of beverages, digestive aids and wound cleaning solutions.

The existence of slaughter house facilities in which large number of animals are regularly processed to provide meat has also facilitated the collection of significant quantities of a particular tissue required as a protein source. Insulin is a classic example of a peptide hormone obtained from the pancreas of cows and pigs till the 1980's. Classical biotechnology required the slaughtering of over 100 pigs or 15 cows to meet the insulin requirements for one diabetic person for about one year. A rough estimate puts the number of diabetics in India by the year 2020 as 20 million! Obviously the requirement of insulin cannot be met by slaughtering pigs and cows. Fortunately, the advent of genetic engineering has ensured the availability of recombinant human insulin expressed in bacteria. Attempts are on to create transgenic animals by direct micro-injection of DNA into ova or stem cells and produce insulin and other proteins in milk on a commercial scale. This technology is called Molecular **Pharming** (Producing pharmaceuticals using genetically modified plants or animals). Advantages of producing recombinant proteins in milk are:

1. High production capacity.
2. Ease of source material collection (milking cows).
3. Moderate capital instrument requirements and low operational cost.
4. Ease of production including purification and scale-up.

Some medically useful peptides such as oxytocin have also been produced by direct chemical synthesis. In spite of different sources of proteins, the general principles of purification are similar. The overall approach and techniques are outlined in Unit-II, Chapter-III of the class XI, Textbook of Biotechnology. The exact details of the purification scheme for any given protein will depend upon a number of factors such as:

1. Exact source material chosen and location of the target protein (intracellular or extracellular).

2. Quantity of protein required and hence amount of raw material processed.
3. Physical, chemical and biological properties of the protein.

Calculation of amount of bacterial ferment required

Question: An *E. coli* cell produces at least 2000 different proteins. One of these is our enzyme of interest produced at a level of 3000 molecules per cell under optimum conditions. If we have to purify 1g of this intra-cellular enzyme, estimate how many cells of bacteria will be required theoretically? It is given that the molecular weight of the enzyme of interest is 1,00,000.

Answer: 1,00,000 g of the protein of interest corresponds to 1 mole of enzyme which corresponds to 6.023×10^{23} molecules (Avagadro no.)

Hence 1 g of enzyme has $1/1,00,000 \times 6.023 \times 10^{23}$ or 6.023×10^{18} molecules .

3000 molecules of the enzyme are present in one cell.

Therefore, 6.023×10^{18} molecules are present in $6.023 \times 10^{18}/3000 = 2.007 \times 10^{15}$ cells.

Question: Assuming that the bacterial cell is a cylinder (d = 1µm, h= 2µm) calculate (a) the total packed cell volume of *E.coli* required to produce 1 g of intra-cellular enzyme (b) the volume of the fermentor required if the maximum cell concentration inside the fermentor is 5% (cells need space to multiply).

Answer: Volume of a single bacterium = πr2h (cylinder volume) = $3.142 \times 0.5 \times 0.5 \times 10^{-12} \times 2 \times 10^{-6}$

(note 1µm = 10^{-6}m, r = ½d) $= 1.57 \times 10^{-18}$ m^3

2.007×10^{15} cells (see previous answer) would have a volume of $2.007 \times 10^{15} \times 1.57 \times 10^{-18}$ m^3

$= 3.15 \times 10^{-3}$ m^3 = 3.15 L Answer (a)

(1L = 10^{-3} m^3)

Answer (b) 100% concentration = 3.15 L

Therefore 5% concentration = 100/5 x 3.15 = 63 L

Volume of the fermentor required would be more than 63 L (30% extra space) about 82 L.

The source material chosen will dictate the range and type of contaminants present in the starting material. If the protein is extracellular, then one needs to separate the cellular components and process the medium to isolate the protein of interest. However if the protein is intracellular then the choice of method of cell disruption will depend on the cell type. Plant and fungal cells require harsher breakage methods; animal cells are easier to break because of no cell wall. Bacterial cells being very small require high pressure techniques. Once the proteins are

released into suitable buffered solutions a variety of physico-chemical techniques are applied to selectively purify the protein of interest from the others.

Genetically engineered proteins are often tagged with certain molecules in order to confer some very pronounced physico-chemical characteristics on the protein of interest. This renders its separation from contaminants more straightforward. The ability to detect and quantify the total protein levels is an essential pre-requisite to the purification and characterisation of any protein. A typical purification scheme can be analysed as follows **(Table 1)**.

Table 1. Typical purification table

Procedure	Total protein (mg)	Activity (units)	Specific activity units/mg
Crude extract	20,000	40,00,000	200
Precipitation (salt)	5,000	30,00,000	600
Precipitation (pH)	1,000	10,00,000	1000
Ion-exchange chromatography	200	8,00,000	4000
Affinity chromatography	50	7,50,000	15,000
Size exclusion chromatography	45	6,75,000	15,000

Note that the last column is a good indication of whether a purification step is useful or not. This is because as a protein is purified its specific activity increases because the denominator should ideally decrease as irrelevant proteins are removed and only the specific protein/enzyme of interest is concentrated. Hence from the given table it is apparent that the step following affinity chromatography, size exclusion chromatography, is redundant as the specific activity does not change. For activity measurements it is also important to choose a proper assay method reflecting the sensitivity required and further the method should be specific. In the case of proteins absorbance measurements at 280 nm is easy, fast and non-destructive procedure for monitoring the concentration.

Bioassays can sometimes be more sensitive than chemical assays. However one needs suitable standards of known bioactivity values to arrive at a correct activity of the unknown sample. Where a protein of biological interest is concerned, example insulin, bioassays are mandatory. If the sample particularily has to be injected other safety tests such as toxicity have to be performed.

Downstream Processing

After cells (bacterial, animal or plant) have grown to their requisite capacity in a fermentor it becomes necessary to harvest the cells or medium depending in which component the

recombinant protein is expressed and then purify the protein from other substances. These processes are part of downstream processing and because of the large amounts of source (a fermentor can be more than 1000 L capacity) bulk separation methods are used which are different from laboratory scale purification although principles involved are similar.

In the case of intracellular microbial proteins, cell harvesting is done by filtration or centrifugation from the fermentation medium, followed by re-suspension of cells in buffer or water with subsequent cell disruption. Most proteins obtained from plant and animal tissues are intracellular in nature. The initial step involves collection of the appropriate tissue, for example collection of blood to obtain proteins, collection of pituitary glands to obtain pituitary hormones etc.

Aqueous two-phase partition

When a crude cell homogenate is added to a biphasic mixture of dextran and polyethylene glycol (PEG) the cellular debris partitions to the lower, more polar and dense phase, dextran. Separation of the two phases achieves effective separation of cellular debris from soluble protein **(Fig. 8)**.

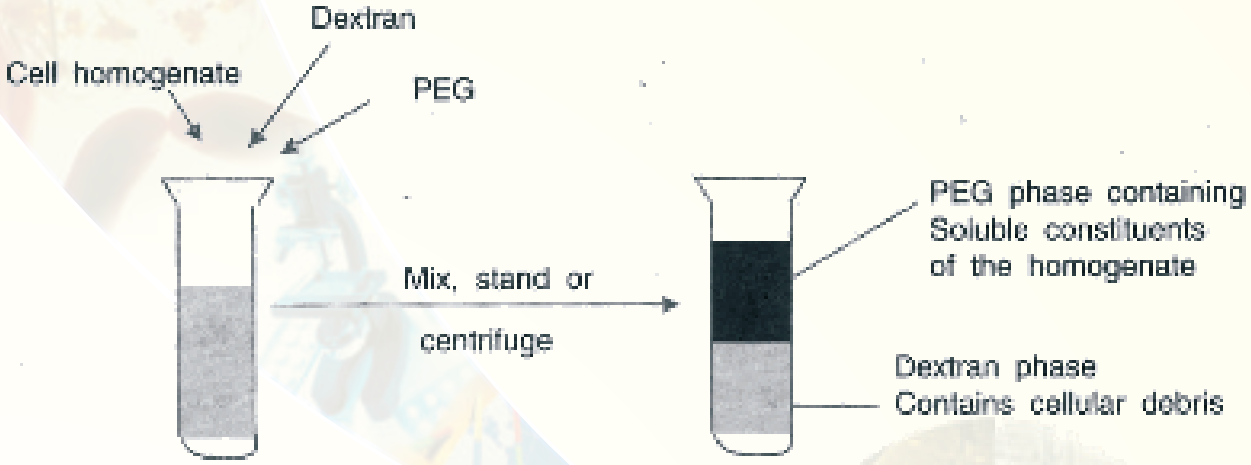


Fig. 8. Two phase separation

In some cases it is desirable and necessary to remove or destroy the lipids and nucleic acid of a cell homogenate as it may be a contaminant and can interfere with subsequent purification steps. The lipid layer can be removed by passage of solution through glass wool or cloth of very fine mesh size. Effective removal of nucleic acids may be achieved by precipitation or by treatment with nucleases.

For large scale application, concentration of extracts is normally achieved by precipitations, ion-exchange chromatography or ultrafiltration.

At any given pH value proteins display either a net positive, negative or no charge. Using these parameters different protein molecules can be separated from one another by judicious choice of pH, ionic strength and ion-exchange materials.

All efforts should be made to maximise protein stability during various steps. Some of the general conditions which may be followed are:

1. Maintenance of a specific pH value range of buffered solutions in which a protein is maximally stable.
2. Maintenance of physiological conditions (%CO_2 for animal cell culture and temperature).
3. Use of inhibitors to prevent the action of proteolytic enzymes.
4. Avoidance of agitation or addition of chemicals which may denature the target protein.
5. Minimise processing time.

Industrial scale production of proteins

The laboratory scale design cannot be scaled up to industrial scale directly. The following points need attention for industrial scale production:

1. Bulk purchase of chemicals and other raw materials would bring down costs.
2. The labour cost decreases sharply with increase in production.

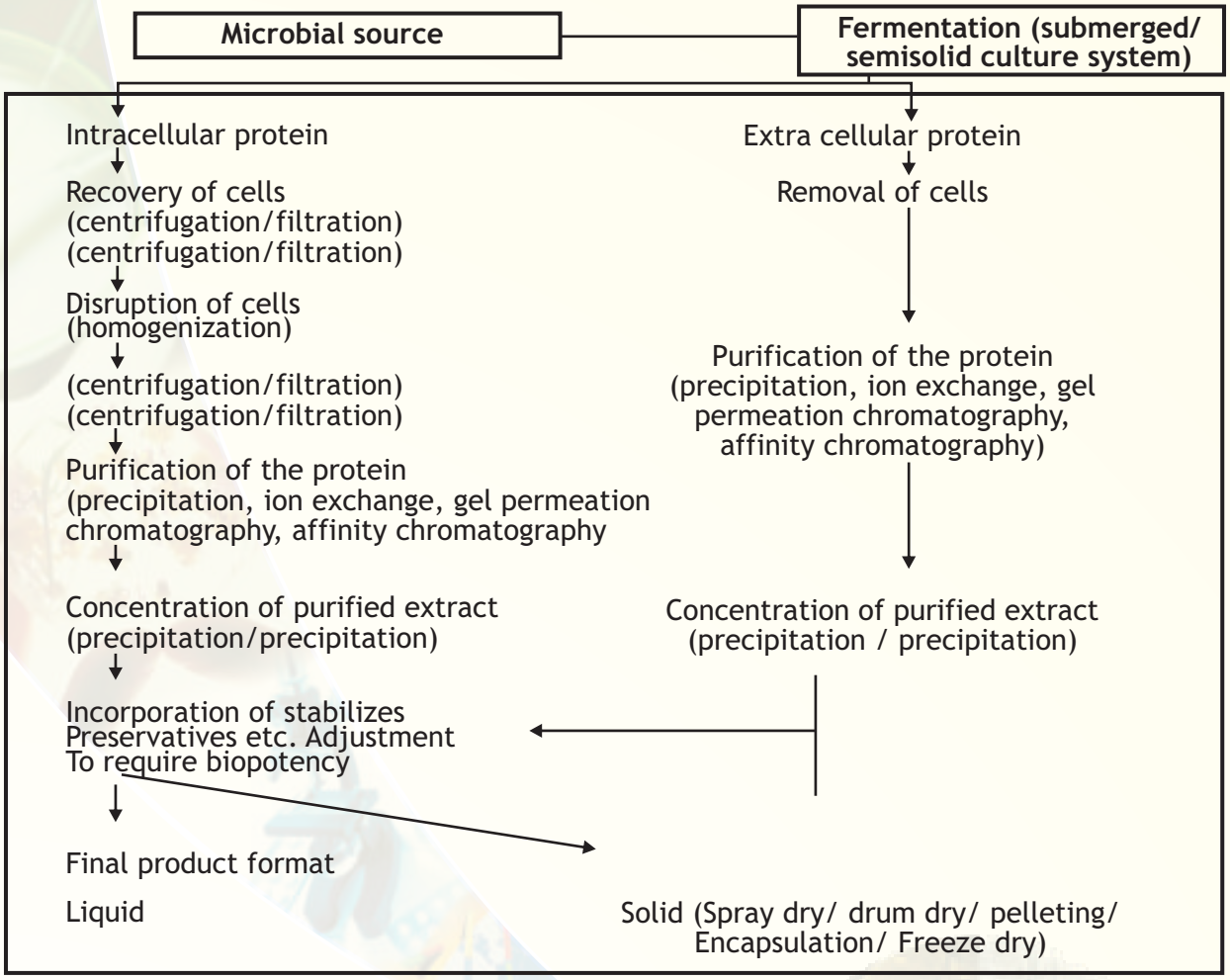
Most large scale process equipment such as holding vessels and transfer pumps are constructed from stainless steel or plastics, such as polypropylene. Glass vessels, so commonly used in laboratory scale culturing techniques are seldom used for large scale preparatory work. Materials used for large scale culturing must be inert and resistant to the corrosive action of any chemical used during the process. They should not allow any leaching of potentially toxic metals or chemicals into the product stream. It is useful to remember that any commercial plant has to have good GMP (good manufacturing practice) . Any downstream processing requires the approval of a regulatory authority in the form of a license to produce and market proteins designed for use in the food or health care industry. This ensures that the processing procedures are based upon established, validated methodologies. Generally 80% of the overall production cost are due to steps in downstream processing and quality assurance.

A generalised downstream processing scheme used in the production of bulk protein/enzyme from microbial sources is given in **Fig. 9.** Similar steps can be applied for animal and plant sources.

Special techniques for therapeutic /diagnostic proteins

These proteins must be purified to a very high degree especially for use in parenteral (injectable) administration. They also have to be sterile products (free of bacterial and fungal contamination) and hence can be administered by injection, infusion or implantation.

Fig. 9. A typical flow sheet from source to product.



5.2.5. Characterisation of Proteins

Techniques listed below characterise proteins with respect to properties such as mass, isoelectric charge, amino-acid sequence etc. Alongside they can also detect impurities as these are very sensitive techniques requiring small amounts, often micrograms of the sample. The first four techniques have already been discussed in this chapter.

1. Electrophoretic techniques, SDS/PAGE.
2. Fingerprinting.
3. Two dimensional gel electrophoresis.
4. Protein sequencing.
5. Mass spectrometry.

Mass spectrometry

Mass spectrometry (MS) has emerged as an important tool in biotechnology. It is extremely useful in obtaining protein structural information such as peptide mass or amino acid sequences. It is also useful in identifying the type and location of amino acid modification within proteins. One of the major attractions of mass spectrometry is that as little as picomoles (10^{-12}) of a protein sample can be analysed. A mass spectrometer is an analytical device that determines the molecular weight of chemical compounds by separating molecular ions according to their mass/charge ratio (m/z) ratios. The molecular ions are generated either by a loss or gain of a charge (e.g. electron ejection, protonation or deprotonation). After the ions are formed they can be separated according to their m/z ratio and finally detected. The process of ionisation, ion separation and detection in a mass spectrum can provide molecular weight or even structural information. A sample M with a molecular weight greater than 1200 D can give rise to multiple charged ions such as (M+nH)n+. Proteins/peptides have many suitable sites for protonation as all the backbone amide nitrogen atoms could be protonated theoretically as well as certain amino acid side chains such as lysine and arginine which contain primary amine functional groups.

A schematic diagram of the various parts of a mass spectrometer is indicated in **Fig.10.** Basically a vapourised sample of a protein or peptide is introduced into the instrument wherein it undergoes ionisation. The charged molecules are then electrostatically propelled into a mass analyser (filter) which separates the ions according to their m/z ratio. The signal received upon detection of the ions at the detector is transferred to a computer which stores and processes the information.

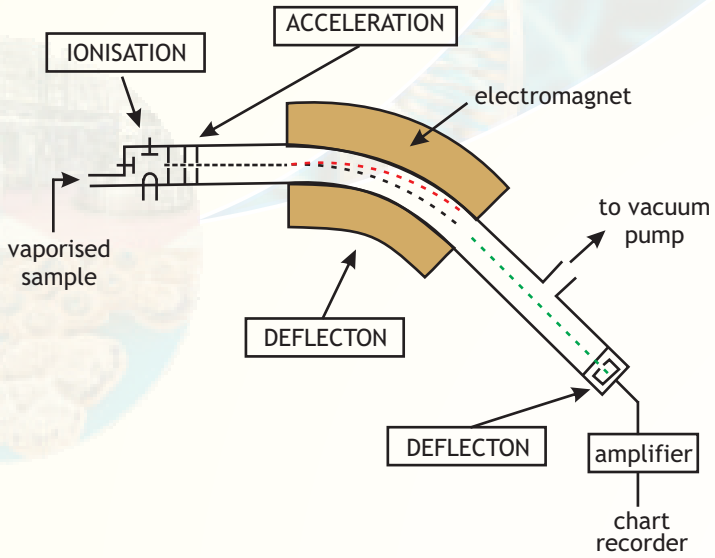


Fig. 10. An outline of a mass spectrometer.

The goal of mass spectrometric analysis of biomolecules like peptides and proteins is to create gas phase ions from polar charged molecules which are generally non-volatile. A popular method called Matrix Assisted Laser Desorption Ionisation (MALDI) is used to volatalise and protonate peptides and proteins. In this procedure, the sample is transferred from a condensed phase to a gas phase with the help of a solid matrix. Ion formation in MALDI is achieved by directing a pulsed laser beam onto a sample suspended or dissolved in a matrix. The matrix plays a key role in this technique by absorbing the laser light energy and causing the matrix material to vaporise. In the gas phase, the matrix plays a role in sample ionisation. The charged molecules are directed by electrostatic lenses from the ionisation source to the mass analyzer.

Multiply charged ions on analysis show patterns as indicated in Fig. from which molecular mass can be deduced as also indicated in the legend below **Fig. 11.**

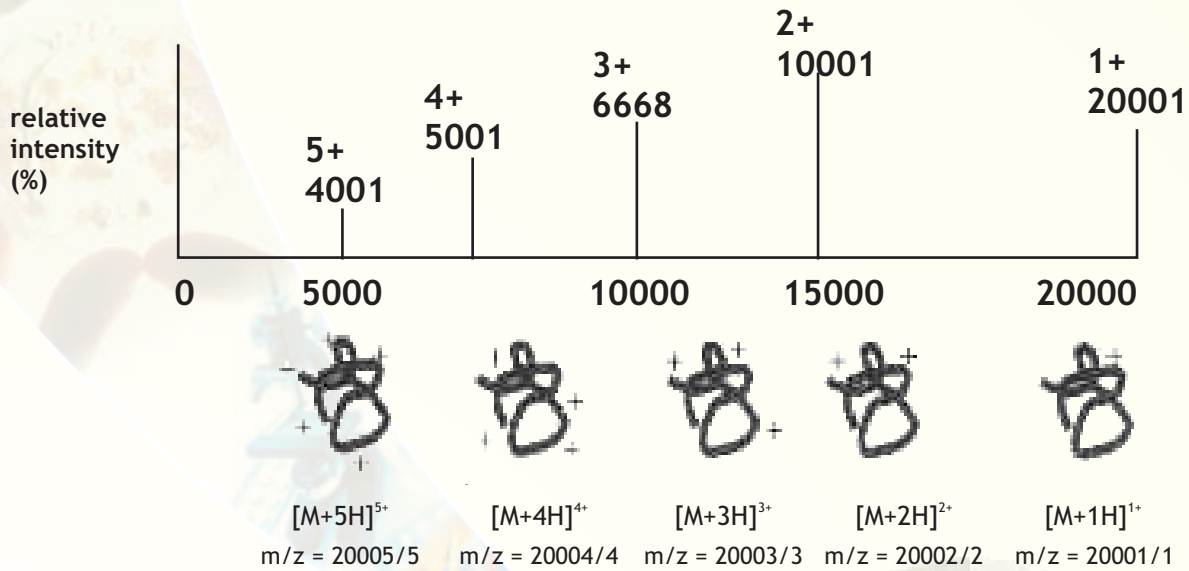


Fig.11. The same protein with a molecular weight of 10,000 contains 5, 4, 3, 2 and 1 charges. The mass spectrometer detects the protein ions at m/z = 2001, 2501, 3334, 5001 and 10,001 respectively.

Typically for protein identification, a crude extract is separated on a 2D gel. After visualising the proteins on the gel, protein spots are excised and used for mass analysis by the MS technique described above. For doing so the protein in the 2D gel after extraction from gel is either used intact or it is cleaved into small peptides with a protease like trypsin which makes the mass analysis easier. These peptides are separated on an on-line liquid chromatography system before introduction into the mass spectrometer. Liquid chromatography techniques like ion exchange, affinity or reverse phase column chromatography can be used to separate the peptides. These peptides are either sequenced directly or the mass of peptides is analysed using database searches (see Bioinformatics unit). With the human genome sequencing project, it is now possible to identify new proteins by combining the mass spectrometric information with the genomic

information using Bioinformatics tools.

5.2.6. Protein Based Products

From the commercial point of view, proteins may be classified into the following categories.

1. Blood products and vaccines.
2. Therapeutic antibodies and enzymes.
3. Therapeutic hormones and growth factors.
4. Regulatory factors.
5. Analytical application.
6. Industrial enzymes.
7. Functional non-catalytic proteins.
8. Nutraceutical proteins.

Blood products and vaccines

Blood carries out several functions and is one of the best mediums for transportation in an animal. A better understanding of haematapoiesis (formation of blood cells) as well as factors responsible for blood coagulation has led to the discovery of several useful proteins. Several proteins from blood and plasma have been commercially available for decades. While these products have traditionally been obtained from blood donated by human volunteers, some are now produced by recombinant DNA technology. For example Factor VIII for treatment of Haemophilia A, Factor IX for treatment of Haemophlia B, Hepatitis B vaccine for prevention of hepatitis etc.

Therapeutic antibodies and enzymes.

Polyclonal antibodies have been used for more than a century for therapeutic purposes. More recently monoclonal antibody preparations as well as antibody fragments produced by recombinant DNA technology have found therapeutic use. For example tissue plasminogen activator (t-PA) is a proteolytic enzyme used to digest blocks in arteries following myocardial infarctions. A monoclonal antibody OKT-3 is used to prevent rejection following kidney transplantation because the antibody blocks those immune cells which attack foreign grafts.

Therapeutic hormones and growth factors

A number of hormone preparations have been used clinically for many decades. Though insulin was prepared from the pancreas of cows and pigs, the ability to genetically transfer human insulin gene into bacteria and the ability to modify amino acid residues (protein engineering) has

facilitated the development of modified forms which are faster acting like humulin. Humulin acts in 15 min unlike pig insulin which takes 3 hours. Another growth factor- platelet derived growth factor has been approved for diabetics who develop skin ulcers. Several other growth factors are under various stages of clinical trials.

Regulatory factors

Several new regulatory factors were discovered that did not fit the classical definition of a hormone. Initially they were known as cytokines. These include interferons, interleukins, tumor necrosis factor and colony stimulating factors. The interferon family of INF alpha, beta and gamma have found widespread therapeutic application; interferon alpha is used for treatment of Hepatitis C, beta for Multiple Sclerosis and gamma for Chronic Granulomatous disease.

Analytical applications

Enzymes and antibodies have found a range of analytical applications in the diagnosis of diseases; hexokinase for quantitative estimation of glucose in serum, uricase for uric acid levels in serum, horse radish peroxidise and alkaline phosphates in ELISA etc.

Industrial enzymes

Proteolytic enzymes constitute an 8000 crore annual market for industrial enzymes. They find application in the beverage industry, detergent industry, bread and confectionary industry, cheese production, leather processing and meat industry. Alcalase is an enzyme used in the soap industry, papain is used in the beverage industry, glucose isomerise in the confectionary industry and chymosin is used in the cheese industry.

Functional non-catalytic proteins

Functional non-catalytic proteins are those which have properties such as emulsification, gelation, water binding, whipping and foaming etc. **(Table 2)**. For example kappa casein, a component of casein is involved in micelle stabilisation of milk proteins and keep the proteins suspended uniformly in milk because it behaves like a lipid molecule (2/3rd of the protein is hydrophobic). The food industry has exploited these non-catalytic proteins as illustrated in the below.

Table for whey protein.

TABLE : 2

Functional Property	Mode of action	Food System
Whipping/Foaming	Forms stable film	Egg less cakes, desserts, whipped topping
Emulsification	Formation and stabilization of fat emulsions	Vegetarian sausages, salad dressings, coffee whiteners, soups, cakes, infant food formulas, biscuits.
Gelation	Protein matrix formation and setting	Meat, baked goods, cheeses
Viscosity	Thickening, water binding	Soups, gravies, salad dressings
Water binding	Hydrogen bonding of water; entrapment of water	Meats, sausages, cakes, breads
Solubility	Protein solvation	Beverages
Browning	Undergoes Maillard reaction (on heating, the amino groups of protein react with aldehyde groups of sugars)	Breads, biscuits, confections, sauces
Flavour/Aroma	Lactose reacts with milk proteins	Baked goods, biscuits, confectionaries, sauces, soups, dairy products.

Nutraceutical Proteins

Nutraceutical is a word coined from combination of nutrition and pharmaceuticals. It has been observed that several nutritional proteins also have therapeutic functions. For example whey protein concentrates, lactose free milk (for lactose intolerant babies) and infant food formulations.

Where does one get the raw building materials such as amino acids needed to make all body proteins? During infancy we depend on milk. Baby milk formulations are also there (Amul, Lactogen etc.) which have been formulated to have similar composition as mother's milk. All these food materials provide the essential components nutritionally for growth and development during the first few months of our existence. A typical composition of milk from buffalo, human and cow sources is given in the **Table 3** from which baby milk formulations can be made to suit an infant.

Table 3. Composition of milk from buffalo, human and cow.

Constituents (per 100 ml of milk)		buffalo	human	cow
1.	Protein (g)	3.8	1.2	3.3
2.	Casein (g)	3.0	1.4	2.8
3.	Lactalbumin (g)	0.4	0.3	0.4
4.	Lactoglobulin (g)	0.2	0.2	0.2
5.	Fat (g)	7.5	3.8	3.7
6.	Lactose (g)	4.4	7.0	4.8
7.	Calorific value (K Cal)	100.0	71.0	69.0
8.	Calcium (mg)	203.0	33.0	125.0
9.	Phosphorous (mg)	130.0	15.0	96.0
10.	Chloride (mg)	112.0	43.0	103.0

From the **Table 3** it can be observed that milk contains several proteins, carbohydrates, lipids, vitamins, antibodies, minerals etc. It is interesting to note that human milk has nearly half the amount of casein as compared to cow and buffalo. Besides use of milk as a nutritional source, claims have been made to the effect that curd is beneficial in the management of some types of intestinal infections according to our ancient Sanskrit scriptures dating back to 6000 BC. Since time immemorial whey (liquid part of curds) has been administered to the sick for the treatment of numerous ailments. In 1603, Baricelli reported on the therapeutic use of cow's or goat's whey, sometimes mixed with honey and herbs. The spectrum of illnesses treated with whey include jaundice, infected skin lesions, genitor-urinary tract infections. Gallen and Hippocrates insisted on a minimum daily drinking of one litre of whey. Using modern scientific research it has been possible to explain these observations. Whey proteins result in the elevation of a tripeptide glutathione(gamma-glutamyl cysteinyl glycine) in cells. This peptide is a reducing compound and has a broad range of functions including detoxification of xenobiotics and protection of cellular components from the effect of oxygen intermediates and free radicals. More recently curd has also been used as a pro-biotic (administered with antibiotics) because it is a good source of beneficial bacteria which can colonise the intestinal tract. **Table 4** gives the useful components of whey.

Table 4. Whey components.

α - Lactalbumin	Lactose
α - Lactoglobulin	Sialic Acid
Bovine Serum Albumin	Lactic Acid
Immunoglobulins	
Lactoferrin	Sodium
Lactoperoxidase	Potassium
	Calcium
	Magnesium
Protease peptones/polypeptides	Chloride
Free amino acids	Phospphate
Urea	Sulfate
	Citrate
Glycomacropeptides	Heavy Metals
Growth Factors	Milk Fat
	Globule
	Free Fat
	Lipoproteins

5.2.7. Designing Proteins (Protein Engineering)

Considerable interest exists in the biotechnology industry for the engineering of proteins with increased stability when exposed to harsh conditions like elevated temperature, organic solvents and reactive chemicals, often encountered in the industrial processes. Besides, it is of great interest to explore biological adaptations to environmental stresses such as high salinity, drought, cold etc. Therefore, in order to stabilise your favourite protein, it is essential to know the cause of inactivation.

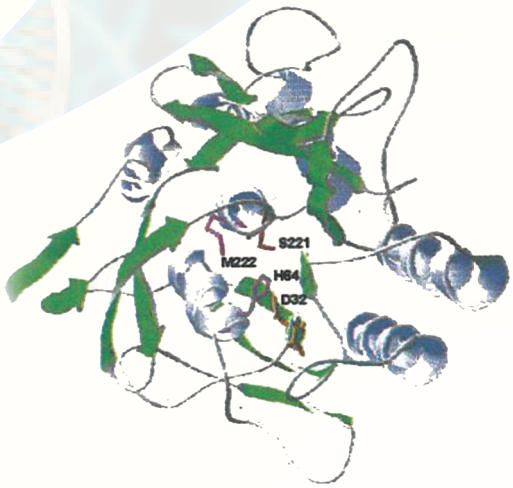


Figure 12. Subtilisin with the catalytic triad

Stability in a folded protein is a balance between the stabilising (mainly hydrophobic) interactions and the tendency towards destabilisation caused by the loss of conformational entropy as the protein adopts the unfolded form. The stability of a protein may however be changed by substituting amino acids that either favour stabilising interactions in a folded protein or destabilising interactions in an inactive protein. Numerous attempts have been made on different

proteins and enzymes in order to improve their properties for thermal and pH stability, solvent tolerance and solubility, catalytic potency etc. Given below is an example which has been successfully used in the detergent industry.

Improving laundry detergent Subtilisin

Subtilisin (27 kD) is a protease produced by bacteria that can digest a broad range of proteins that commonly soil clothing, see **Fig. 12**. The enzymatic activity of subtilisin is contributed by a catalytic triad, i.e., Ser221, His64 and Asp32 similar to chymotrypsin. Replacement of all three residues with alanine either singly or in combination results in significant loss of activity. Subtilisin represents the largest industrial market for any enzyme. To improve the efficiency of laundry detergents, detergent manufacturers supplement subtilisin in their products with various catchy slogans on the detergent box such as "stain cutter" or "biologically active enzymes".

The native enzyme subtilisin is easily inactivated by bleach (up to 90%). Careful studies showed that this inactivation was due to oxidation of the amino acid residue Methionine222 in the protein molecule **(Fig.12)**. Using site-directed mutagenesis of the subtilisin gene in *E.coli*, this methionine was substituted by a variety of other amino acids and the enzyme activity measured in the presence of bleach **(Table 5)**. It was observed that substitution of Met222 with Ala222 was the best in terms of activity and stability. Nowadays, many laundry detergents contain cloned, genetically engineered or recombinant subtilisin.

Table 5. Site-directed mutagenesis at codon position 222.

Codon-222	% activity w.r.t wild type	Codon-222	% activity w.r.t. wild type
Cys	138.0	Gln	7.2
Met	100.0	Phe	4.9
Ala	53.0	Trp	4.8
Ser	35.0	Asp	4.1
Gly	30.0	Tyr	4.0
Thr	28.0	His	4.0
Asn	15.0	Glu	3.6
Pro	13.0	Lie	2.2
Leu	12.0	Arg	0.5
Val	9.3	Lys	0.3

Creation of Novel Proteins

Conventional vaccines have utilised heat inactivated bacteria/viruses or their surface proteins to generate immunity against various specific diseases. Often it has been observed that some of the components of the vaccines have undesirable effects such as fever and rarely one has also heard of some of the components like virus actually causing the disease due to incomplete inactivation. Since proteins are the main molecules which provide the stimulus for immunity, attempts have been made to engineer proteins having minimum deleterious effects. The specific sequences of amino acids in the protein which stimulate immune response are known as epitopes. A recombinant vaccine based on selected epitopes may provide optimal design, scope for micromanipulation, unhindered supply and safety needed for an effective vaccine. Working on these lines, a novel synthetic gene has been assembled as a first step towards developing a subunit vaccine against Hepatitis B virus.

Improving nutritional value of cereals and legumes

The cereal grains and seeds of legumes constitute a major chunk of dietary protein requirement. The seed storage proteins are synthesised and accumulated throughout seed development to serve as source of amino acid reserves at the time of seed germination. High levels of such proteins in seeds would provide an enriched amino acid source for human consumption. However deficiencies in seeds of certain essential amino acids render the cereal grains or legumes unsuitable for a balanced diet. Supplementation of diet with essential amino acids from other sources therefore becomes essential. **Fig.13** gives the essential amino acid content of various cereals and commonly eaten food proteins. Essential amino acids are those which have to be obtained from food and cannot be made in our cells. From the data in the figure it is apparent that whey protein is superior to other sources especially with regard to branched amino acids- ile, leu, val, lys and trp. The branched chain amino acids (BCAA) are essential for the biosynthesis of muscle proteins. They help in increasing the bio-availability of high complex carbohydrates intake and are absorbed by muscle cells for anabolic muscle building activity. One of the theories is that during exercise the BCAAs are released from the skeletal muscle; the carbon skeleton part is used as fuel and the nitrogen part is used to make alanine which then goes to the liver where it is turned into glucose for energy. So for athletes who want to protect their existing mass, the idea is to take BCAA enriched foods before and after excercise. BCAAs reduce muscle breakdown and act as an energy source before and after exercise. Hence while maintaining exercise performance and delaying exhaustion BCAAs are very important for muscle growth. Nowadays an entire new area of sports medicine and nutrition prepare and recommend special nutrient drinks etc. which incorporate these principles. In the unit on plant tissue culture you will read how plant cereals have been genetically engineered for higher nutrient value in terms of proteins, vitamins etc.

Biological value (BV) measures the amount of protein nitrogen that is retained by the body from a given amount of protein nitrogen that has been consumed. It has been observed that the BV of whey proteins is the highest compared to rice, wheat, soya and egg proteins. Another index of protein value is the **protein efficiency ratio** (PER). PER is used as a measure of growth expressed in terms of weight gain of an adult by consuming 1g of food protein. The PER value of the following proteins are arranged in decreasing order- whey, milk, casein, soya, rice, wheat. The modern day approach for overcoming the nutritional deficiencies of seeds would be to engineer genes that would encode storage proteins with more of the nutritionally desirable amino acids either by inserting additional amino acids or substituting existing amino acids with new ones. Attempts are already being made on zein storage protein genes of maize to enhance its nutritional value. Introduction of entirely novel proteins that are highly enriched in specific amino acids is also being considered.

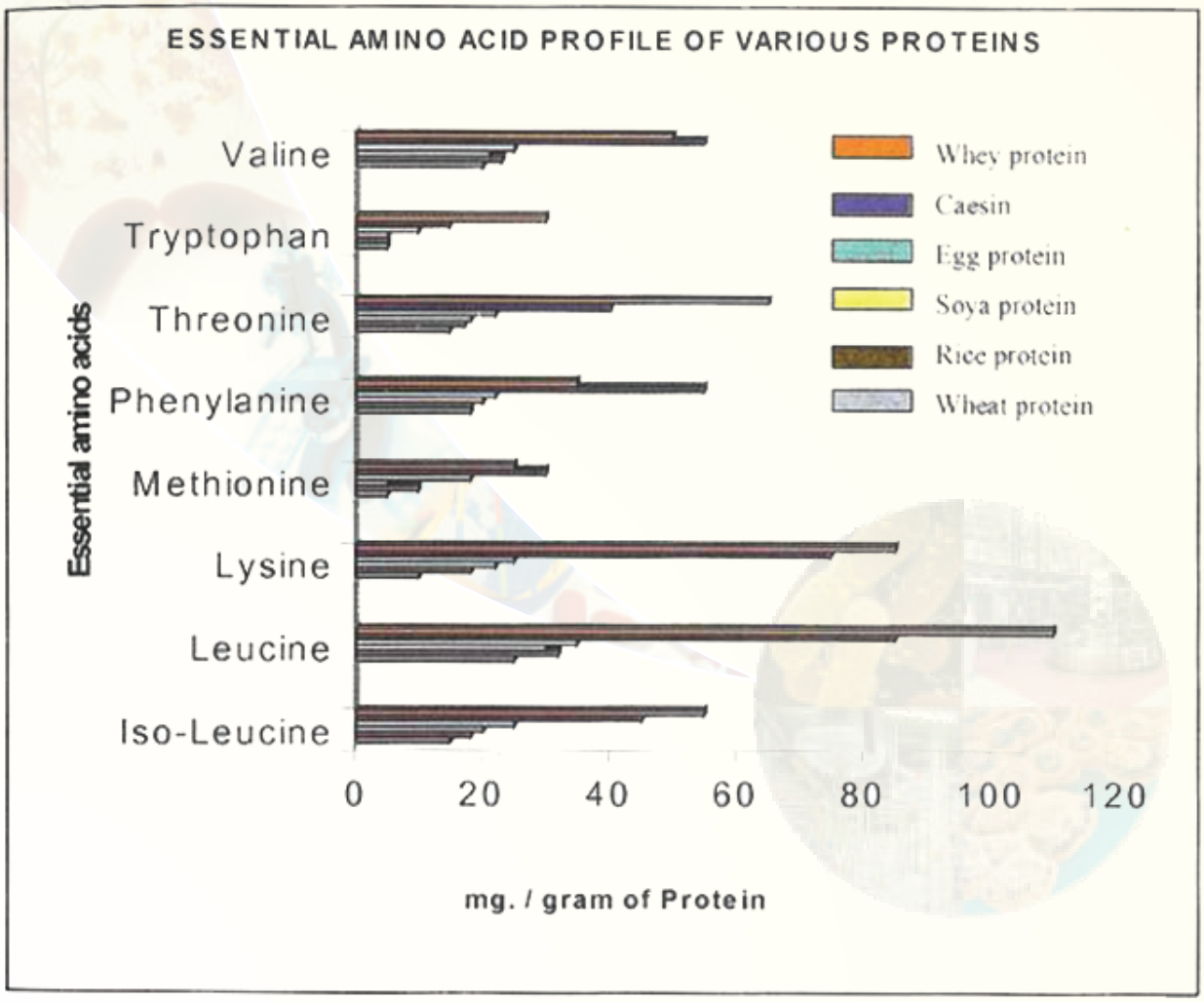


Fig. 13. The essential amino acids profile.

Review Questions

1. Name two human diseases caused by the absence of a protein.
2. What is the consequence if a protein is incorrectly folded? Give an example to illustrate your answer.
3. Distinguish between chymotrypsinogen and chymotrypsin.
4. Briefly explain how the serine residue in some enzymes can become acidic (reactive). Also suggest how you can confirm that a serine residue is involved in the catalysis.
5. Why is sickle cell anaemia called "Molecular disease"? How can sickle cell haemoglobin be identified?
6. What are the principles behind Isoelectric Focusing and SDS/PAGE techniques? Why is 2-D electrophoresis better than single dimension electrophoresis?
7. Define subunit, domain, and quaternary structure in proteins.
8. With an example explain the development of one protein based product.
9. What are non-catalytic functional proteins, therapeutic proteins and nutraceutical proteins? Give one example each.
10. Briefly discuss the use of designing a protein for any product.
11. What is the principle of MALDI-TOF? What is its main use in protein studies?
12. *E. coli* is a rod shaped bacteria about 2 µm long and 1 µm in diameter. The average density of a cell is 1 .28 g/ ml. Approximately 13.5% of the wet weight of E. coli is soluble protein. Estimate the number of molecules of a particular enzyme per cell if the enzyme has a molecular weight of 100,000 and represents 0.1 % of the total soluble protein. (Answer: 1626 molecules per cell).

References

1. Protein: Biochemistry and Biotechnology by Gary Walsh (2002), John Wiley & Sons, Ltd.
2. Practical Biochemistry: Principles and Techniques Edited by Keith Wilson and John Walker (2000), Cambridge University Press.
3. Biochemistry by Lubert Stryer, J.M, Berg and J.L. Tymoczko (2002). W.H. Freeman.
4. Principles of Biochemistry by Albert L. Lehninger, D.L. Nelson and M.M. Cox (2000), W.H. Freeman.
5. Biochemical calculations by Irwin H. Segel (1985), John Wiley & Sons, Ltd.
6. Proteins: Structures and Molecular Properties by Thomas E. Creighton (1992), W.H. Freeman.

CHAPTER

3 GENOMICS AND BIOINFORMATICS

5.3.1. Introduction

The term "GENOMICS" was coined in 1986 by Thomas Roder, to describe the scientific discipline of mapping, sequencing and analyzing genomes. H. Winkler in 1920 had coined the term genome to implicate the complete set of chromosomal and extra chromosomal genes of an organism, a cell, an organelle or a virus.

The field of genomics relies upon bioinformatics, which is the management and analysis of biological information stored in databases. During the mid-1980s to late 1980s, researchers started to use computers as central sequence repository, from where the data could be accessed remotely. Later in the early 1990s, genomics was transformed from an academic undertaking to a significant commercial endeavor, a course followed by bioinformatics a few years later.

In retrospect, the genomics really began with the conception of the Human Genome (HGP) in the mid-1980s. In the United States, the Human Genome Project officially started on October 1, 1990, as a 15-year program to map and sequence the complete set of human chromosomes, as well as those of several model organisms. The goal of sequencing an estimated three billion base pairs of the human genome was ambitious, considering that few laboratories in 1990 had sequenced just 100, 000 nucleotides. By 1993, the Human Genome Project had become an established international effort. The strategy of this international project was to make a series of maps of each human chromosome at increasingly finer resolutions. In this approach, chromosomes were divided into smaller fragments that could be cloned and then the fragments were arranged to correspond to their locations on a chromosome. After mapping, each of these ordered fragments would be sequenced.

Progress in stages

J. Craig Venter, a researcher at the National Institutes of Health, and his colleagues~ in early 1990s devised a new way to find genes. Rather than taking the Human Genome Project strategy of sequencing chromosomal DNA-one base at a time, his group isolated messenger RNA molecules, copied these RNA molecules into DNA, and then sequenced a part of these DNA molecules to create expressed sequence tags, or "ESTs." These ESTs could be used as handles to isolate the entire gene. Venter's method, therefore, focused on the "active" portion of the genome, which was producing messenger RNA for protein synthesis. The EST approach has generated enormous sized databases of nucleotide sequences, and facilitated the construction of a preliminary transcript map of the human genome. The development of the EST technique is considered to

have demonstrated the feasibility of high-throughput gene discovery (screening of all possible gene candidates from the EST library), as well as provided a key impetus for the growth of the genomics industry. After the success of these projects, Craig Venter moved again to sequence entire genomes.

Evolving approaches

He devised the "whole-genome shotgun strategy," which involves randomly breaking DNA into segments of various sizes and cloning the fragments into vectors. Since the fragments are randomly cleaved from the genome, they tend to overlap, and a genome assembly program is used to fit contiguous pieces by matching overlapping ends. This method was validated by sequencing the entire genomes of a few selected microorganisms.

This is how, the first set of whole genome sequences of the smallest genome Mycoplasma genitalium and Haemophilus influenzae Rd were released. To analyze the data, several computer programs had to be adapted to fast computers, Later several new programs were also written to accomplish the task of sequence assembly. Craig Venter established an organization called "The Institute of Genomics Research (TIGR)" located in Maryland U.S.A., and soon whole genome sequences were determined for many other bacteria including those that live in exotic environments such as hot temperature or deep sea vents. Several bacteria of medical importance were also sequenced. During this time, several groups from Europe also initiated whole genome sequencing of bacteria such as Mycobacterium tuberculosis and *Bacillus subtilis* at Pasteur Institute. Generally, in Europe large consortiums (group of organizations in various countries) were formed to complement each other's strengths.

The exciting commercial era of genomics began with the establishment of Celera Genomics that was dedicated to sequencing the human and the mouse genomes, compared to microbial genomes, the human genome is large ~ 3×10^9 bps and also contains lots of repeated sequences. These repeated sequences present difficulties in sequence assembly because doubts arise with regard to their true order of arrangement in the genome. The parts containing the genes were somewhat easier to assemble. The problem of sequence assembly of repeats and of unique sequences by the computer is akin to this example. Suppose you were blindfolded and asked to pick two balls of different size from a lot of mostly identical balls, you would make several attempts but end up with failures most of the time. Further, you may not be able to distinguish one ball from another. However, if you were given the same assignment of drawing two different balls from a lot balls of all different sizes and shapes, then there is a good possibility of you picking up two different balls at much fewer attempts, perhaps the very first attempt itself may be enough.

But the unraveling of the human genome sequence gave us a surprise. Initial EST sequencing had led to an estimate of over 100,000 genes being present in the human genome. The genome sequence however, revealed that there are only about 30,000 genes. This number is only twice

that of the fruitfly *Drosophila melanogaster,* a simple organism compared to the immensely complex human being. Possessing only twice the number of genes of fruitfly challenges us to search for other explanations that underlie the complexity of the humans. It turns out that humans can achieve this through combinations of the genes. You can understand this by an example.

Suppose, a mechanic has a set of 20 or 30 tools, each one dedicated to carry out a specific task. Then the mechanic can accomplish 20 or 30 tasks. However, if the same set of tools had flexible parts, then the same mechanic can generate several 'new combinations' of these tools to carry out hundreds of tasks. Naturally through combinations, this mechanic will have more business, earn more money and will be most sought after compared to using the dedicated 'one job specific' tools.

Below, the sections on genomics explain the different branches of this exciting new area. With the development of automated sequencing, it has been possible to sequence genomes of many organisms. According to the latest list displayed at the NCBI (National Centre for Biotechnology Information) site, there are 1409 complete genome sequences of bacteria and archaea, 40 complete genome sequences of eukaryotes and 2537 complete genome sequences of viruses. The sequencing projects have shown several interesting and unexpected findings. The term genomics itself has undergone expansion in last few years and in the present context also includes genome function. Genomics can be broadly divided into structural genomics and functional genomics.

Structural Genomics

Structural genomics primarily involves high-throughput DNA sequencing followed by assembly, organization and management of DNA sequences. It represents an initial phase of genome analysis, which involves the construction of high-resolution genetic, physical or transcript maps of the organism. The ultimate physical map of an organism is its complete DNA sequence. Although, in the last few years with the completion of several genome-sequencing projects, the term structural genomics has also undergone transition. Several structural genomics initiatives now encompass systematic and high-throughput determination of three-dimensional structures of all proteins. The information and reagents provided by structural genomics are used to design global (genome-wide) experiments to identify functions of proteins.

Functional genomics

Functional genomics represents a new phase of genome analysis and deals with the reconstruction of the genome to determine the biological function of genes and the interactions between genes. The fundamental strategy in a functional genomics approach is to expand the scope of biological investigation from studying single genes or proteins to studying all genes or proteins at once in a systematic manner. Functional genomics is therefore characterized by high-throughput or large scale experimental methodologies combined with statistical and computational analysis of results.

5.3.2. Genome Sequencing Projects

There are several reasons for completely sequencing a genome.

- First it provides a means for the discovery of all the genes and thus provides an inventory of genes.
- Second, the sequence shows the relationships between genes.
- Third, it provides a set of tools for future experimentation.
- Fourth, sequencing provides an index to draw and organize all genetic information about the organism.
- Fifth, and very important over time, is that the whole genome sequence is an archive for the future containing all the genetic information required to make the organism.

There are several methods for small-scale sequencing, although most of these do not scale well to sequence entire genomes. The two main methodologies used for genome sequencing are discussed here. These have also been briefly discussed in the introduction.

Directed sequencing of Bacterial Artificial Chromosome (BAC) contigs

You have already learnt in the previous chapter that Bacterial Artificial Chromosome (BAC) vectors are capable of stably propagating large, complex DNA inserts in *Escherichia coli*. These vectors are used to make genomic libraries in which the insert size is 80-100 kb. This library is then screened by finding common restriction fragments. These BAC clones are then mapped to find overlapping arrays of contiguous clones called contigs. The mapped contigs are sequenced by breaking large DNA fragments into small pieces. Therefore, in this directed sequencing strategy, pieces of DNA from adjacent stretches of a chromosome are sequenced.

Random shotgun sequencing

Random shotgun sequencing is one approach to sequence genomic DNA. Genomic DNA macromolecules are very long and they contain many genes and other sequences required to build the whole organism. Even with the best of sequencing techniques we get a maximum of 700 bases of sequence information from one single run of an experiment. Therefore, we need a strategy to sequence the whole DNA. The random shotgun sequencing approach follows a very well known common theme "divide a big problem into small tasks. Solve these small tasks individually. Finally add up all these solutions to get the full final solution". Big genomic DNA molecules are broken down into small fragments, which are cloned in small (2.0 kb) and medium (10 kb) plasmid vectors. Plasmids have specific sites where these molecules can be inserted through enzymatic procedures. Thus, a library is constructed. Now each clone is picked up randomly and sequenced from both ends. By picking many clones and sequencing them, we get large amounts of sequences. Observations show that several of these sequences are identical,

some are similar to each other in parts called overlapping parts, whereas, a few may be just unique. After we feed all these data into a computer program, these sequences are joined by finding overlapping parts. The result is, we get long pieces of DNA sequences. This process of assembling continues until all overlapping parts are exhausted. Finally, we would get a large portion of the genomic DNA sequence.

Even though in theory, the entire genomic DNA sequence can be obtained in this way, in practise, this is not so. Some gaps in genomic DNA sequence do arise and these gaps need to be closed by specific cloning of those regions and additional sequencing.

5.3.3. Gene prediction and counting

Gene prediction is an important problem for computational biology and there are various algorithms that do gene prediction using known genes as a training data set. Since most of the knowledge to carry out these predictions comes from experimentally identified genes, this becomes a limitation. Even if we know where the genes are in the genome, it is not entirely clear how to count them. Due to the existence of overlapping genes and splice variants it is difficult to define the parts of the DNA that should be regarded as the same or several different genes. Nevertheless, for practical purposes (allowing for some 'experimental error') we can count the number of genes in an organism. Some of the results of counting predicted genes have turned to be quite surprising **(Table 1)**.

Table 1. Genome size and gene predictions between several organisms.

Organism	No. of chromo somes	Genome size in base pairs	The Number of Predicted genes	Part of the genome that encodes for protein
Bacteria *Escherichia coli*	1	500,000	5000	90%
Yeast *Saccharomyces cerevisiae*	16	12,068,000	6340	70%
Worm *Caenorhabditis elegans*	6	100,000,000	19,000	27%
Fly *Drosophila melanogaster*	4	175,000,000 - 196,000,000	13,600	20%
Weed *Arabidopsis thaliana*	5	157,000,000	25,498	20%
Human *Homo sapiens*	23	3,000,000,000	20,000 - 25, 000	< 5%

One of the surprises is the relatively small number of genes in a human genome (20,000 - 25,000 genes) in comparison to worm (19,000 genes). In fact some experts still think that there must be at least 40,000 - 50,000 genes in the human genome, and that 30,000 just reflects the

unreliability of in silico (i.e., computational) gene prediction. Still, it seems that there is no simple correlation between the intuitive complexity of an organism and the number of genes in its genome.

5.3.4. Genome Similarity, SNPs and Comparative Genomics

The human genome project has often raised two interesting questions - whose genome is being sequenced and how similar are genomes between two different individuals. It is understood that several anonymous samples were collected and pooled for human genome project. Since each person's genome is believed to be 99.8% identical to everyone else's, we can talk in terms of the consensus human genome. SNPs are DNA sequence variations, which occur when a single base (A, C, G, or T) is altered so that different individuals may have different bases at these positions **(Fig. 1).**

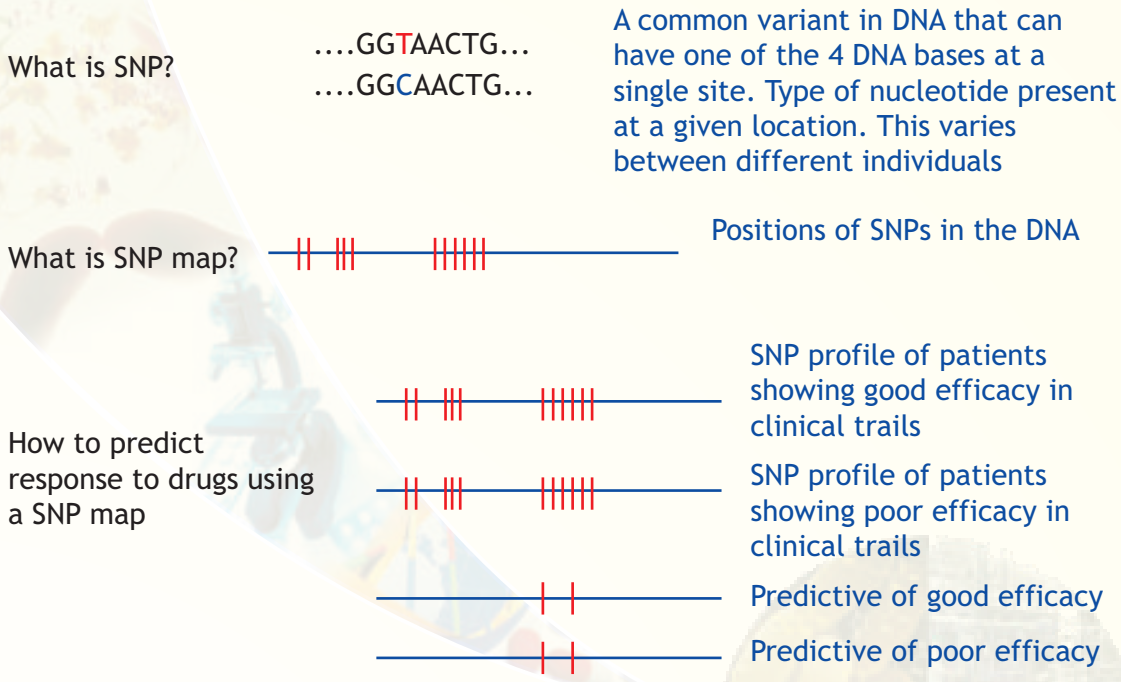


Fig. 1. Definition of SNP, and an illustration to show how physicians can use SNP map to determine how patients are likely to respond to a particular drug. The vertical bars indicate various SNPs on the human DNA.

However, the other way to look at it is that the 0.2% difference in DNA sequence is enough to make each individual unique. It is understood that on an average one in a thousand nucleotides are different in genomes of two different individuals. Particularly, important variations in individual genomes are the **single nucleotide polymorphisms** or SNPs, which can occur both in coding and non-coding regions of the genome. It is believed that SNPs occur at 1.6 million to 3.2 million sites in the human genome, and may affect gene function, depending upon exact base change and where it occurs. It would be interesting to note the following research based observation.

1. The genetic variations between individuals (particularly, in the non-coding parts of the genome) are exploited in DNA fingerprinting, which is used in forensic science.

2. However, not all genetic variations are beneficial **(see Table 2)**. Genomic variations underlie differences in our susceptibility to, or protection from all kinds of diseases. The severity of illness and the way our bodies respond to treatments are also manifestations of genetic variations. For example, a single base difference in the ApoE gene is associated with Alzheimer's disease, and that a simple deletion within the chemokine-receptor gene CCR5 leads to resistance to HIV (Human Immunodeficiency Virus) infections and the development of AIDS (Acquired Immunodeficiency Syndrome). SNP analysis is therefore important for diagnostics and a SNP database has been developed to aid these applications. An example of how a physician can decide if a medicine prescribed will be effective to a patient is illustrated in **Fig. 1** (vertical bars denote various SNPs in patients' genomes).

Table 2. Genes and diseases

Single-gene mutations which follow mendelian inheritance	**Gene polymorphisms which has complex inheritance**
Cystic Fibrosis (Cystic Fibrosis Transmembrane Conductance Regulator CFTR gene) 1. Inheritance: autosomal recessive disease	Common late-onset Alzheimer's disease 1. Inheritance: Major cause is epsilon4 allele of the gene coding for apolipoproteinE (APOE)
2. Genomic location: Chromosome 7 (7q31.2)	2. Genomic location: Chromosome 19 (19q13) and recently Chromosome 10 (10q21).
3. Mutation: The most common mutation is a deletion of 3 bps resulting in the loss of codon no. 508, which codes for phenylalanine	
Huntington disease (Huntingtin gene HTT) 1. Inheritance: autosomal dominant 2. Location: Chrosome 4 (4p16.3) 3. Mutation: increased number of CAG repeats more than 35 times	Migraine 1. Susceptibility locus: Chromosome 6p12.2 - 6p21.1 and Chromosome 1q31

3. Physicians can use patients DNA sample to determine the pattern of SNP genotype profile and from that they can predict how patients are likely to respond to a particular drug. SNP analysis can also be used in population genetics, as some SNPs vary in different frequencies between populations.

4. The genome sequencing projects have revealed that the genomes of organism otherwise quite different in appearance are quite similar for example mouse and man, are quite similar. Another example is that, among the conserved elements between different species such as the worm and the yeast, substantial portion belongs to genomic regions coding for proteins

5. It is estimated that the difference between human and chimpanzee genomes is only 1 to 3%, while human and mouse share about 97.5% of their working DNA. These similarities suggest that none of these genomes has changed much since we shared a common ancestor 100 million years ago.

5.3.5. Functional Genomics

Functional genomics dissects the emerging knowledge about genomes to understand the gene and their product functions and interactions. Two exciting new developments are now enabling scientists to get a wealth of clues to this complicated story. The new technique, microarray technology and proteomics provide snapshots of all the genes expressed in a cell or tissue under different environmental conditions. The DNA microarray technology is used for analysing the expression of thousands of messenger RNA molecules.

Fluorescence *in situ* hybridization

It is possible to introduce colours into DNA by a technique called Nick Translation developed in 1977 by Rigby and Paul Berg. The enzymes, DNA polymerase I makes DNA and DNase I, which cuts DNA are combined in a buffered reaction with dNTP's, including dUTP labelled with a red or green fluorescence. The DNA polymerase I adds nucleotide residues to the 3-prime hydroxyl terminus that is the result of nicks (breaks) created by the DNase I in the DNA. In the process, the fluorescence labelled nucleotide in the free nucleotide mixture becomes incorporated into the newly synthesized strands of DNA **(Fig. 2)** .

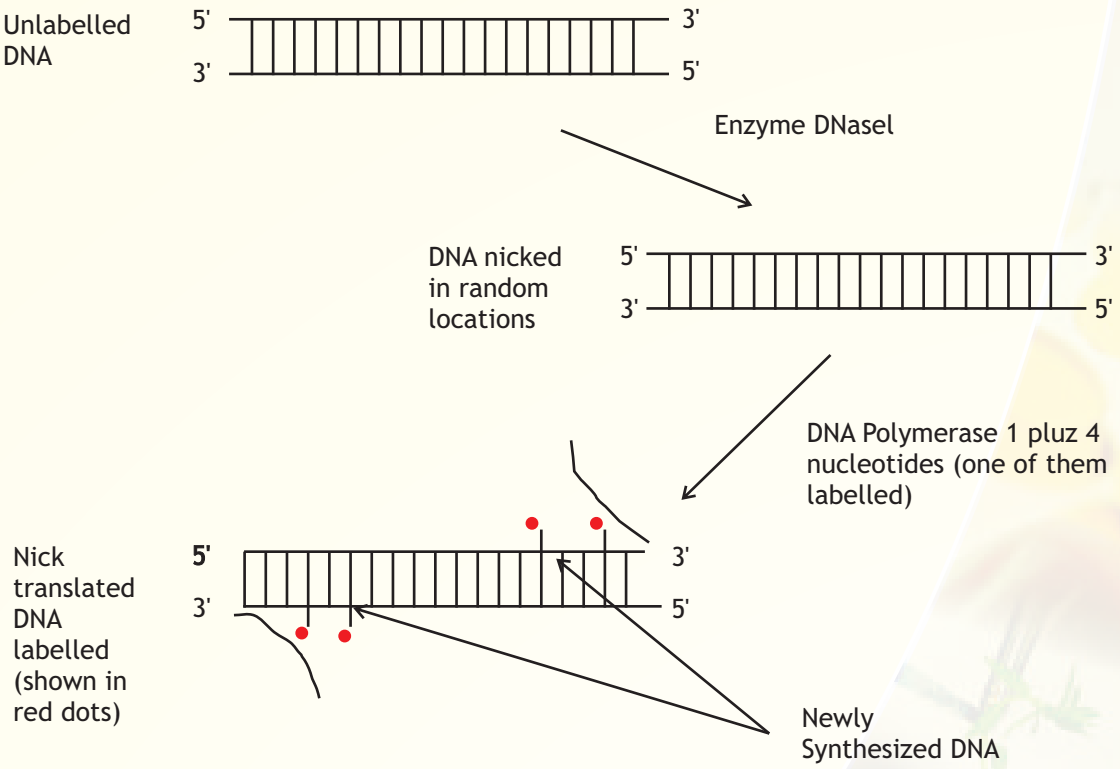


Fig. 2. Nick Translation. Nicks (breaks) are created in the DNA using DNase I. Subsequently, DNA Polymerase I synthesizes new DNA using the older one as template and incorporates the labelled nucleotides in the process. Finally we get labelled DNA.

The DNA fragment size with fluorescence probe after nick translation depends upon the amount of enzyme and the incubation time of reaction. The size range can be 300 to 3000 bp. The application of FISH can be illustrated by taking an example of chronic mylogenous leukemia (CML). It was observed from the karyotype analysis of the lymphocyte preparation made from blood samples of CML patients that there was a 9-22 translocation in the chromosome (also called 'Philadelphia chromosome'). Although by counting the number of such cells it was possible to find out the severity of the disease, it was not an easy procedure. The regions on the chromosomes involved in translocation were identified on chromosomes 9 and 22. From the DNA library it was possible to pick up clones carrying the particular genes involved in CML. Using nick translation it was possible to flourescently label chromosome 9 region with red colour and chromosome 22 region with green colour and prepare the probe **(Fig. 3)**.

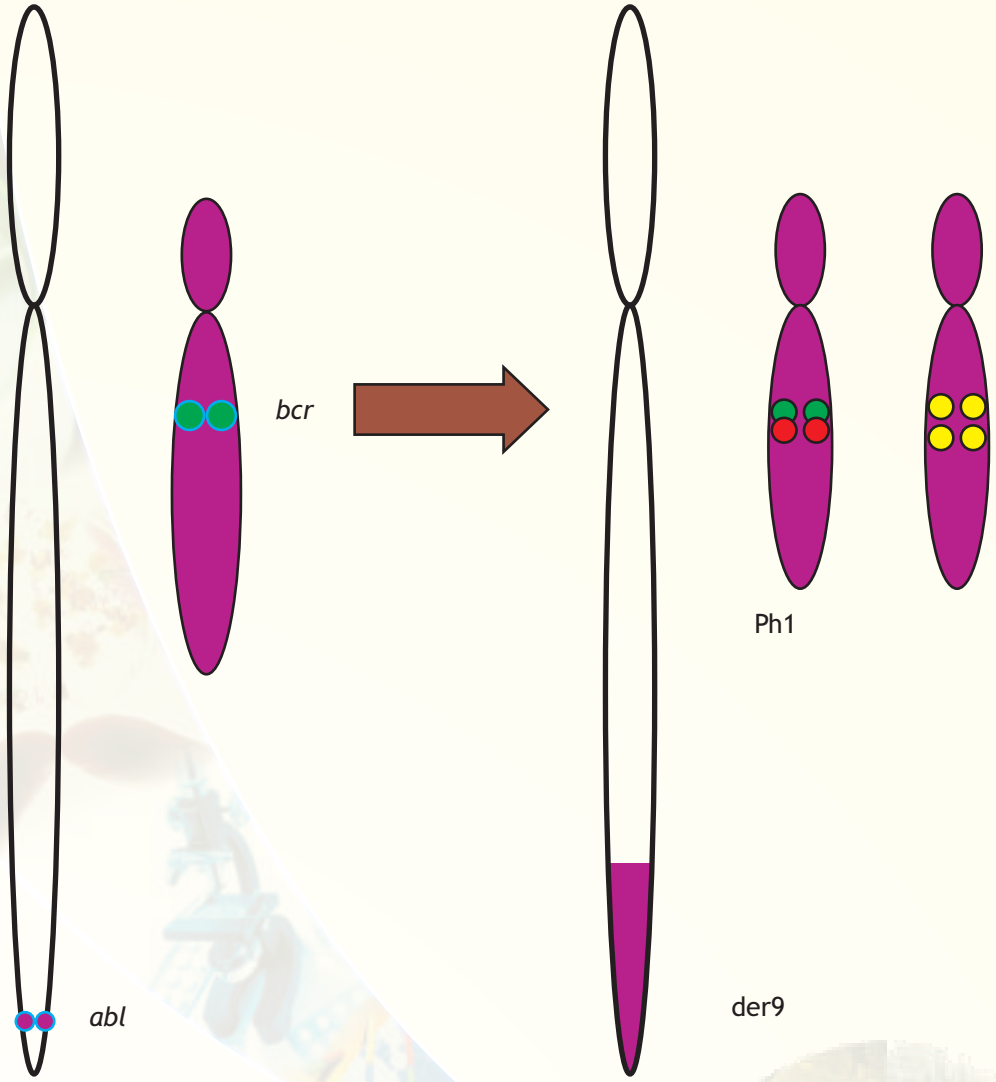


Fig. 3. Reciprocal translocation between Chromosome 9 and Chromosome 22 forms an extra-long chromosome 9 (der 9) and the Philadelphia chromosome (Ph1) containing the fused a-bcrg ene. This is a schematic view representing metaphase chromosomes.c

It was observed that when CML lymphocytes smear cells were hybridized with the two probes *in situ* and when observed under fluorescent microscope, the cells, which were affected, appeared yellow (mixing of green and red colour produces yellow colour). The unaffected cells appeared as red and green **(Fig. 3)**. This technique known as Fluorescence *in situ* Hybridization (FISH) allows knowing the status in the interphase unlike in karyotyping where you need a metaphase chromosome. The status of the disease could easily be identified by counting the number of cells, which appeared yellow. Further, it was possible to monitor the effect of chemotherapy and drugs by taking out samples and counting the number of cells appearing yellow.

Microarray Technology

It is widely believed that thousands of genes and their products (i.e., RNA and proteins) in a given organism function in a complicated and orchestrated way that creates the mystery of life. However, traditional methods in molecular biology generally work on a "one gene - one experiment" basis, which means that the throughput is very limited and the "whole picture" of gene function is hard to obtain. In the recent years, a new technology, called DNA microarray, has attracted tremendous interests among biologists. This technology promises to monitor the whole genome on a single chip so that researchers can have a better picture of the interactions among thousands of genes simultaneously.

Principle: Microarrays consist of large numbers of DNA molecules spotted in a systematic order on a solid substrate, usually a slide **(Fig. 4)**. The base pairing or hybridization is the underlying principle of DNA microarray. Microarray exploit the preferential binding of complementary single-stranded nucleic acids. A microarray is typically a glass (or some other material) slide, onto which DNA molecules are attached at fixed locations (spots). The type of molecule placed on the array units also varies according to circumstances. The most commonly used molecule is cDNA, or complementary DNA, which is derived from messenger RNA. Since cDNA are derived from a distinct messenger RNA, each feature represents an expressed gene. In order to detect cDNA bound to the microarray, they must be labeled with a reporter molecule that identifies their presence. This technique of introducing fluorescent dyes in DNA and its use in detection of target molecule by hybridization has been previously applied in fluorescent *in situ* hybridization (FISH).

Procedure: Comparative hybridization experiments compare the amounts of many different mRNA in two cell populations. If one wanted to compare a normal cell and a cancerous cell, the following experiments are needed to be carried out. mRNA is first purified from total cellular contents. mRNA accounts for only about 3% of all RNA in a cell so isolating it in sufficient quantity for an experiment (1-2 micrograms) can be a challenge. Since free RNA is quickly degraded, to prevent the experimental samples from being lost, they are reverse transcribed back into more stable DNA form. The products of this reaction are called complementary (cDNA) because their sequences are the complements of the original mRNA sequences. The reporters currently used in comparative hybridization to microarrays are fluorescent dyes (fluors), represented by the red and green circles attached to the cDNAs in **Fig. 4**. A differently-colored fluor is used for each sample so that we can tell the two samples apart on the array.

The labelled cDNA samples are called probes because they are used to probe the collection of spots on the array. The two cDNA probes are tested by hybridizing them to a DNA microarray. The array holds hundreds or thousands of spots, each of which contains a different DNA sequence. If a probe contains a cDNA whose sequence is complementary to the DNA on a given spot, that cDNA will hybridize to the spot, where it will be detectable by its fluorescence. In this way, every spot on an array is an independent assay for the presence of a different cDNA. Microarrays are made

from a collection of purified DNA molecules typically using an arraying machine. The choice of DNA to be used in the spots on a microarray determines which genes can be detected in a comparative hybridization assay. The end product of a comparative hybridization experiment is a scanned array image. A small piece of such an image is shown in **Fig. 4**.

Interpretation: The measured intensities from the two fluorescent reporters have been false-coloured red and green and overlaid (in reality, you observe colours when the fluors are stimulated by a laser). Spots, whose mRNA is present at a higher level in one or the other cell population show up as predominantly red or green. Yellow spots have roughly equal amounts of bound cDNA from each cell population. Therefore yellow spots correspond to genes expressed approximately equally in both normal and cancerous cells whereas red spots correspond to genes expressed in high amounts in normal cells. Similarly green spots correspond to genes expressed in high amounts in cancerous cells. By drawing this distinction, we would be able to understand the altered gene expression patterns in cancerous cells. This allows us to further understand the mechanism and make attempts to develop cures.

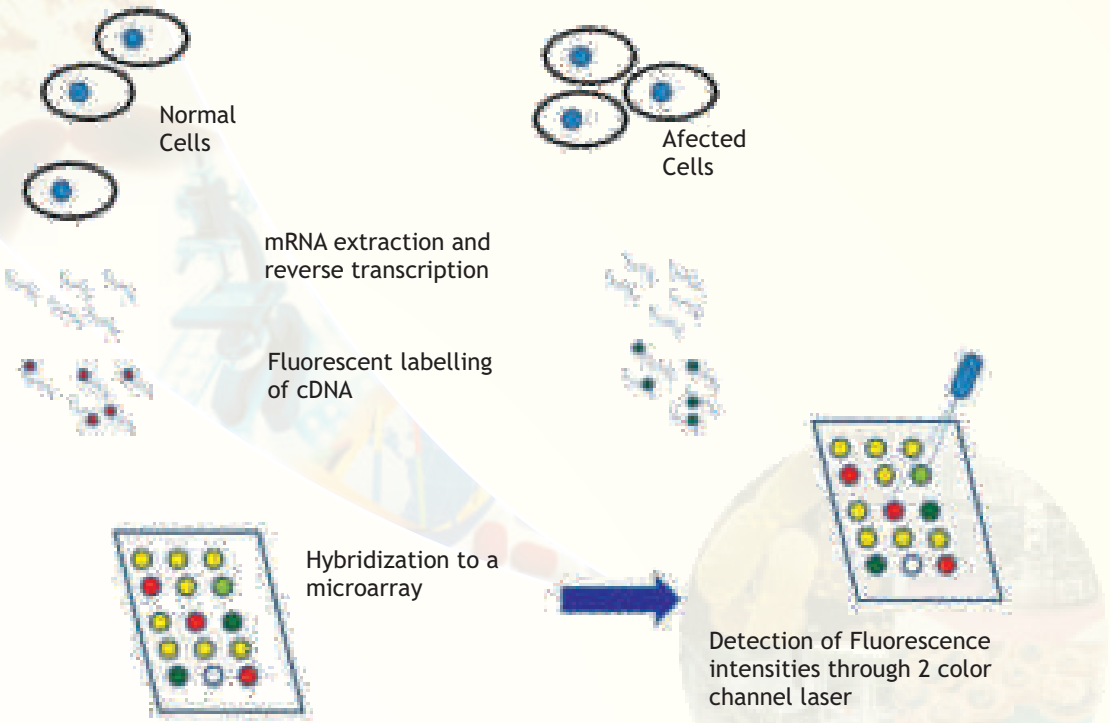


Fig. 4. Major steps involved in comparative microarray hybridization experiments between normal and affected (for example, cancerous) cells are illustrated. Observe that some spots are yellow, meaning that the particular gene is expressed in equal amounts. Some spots are clearly red or green indicating that the particular genes are expressed in only normal or affected condition. Some spots are more greenish or orange meaning that the expression status is not clearly tilted to either side but there is a trend towards either extremities. A few spots may appear blank (no colour).

This microarray technology promises to monitor the whole genome on a single chip so that researchers can have a better picture of the interactions among thousands of genes simultaneously. There are several names to this technology - DNA arrays, gene chips, biochips, DNA chips and gene arrays. In the case of gene chips, the substrate for immobilization is a silicon wafer and the probes are oligonucleotides spotted through photolithographic etching. In this case of gene chip only 1 colour hybridization is performed per chip. Comparisons are done by matching data from one chip to another through a special data normalization procedure. The principle used in this technology is being extended to develop protein arrays also. This technique has been used to study the following:

1. Tissue specific genes
2. Regulatory gene defects in a disease
3. Cellular responses to environment
4. Cell cycle variations

The strength of genomic studies lies in its global comparisons between biological systems. Genomics studies provide initial guidelines to identify areas for deeper investigation and to see how these results fit in the biological context.

5.3.6. Proteomics

The term Proteome refers to the complete protein set of a cell. Proteomics refers to the large scale characterization of the entire protein complement of cells, tissues and even whole organisms. Modern proteomic studies involve many different areas, which are illustrated in **Fig. 5**. These include protein-protein interaction studies, protein function, and protein localization. The growth of proteomics is a direct result of advances made in large-scale nucleotide sequencing of various genomes. Without this development, protein identification would have been difficult.

It is important to have information about the proteins simply because they are responsible for the phenotype of the cells. It is impossible to understand mechanisms of disease, ageing etc. solely by studying the genome. Only by understanding protein function and their modifications, drug-targets for various diseases can be identified. One of the major aims of proteomics is to create a three dimensional map of a cell indicating the location of proteins.

The proteome of a given cell is dynamic. In response to internal and external cues biochemical machinery of the cell could be modulated. This could lead to several changes in the proteins such as post-translational modifications, changes in cellular localization, effect on their synthesis or degradation. Thus examination of a proteome is like taking a snapshot of the protein environment at a given time.

It is speculated that no function can be assigned to about one-third of the gene sequences in organisms for which the genome sequence is known. The complete identification of proteins in a genome will help structural genomics projects. The aim of these projects is to obtain 3D structure of all proteins in a genome. The structural analysis would be helpful in assigning function to many of these proteins. In addition to identification of proteins, one of the major goals of proteomics is to characterize post-translational modifications on proteins.

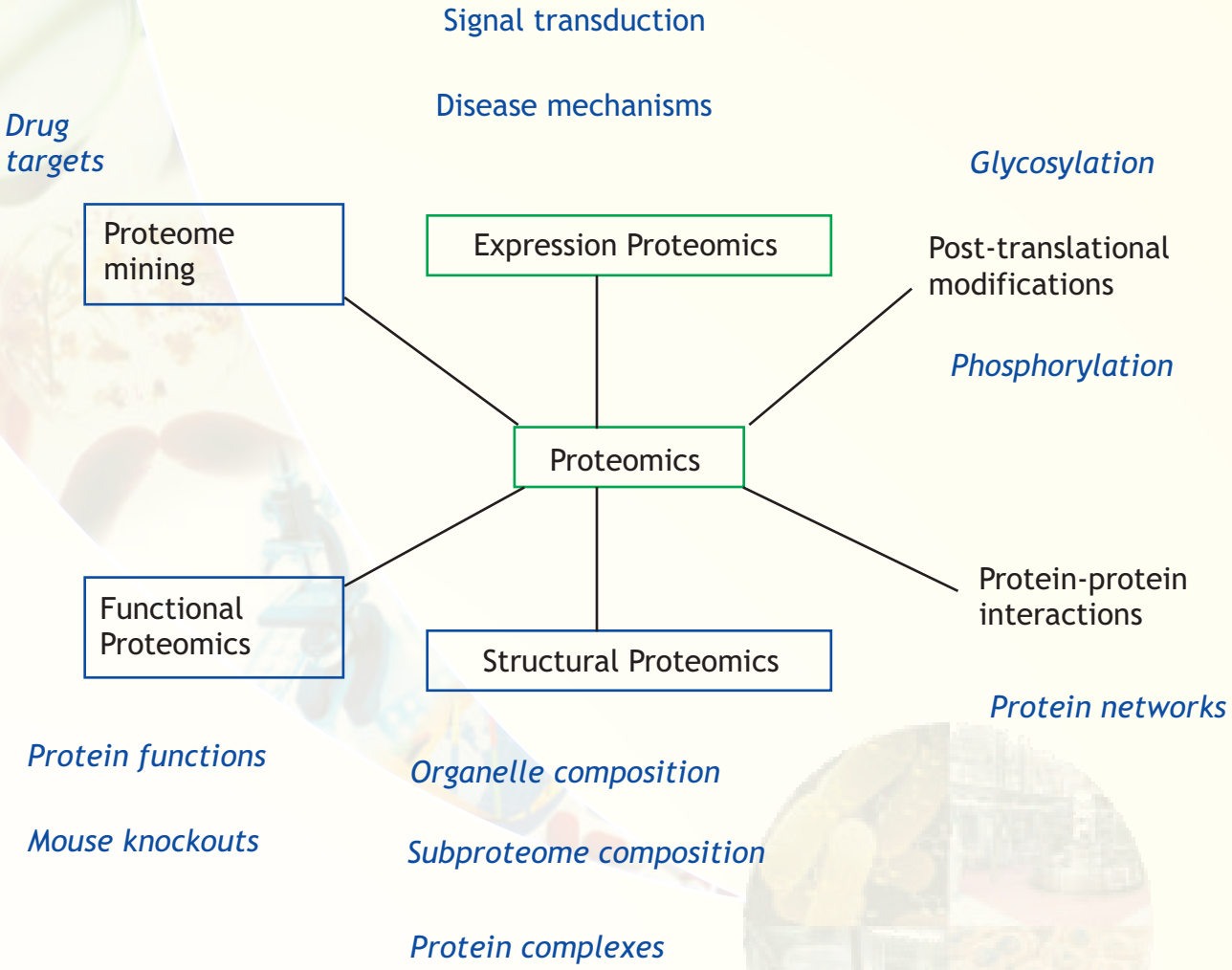


Fig. 5. Types of proteomics and the scientific knowledge that can be gained from them.

Types of Proteomics

Expression proteomics: The quantitative study of protein expression between samples that differ by some variable is known as expression proteomics. Using this approach, protein

expression of the entire proteome or of subproteomes between samples can be compared. This could be useful in identification of disease specific proteins. For example: tumor samples from a cancer patient and a similar tissue sample from a normal individual could be analyzed for differential protein expression. Using two dimensional gel electrophoresis, followed by mass spectrometry, proteins, which are over or under expressed in the cancer patient compared to the normal individual can be identified. This could be compared with the microarray data **(Fig. 4)**. Identification of these could provide a lead in understanding the basis of tumor development.

Structural proteomics: Unlike comparing the same cell or tissue in normal and diseased state in expression proteomics, structural proteomics are directed to map out the structure and nature of protein complexes present specifically in a particular cellular organelle. The aim is to identify all proteins present in a complex and to characterize all protein-protein interactions occurring between these proteins. Isolation of specific sub cellular organelles or protein complexes by purification can help assembling information about architecture of cells and explain how expression of certain proteins gives a cell its unique characteristics.

Functional proteomics: Functional proteomics is a very broad term for many specific, directed proteomics approaches. It can be defined as the use of proteomics methods to analyze the properties of molecular networks involved in a living cell. One of the major objectives is to identify molecules that participate in these networks. One of the successes of functional proteomics is identification and analysis of protein networks involved in the nuclear pore complex. This discovery has led to the identification of novel proteins which are important for translocating important molecules from the cytoplasm of a cell to the nucleus and vice versa.

Genes and Proteins

Number of genes vs Number of proteins

Analysis of mRNA does not provide a direct reflection of the protein content in the cell. One reason is that mRNA and protein expression levels do not always correlate. The formation of mRNA is only the first step in a long sequence of events resulting in protein synthesis **(Fig. 6)**. mRNA could undergo various post-transcriptional modifications like, polyadenylation and mRNA editing. Some of these modifications could lead to the generation of various protein forms from a single gene. Subsequently, translational regulation of mRNA could take place. Proteins, after synthesis, could undergo post-translational modifications. It is estimated that proteins could undergo as many as 200 different types of these modifications. Due to these reasons the relationship between number of genes and number of proteins is not linear. In other words, the number of proteins could easily outnumber the number of genes.

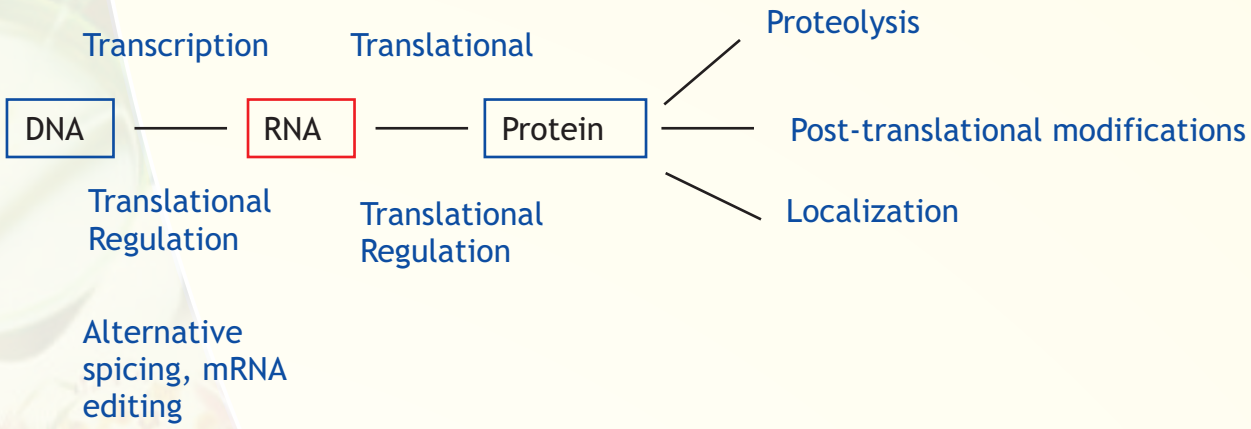


Fig. 6. Processes through which genes can give rise to multiple protein products with differing functions.

5.3.7. History of Bioinformatics

Bioinformatics has emerged as a scientific discipline that encompasses the application of computing science and technology to analyze and manage biological data. All this began when it was demonstrated by Ingram that there is homology between sickle cell haemoglobin and normal haemoglobin. This led to comparison of other proteins with similar biological function. As more and more proteins were sequenced, it became necessary to have databases which enabled a quick comparison using computational softwares. With the advent of rapid nucleic acid sequencing techniques, a large number of sequences started accumulating which again required computing facilities.

In 1962, Zuckerkandl and Pauling proposed a new approach of studying evolutionary relations using sequence variability. This initiated a new field called 'molecular evolution'. The approach was based on the observation that functionally related or homologous protein sequences were similar. Subsequently, sequence comparisons, analysis of functional relatedness and inference of evolutionary relationships became possible. Margaret Dayhoff observed that protein sequences undergo variation during evolution according to certain patterns. She noted that :

- amino acids were not replaced at random but were altered with specific preferences. For example, amino acids with similar physico-chemical characteristics were preferred, one for another.

- some amino acids such as tryptophan, was generally not replaced by any other amino acid.
- based on several homologous sequences, a point accepted mutation (PAM) matrix could be developed.

This laid the first foundation for subsequent work on sequence comparisons using quantitative approaches.

The National Biomedical Research Foundation (NBRF) compiled the first comprehensive collection of macromolecular sequences in the "Atlas of Protein Sequence and Structure' published from 1965-1978 under the editorship of Margaret O. Dayhoff. Dayhoff and her research group pioneered the development of computer methods for the comparison of protein sequences, for the detection of distantly related sequences and duplications within sequences, and for the inference of evolutionary histories from alignments of protein sequences.

In 1980, the data library was established at the European Molecular Biology Laboratory (EMBL) to collect, organize, and distribute nucleotide sequence, data and related information. Now its successor is the European Bioinformatics Institute (EBI) located at Hinxton, U.K. The National Centre for Biotechnology Information also started in USA as a primary information databank and provider at about the same time. Later, the DNA Data Bank of Japan was initiated. The Protein Information Resource (PIR) was established in 1984 by the National Biomedical Research Foundation (NBRF) as a resource to assist researchers in the identification and interpretation of protein sequence information. Today, all these databanks are in close collaboration with each other and they exchange data on a regular basis.

As the sequence data began to accumulate rapidly, new powerful sequence analysis softwares were needed. In parallel, firm mathematical basis was also required to develop algorithms. Scientists from the field of mathematics, biology, and computer science entered the emerging field of bioinformatics.

The databanks through their wide network of distribution of information are very important sources for all researchers who take interest in asking fundamental questions in biology. Thus, a major primary aim of bioinformatics is to spread scientifically investigated knowledge for the benefit of the research community. Other aims include the development of softwares for data analysis.

The word "bioinformatics" is a combination from biology and informatics. As it became clear that biological polymers, such as nucleic acid molecules and proteins, can be transformed into sequences of digital symbols informatics approaches can be used for analysis. Moreover, only

limited set of letters is required to represent the nucleotide and amino acid monomers. It is the digital nature of this data that differentiates genetic data from many other types of biological data, and has allowed bioinformatics to flourish. Another key point is that the use of sequence data relies upon an underlying reductionist approach: sequence implies structure which in turn implies function. In the subsequent sections we will see the details of these activities.

5.3.8. Sequences and nomenclature

The nomenclature system we adopt in Bioinformatics work is based on the International Union of Pure and Applied Chemistry (IUPAC) recommendations. It is useful to follow this nomenclature system so that data sets from different laboratories situated around the world can be compared easily and uniformly. The database institutions and the journals that publish research reports follow these recommendations strictly to ensure uniformity and to aid rapid reproducibility. We will go through the basic nomenclature system for nucleic acids and proteins in this section. Details of modifications of the nucleotides may be touched upon but we suggest you refer to the IUPAC website for these details. For routine work using the nucleic acid and protein sequence data we discuss the following system of IUPAC nomenclature **(Fig. 7)**.

Human communication language	Biological language
Letters	Nucleotide bases
Words	Genes, Exons, Introns
Sentence	Operons
Punctuation	turns, kinks, bending
Chapter	Chromosome

Fig. 7. Comparison between Human communication and biological system. The biological language is used in Bioinformatics.

DNA and protein sequences

Table 3. Summary of single-letter code IUPAC recommendations.

Symbol	Meaning	Base(s)
G	G	Guanine
A	A	Adenine
T	T	Thymine
C	C	Cytosine
R	G Or A	puRine
Y	T Or C	pYrimidine
M	A Or C	Amino
K	G Or T	Keto
S	G Or C	Strong (3 Hydrogen bonds)
W	A Or T	Weak (2 Hydrogen bonds)
H	A Or C Or T	not-G, H follows G in the alphabet
B	G Or T Or C	not-A, B follows A in the alphabet
V	G Or C Or A	not-T (not-U), V follows U in alphabet
D	G Or A Or T	not-C, D follows C in the alphabet
N	G Or A Or T Or C	Any

The symbols, their meaning and the bases for the nucleic acid sequences are presented in **Table 3.** The first 4 bases G,A,T,C, their symbols and the basis for nomenclature is clear. While determining sequence data through experiments, sometimes, the sequence identity at a particular position may not be clearly identifiable due to compression artifacts or other secondary structure related problems. In most cases the problem can be solved by repeating the experiment and also by sequencing the complementary strand. In a few cases, ambiguities may persist. In such cases, the most probable results are inferred from the chromatograms.

For instance, at a position where the ambiguity is not resolvable between a 'G' or a 'C' but one can be sure that there is no possibility of "A' or 'T' in the same position, then the symbol to be used is 'S'.

In most organisms, DNA is present as double stranded. The two strands are anti-parallel and complementary to each other (following Watson-Crick base-pairing). However, the problem

arises when we start encountering the symbols that mean more than one base at a given position. Again, the IUPAC system comes to aid. The symbols to be used in the complementary strand corresponding to the symbol at the same position in a given strand are specified in **Table 4**. In certain cases, the complementary symbols are same as in the given strand because in both cases they mean the same set of bases.

Table 4. Definition of complementary symbols.

Symbol	A	B	C	D	G	H	K	M	S	T	V	W	N	R	Y
Complement	T	V	G	H	C	D	M	K	S	A	B	W	N	Y	R

The symbols and their meaning for the protein sequences are presented in **Table 5.** It is evident that the number of symbols that mean more than one amino acid is very few.

Table 5. Symbol definitions for the amino acids.

Single letter code	Three letter code	Full name
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic acid
C	Cys	Cysteine
Q	Gln	Glutamine
E	Glu	Glutamic acid
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Asx
Z	Glx	Glx
X	Xaa	Xaa

The concept of directionality

In the biological systems, the usual direction in which the DNA and RNA are synthesized is the 5'-3' direction. This is universal and therefore it is helpful to adopt this fact as a way to collect and store data in the sequence databases. The nucleotide sequence are generally present in the database as they have been submitted or published, subject to some conventions which have been adopted for the database as a whole. The sequences are always listed in the direction 5' to 3'. Bases are numbered sequentially beginning with 1 at the 5' end of the sequence. The complementary sequence is described with a 'c' indicated next to the position of the sequence. Complementary sequence also runs 5'-3' but in the opposite direction to the given strand. Only one strand of the DNA sequence is given in a database entry. The complementary strand will have to be inferred using programs available in various packages or from various websites. We will see the details of a database entry in the next section. In the case of proteins, they are synthesized in the cell from N-terminus to the C-terminus. It is useful to adopt this convention in database entry for protein sequences. Thus, the concept of directionality used in biological systems is useful in describing the conventions to be adopted by the Database institutions. The advantage here is the universality of these fundamental biological processes in almost all living organisms.

Question : If you are given a sequence without any label, how will you find out whether it is a DNA sequence or a RNA sequence or a Protein sequence?

Answer : The usual approach taken by standard computer programs like sequence search programs scan the first 20 symbols. If the symbols encountered switch between any of the 4 bases only, then the sequence at hand is taken as a DNA sequence. Instead of T if U is encountered, then it is a RNA sequence. But if the symbols switch between any of the 20 (greater than 4), then it is taken as protein sequence.

Different types of sequences

cDNA : A large number of sequences deposited in the Databases were determined from cDNA molecules. While filling up the sequence entry form you must tick at the right position to indicate whether the sequence being deposited is a cDNA sequence. This data will also be provided when a sequence is retrieved. Thus in the case of cDNA sequences one is looking at the expressed part of the genome.

Genomic DNA: Sequencing of genomic DNA has become very routine nowadays. The genomic DNA is the store-house of information of which expressed part is represented in the cDNA sequences also.

ESTs : It is an abbreviation for Expressed Sequence Tags. Dr. Craig Venter initiated sequencing of a large number of cDNA molecules by sequencing one end of each of the randomly picked cDNA clones. Millions of ESTs have been deposited in a special database called dbEST. EST data is used

to infer expression patterns by counting the number of ESTs corresponding to each gene divided by the total number of ESTs.

GSTs : In *Plasmodium falciparum* the enzyme Mung Bean Nuclease (MNase) cleaves in between the genes. A genomic DNA library generated by digestion with MNase was used for gene identification in *P. falciparum*. The approach used was similar to ESTs. One read of sequence was obtained from either ends. This data is referred to as genome sequence tags (GSTs). Usually, genomic DNA sequence refers to the nuclear DNA.

Organelle DNA: Eukaryotic cells have organelles such as mitochondria and chloroplast. These organelles have their own store house of information in the form of organelle DNA. Organelle DNA codes for a few genes. The coding information for the rest of the genes reside in the nuclear DNA of the same cell. If an organelle DNA has been sequenced the appropriate position in the sequence submission form must be mentioned.

Other molecules: In addition to these molecules, the databases contain the sequences of other molecules such as tRNA, and other small RNAs.

5.3.9. Information Sources

Major databases

The National Center for Biotechnology Information (NCBI): The NCBI at the National Institutes of Health was created in 1988 to develop information systems in molecular biology. In addition to maintaining the GenBank nucleic acid sequence database, NCBI provides data retrieval systems and computational resources for the analysis of GenBank data and the variety of other biological data made available through NCBI. Since then, the tools and the services provided by the NCBI have grown so large that it would require a full chapter to describe about NCBI alone.

The resources available from the NCBI have been classified into the following heads: (1) Database retrieval tools, (2) BLAST family of sequence similarity search programs, (3) Gene level sequences, (4) Chromosomal sequences, (5) Genome analysis, (6) Analysis of gene expression patterns, (7) Molecular structure.

All these Web based tools are available free. We shall learn about the practical aspects of some of these tools in practical classes. Below we discuss the first three set of resources. Learning these can help you in most cases. A great majority of bioinformatics activity is carried out using these resources. Other resources are useful for advanced studies.

Database retrieval tools

Among the database retrieval tools are ENTREZ, TAXONOMY BROWSER, LOCUS LINK. Entrez is an integrated database retrieval system. Through this system one can access literature (in the form

of abstracts), sequences and structures. Entrez is an excellent system for obtaining comprehensive information on a given biological question. The taxonomy browser provides information on taxonomic classification of various species.

The taxonomy database has information on over 79, 000 organisms. Locus link carries information on the official gene names and other descriptive information about genes. Additionally, through Locus link one can access information on homologous genes. For example, it is very convenient to obtain information on the mouse homologue of a given human gene. Homologues from other organisms are also available.

BLAST family of search tools

Among the BLAST (Basic Local Alignment Search Tool) family of similarity search programs are several tools to analyze sequence information. These tools are designed to answer the question "Which sequences in the database are similar (or homologous) to my sequence?" The theory on which BLAST systems were developed is somewhat complex and is out of the scope of discussion here. The principles involved are-

(a) A given sequence is compared with sequences in the database using substitution matrices that specify scores to either 'reward' a match or 'penalize' a mismatch.

(b) Top scoring matches are ranked according to set criteria that serve to distinguish between a similarity due to ancestral relationship or due to random chance. In most analysis these criteria are not changed. However, if the user wishes, criteria can be changed.

(c) True matches are further examined thoroughly with other details accessible through Entrez and other tools available at NCBI.

Note: Two sequences being similar does not mean that they are homologous. Homology is defined as similarity due to common ancestry. Two sequences each from species A and species B are said to be homologous if they have descended from a common ancestor to species A and species B. Duplicated genes within a genome also may have similarities but these are referred to as '**paralogs**'. Homologues will have the same function whereas paralogs may differ in functions.

Resources for gene level sequences

Among the resources for gene level sequences are several tools such as the UniGene, HomoloGene, RefSeq and others. We mentioned about the ESTs in the previous section. The method described therein produces many redundant ESTs because several cDNA clones represent the same gene. To manage the redundancy in EST data, UniGene database was created. The objective is to group ESTs into sets called clusters that belong to 'one' gene (Uni meaning one).

Homologene is a database of orthologs and homologs for several organisms like human, mouse, rat, zebrafish and cow genes represented in UniGene and Locus Link. It is easy to infer homologous relations using this database. RefSeq is a curated database of mRNAs and proteins of organisms like human, mouse and rat. The data provided in RefSeq has been used in many cases such as designing gene chips and describing the sequence features of the human genome.

There are many other resources provided by the NCBI. Discussing all of these is not possible within the space limitations. Now, we mention a few other databases of importance to bioinformatics work (see **Table 6** below)

Table 6. Examples of other useful databases for Bioinformatics.

Database	Information available
EMBL(European Molecular Biology Laboratory)	Nucleotide sequence
UniProtKB	Annotated protein sequence
PDB (Protein Database)	Three dimensional structure of proteins
Ribosomal RNA database	rRNA subunit sequences
PALI database	Phylogenetic analysis and alignment of proteins

Curator: A curator is one who reviews and checks newly submitted data ensuring all mandatory information has been provided, that biological features are adequately described and that the conceptual translations of any coding regions obey known translation rules. This process is called curation.

Analysis using Bioinformatics tools

Many kinds of analysis can be made using various bioinformatics tools. These include:

Processing raw information: The experimentally determined sequence (raw information) is processed using bioinformatics tools into genes, the proteins encoded and their function, the regulatory sequences, and inferring phylogenetic relationships.

Genes: Gene prediction can be done by using computer programs like GeneMark for bacterial genomes and GENSCAN for eukaryotes.

Proteins: Protein sequences can be inferred from the predicted genes by using simple computer programs.

Regulatory sequences: Regulatory sequences can also be identified and analysed by using bioinformatics tools.

Inferring phylogenetic relationships: Information regarding the relationships between organisms can be obtained by aligning multiple sequences, calculating evolutionary distance and constructing phylogenetic trees.

Making a Discovery: Using the bioinformatics tools and databases, the functions of unknown genes can be predicted.

Review Questions

1. Why was it necessary to create Bioinformatics database?
2. List the important databases used in routine bioinformatics.
3. What are the IUPAC codes for (i) 'G' or 'C', (ii) A or T, (iii) A or C, (iv) C or T, (v) A or G
4. What are the conventions adopted by the Database personnel to store nucleic acid data and protein sequence data with regard to the direction of the sequence? What is the basis of the convention?
5. What are the single letter IUPAC codes for alanine, glycine, tryptophan, tyrosine, serine, methionine?
6. What are the different types of molecules on which sequence data is obtained and deposited in the database?
7. Name some of the database retrieval tools. What is their purpose?
8. Suggest one possible way for going about analyzing a given sequence using bioinformatics.
9. Using microarrays one can identify the genes expressed differently in normal vs cancer cell types. Explain.
10. What is random shotgun sequencing? What are the difficulties with assembling sequences with repeats?
11. What were the surprises revealed from genome sequencing? What underlies the accrual of complexity in humans even though the number of genes are low?
12. How genes are linked to diseases? Explain with 2 examples.
13. What is BLAST? Describe the principles that underlie BLAST search.
14. What is proteomics? How we can benefit from proteomics?

References

1. Discovering Genomics, Proteomics & Bioinformatics (2002). A Malcolm Compbell and Laurie J. Heyer, published by Benjamin Cummings.

2. Proteomics-from protein sequence to function (2002). S.R. Pennington and M.J. Dunn BIOS Scientific Publishers Limited.

3. Bioinformatics A practical Guide to the Analysis of Genes and Proteins (2002). A.D. Baxeuarus and B.F Francis Ouellehe. A John Wiley & Sons, Inc.

UNIT 6

CELL CULTURE AND GENETIC MANIPULATION

CHAPTER 1

MICROBIAL CELL CULTURE AND ITS APPLICATIONS

6.1.1. Introduction

In your previous class you have studied that microorganisms like bacteria and fungi are cultured to obtain a number of products, which are beneficial for human beings. Some of the commonest examples are antibiotics, ethanol and enzymes. With the advent of biotechnology and genetic engineering techniques, microorganisms are also being used for the production of recombinant molecules such as insulin, hepatitis B vaccine, growth hormones and interferons. Microorganisms are also grown or cultured for production of biomass per se such as the single cell protein (SCP). Besides these, scientists culture microbes in their laboratories for research proposes. It is therefore very important to understand how microbes are cultured in the laboratory or in the industry for production of various important molecules.

A microbial culture works as a factory in which the metabolism of a microorganism is exploited to convert raw material into products. There is a maximal capacity for a single microbial cell to convert raw material into products in a given period of time. Hence, to produce a given amount of product within a given time, it is necessary to calculate the number of cells as well as doubling time of the particular bacterium used in the process. Like any other chemical reaction, which requires an appropriate temperature, pressure, pH and solvent, microbes also grow in an appropriate environment of pH, temperature, nutrients (provided by the growth medium) and the substrate (raw material), which is converted by the bacterium into the desired product.

This chapter deals with basic principles of culturing microbes in the laboratory, large-scale industrial productions, and their applications. An important point that should be kept in mind is that, the requirements for culturing microbes on a large or industrial scale are different from culturing them in a laboratory.

6.1.2. Microbial culture techniques

Nutrients for microbial culture:

All growing microorganisms require water, sources of energy, carbon, nitrogen, oxygen and mineral elements. For culturing microbes on a small scale such as in a laboratory, it is relatively simpler to devise nutritive medium using pure chemicals such that full composition of the medium is known. Such media are called **synthetic media**. Alternatively, we may use media commonly available in the market such as nutrient broth, typticase soya broth (TSB) or brain heart infusion (BHI) broth. These media contain highly complex components such as peptone,

beef extract, yeast extract or casein digest. Such media are called **semi-synthetic media**. These media are well suited for culturing most microbes in the laboratory. However, when one wants to culture microbes on a large scale for production of useful metabolites, one uses sources of nutrients, which are economical and available readily. In some cases, these nutrient sources or raw materials may need a pretreatment before use. Other considerations while selecting the growth medium are:

- Should yield maximum product or biomass per gram of the substrate used.
- Should be of consistent quality and available throughout the year.
- Cause minimum problems during preparation and sterilization.
- Give minimum problems during production process particularly aeration, agitation, extraction and purification of the product.

Carbon sources: Cereal grains, starch, cane molasses, glucose, sucrose and lactose are commonly used as carbon sources.

Nitrogen sources: Ammonium salts, urea, corn steep liquor or slaughterhouse waste are used as nitrogen sources.

The carbon and nitrogen sources mentioned above are cheap and easily available.

Trace elements: Trace elements like Fe, Cu, Zn, Mn, Mo are also needed in small quantities.

Growth factors: Some microorganisms cannot synthesize specific nutrients like some amino acids or some vitamins. Such growth factors are incorporated into the medium in adequate amounts as a pure compound or as a component of complex mixture.

Antifoaming agents: In most microbiological processes, foaming is a problem. Foaming may be due to the components of the culture medium or some molecules produced by the microbes. The most common cause of foaming is the presence of proteins in the culture medium. Excess foaming denatures proteins and provides hindrance to free diffusion of oxygen in the medium. Commonly used antifoams are fatty acids, such as olive oil or sunflower oil. Silicones are also used as antifoams.

Energy sources: The carbon sources such as carbohydrates, lipids and proteins in the culture medium are the major sources, which provide energy for the growth of microbes.

Water: Water is the major component of the culture media. When required for culturing microbes in the laboratory, single distilled or double distilled water would suffice. When assessing its suitability for large-scale microbial culture as in industry, we should consider the pH and dissolved salts. Water is also required for ancillary services such as heating, cooling and rinsing. Clean water of consistent composition is therefore a prerequisite for large-scale cultivation of microbes.

Culture Procedures

Sterilization procedures: When microbes are cultured in laboratory in 100 -1000 ml flasks, sterilization of the nutrient medium can be easily done in an autoclave at a pressure of 15 pounds psi (per square inch) for 15-20 minutes. However, when microbes are cultured on a large scale such as for fermentation processes in the industry, we need to sterilize thousands of liters of culture medium or the substrate. There is also the problem of sterilizing the huge culture vessel i.e. the fermentor. Steam is used almost universally for sterilization of the fermentation media. Medium may be sterilized *in situ* in the fermentor itself. But if the medium is sterilized in a separate vessel, the fermentor must be sterilized before sterile medium is added to it. The fermentor is sterilized by passing steam through its jacket or coils. The steam is also sparged into the vessel through all entries allowing it to exit slowly from the air outlet. Steam pressure is held at 15 lbs psi in the vessel for 20 minutes.

The air used in the fermentation process should also be sterilized. This is done by filter sterilization.

Aeration and mixing: When microbes are cultivated in the laboratory, aeration and mixing can be easily achieved by putting the flasks on shakers (shake culture). This may be further augmented by the use of baffle flasks **(Fig. 1)**. In large-scale bioreactors however, transfer of oxygen to microorganisms is particularly difficult because the microorganisms must be well mixed and the oxygen dispersed to achieve relatively uniform concentration. Many fermentor designs have mechanical stirrers to mix the medium, baffles to increase turbulence, which ensure adequate mixing. Forced aeration also provides mixing and the needed oxygen.

Equipment for microbial culture

In the laboratory, microbes can be cultured in test tubes or ordinary Erlenmeyer flasks. Such cultures are usually carried out in 100 -1000 ml volumes. Growth of microbes in the laboratory can be augmented by simple improvements in the design of the flasks or by the use of shakers.

Baffle flask: One of the simplest ways is to produce a V- shaped notch or indentation in the sides of the flask. Such flasks are called baffle flasks **(Fig. 1)**. This improves the growth of the microbes by improving the efficiency of oxygen transfer due to increased turbulence of the agitated culture medium.

Fig. 1. A Baffle flask.

Shakers: Continuous agitation of the culture medium also greatly improves the efficiency of the oxygen transfer and this improves the growth of the microbes. In the laboratory, this is done by the use of shakers **(Fig. 2)**. Shakers may be end-to-end type or rotatory type. These may be designed for use at the ambient temperature or in a controlled temperature environment (incubator shaker).

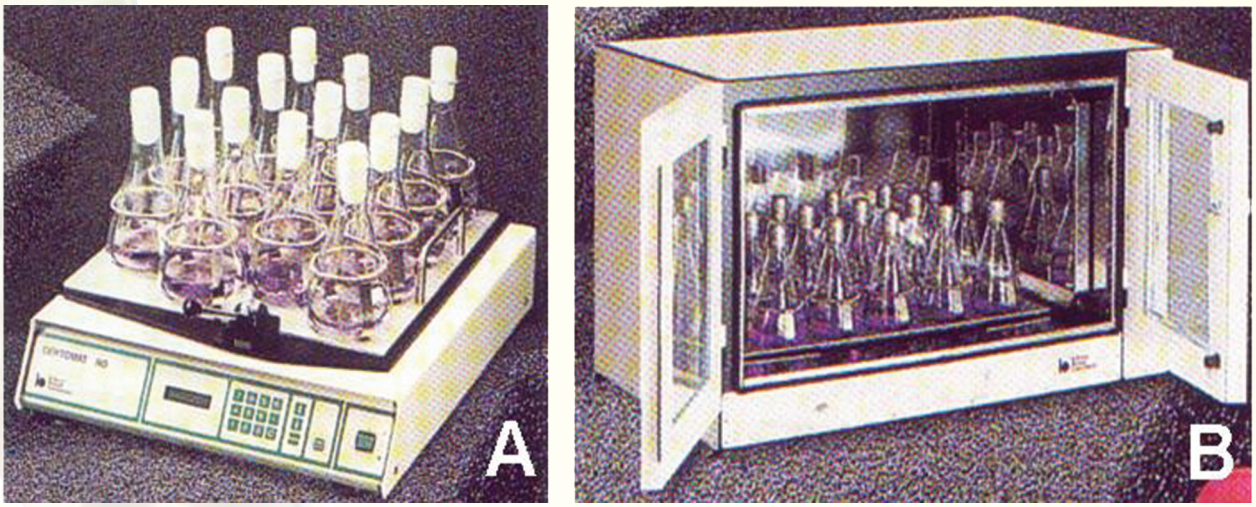


Fig. 2. (A) Shaker; (B) Incubator Shaker.

Fermentors (bioreactors): These are vessels, which are used for large-scale growth of microorganisms under a controlled environment. You have read about the basic design of a fermentor in your previous class in Unit I. It is a closed vessel with adequate arrangement for aeration, agitation, temperature and pH control, and drain or overflow vent to remove the cultured microbes along with their products. The vessel is capable of being operated aseptically for a number of days. These are also equipped with sampling ports, which allow withdrawal of the fermentation broth at regular intervals while the fermentation is in progress.

There are four main types of fermentor or bioreactor designs. The most common of these is a stirred tank bioreactor in which the culture medium is stirred with an impeller. In the bubble column reactor, the air is forced through a bottom sparger that creates enough agitation to ensure proper aeration. Basic features of a fermentor are shown in **Fig. 3.**

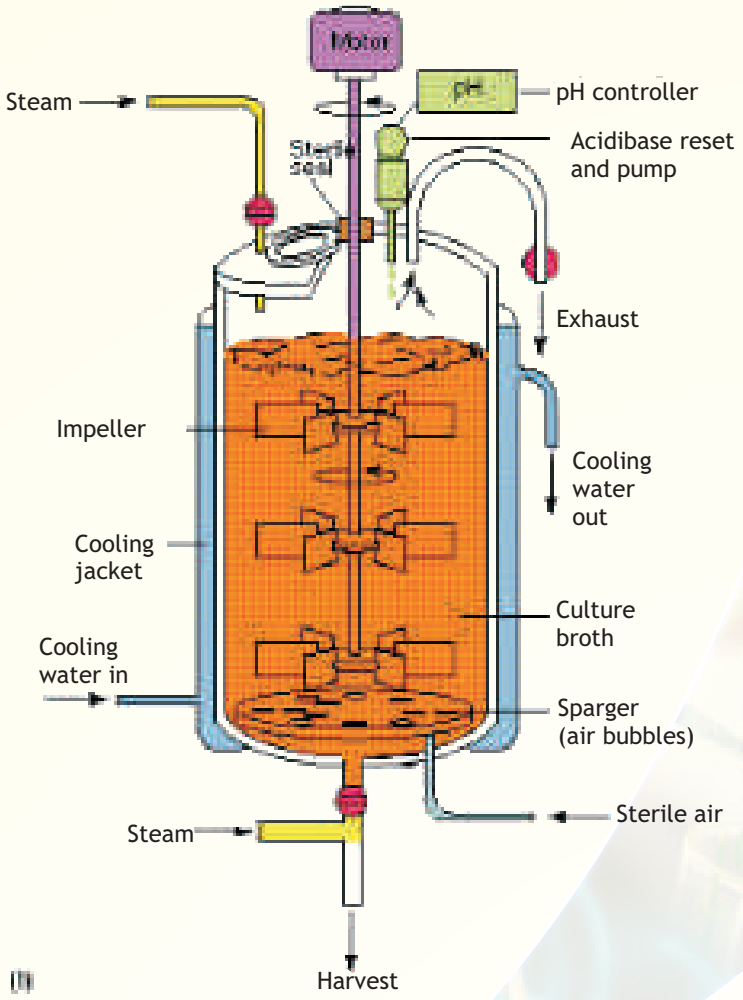


Fig. 3. Basic features of a fermentor (Bioreactor).

It should be noted that a high concentration of the microbial cells, as is achieved in a fermentor, could rapidly deplete the soluble oxygen in the medium, creating anaerobic conditions that may not be favorable to the growth of microorganisms and/or production of the desired products. So, forced aeration is done. The rapid growth of the microorganisms in the fermentor quickly alters the pH of the medium. So pH of the growing culture is continuously monitored and acid or alkali is added as required.

However, you should not harbor any notion that fermentors are used only for large-scale growth of microbes on an industrial scale **(Fig. 4)**. Small scale fermentors of capacity 10 -100 liters are also used in research laboratories These are used by the scientists in research, to optimize various parameters for the growth of microbes. The laboratory scale fermentors are also used by scientists, to produce enough quantities of metabolites from microbes for research purposes **(Fig. 4)**.

Fig. 4. (A) Laboratory-scale fermentor; (B) Industrial-scale fermentor.

Types of microbial culture

Batch culture: Batch culture is a closed culture system, which contains an initial limited amount of nutrients. After the medium is inoculated with the bacterial inoculum, the organism will grow and show usual growth phases viz. lag, log (exponential), stationary and decline. You have read about these in your class XI. Growth results in the consumption of nutrients and excretion of microbial products. At stationary phase, the growth is zero. This means, that in such a culture, growing cells are exposed to continually changing environment due to gradual consumption of nutrients and the accumulation of metabolites. The cell density [X], the concentration of substrate [S] and cell-specific substrate turnover rate [QS] during such a system are shown in **Fig.5**. Culturing microbes in the laboratory, in an ordinary flask, is nothing but an example of batch culture.

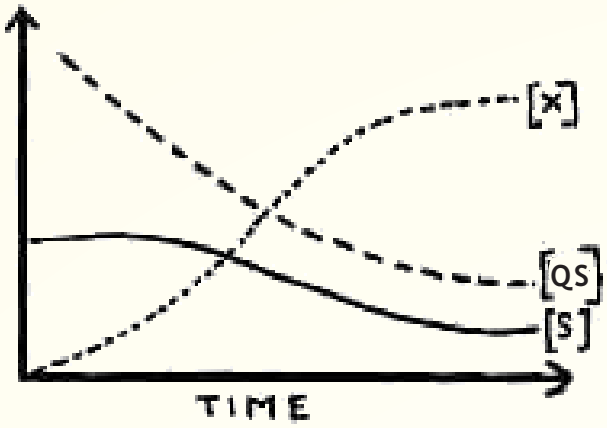


Fig. 5. Characteristic features of a batch culture.

Fed-batch culture: If a batch culture described above is continuously or sequentially fed with fresh medium without removing the growing culture, it is called fed-batch culture. In this system, substrate concentration remains constant and the cell density keeps increasing over time **(Fig. 6)**. This also means that over the period of time, the volume in the culture vessel goes on increasing.

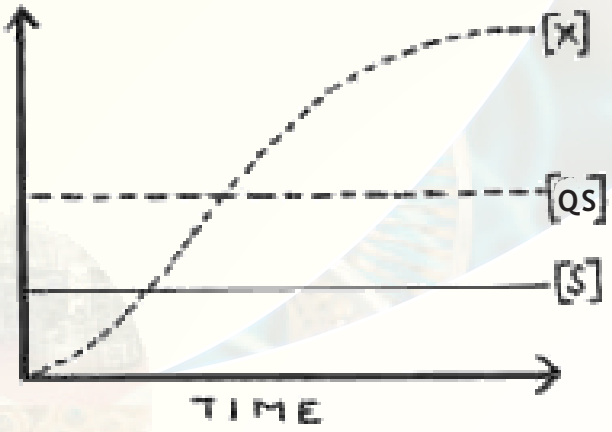


Fig. 6. Characteristic features of a fed-batch culture.

Continuous culture: This is a very interesting way of getting a continuous supply of microbial growth and/or products. The growth medium is designed in such a way that one of the nutrients is in limited quantity. Thus, during the exponential growth, as this nutrient is exhausted the growth will stop. However, just before the nutrient is fully exhausted, fresh medium containing the limited nutrient is added. This is repeated every time the limited nutrient is about to exhaust. This system is also fitted with an overflow device. This means that the added volume displaces out an

equal volume of culture from culture vessel. That is, formation of new biomass by the culture is balanced by the loss of culture from the vessel. In continuous culture, cells can be grown at a particular growth rate for an extended period of time. Most of the time, the chemical environment inside a continuous culture is constant. In a **chemostat**, constant chemical environment is maintained whereas in a **turbidostat** constant cell concentration is maintained. If medium is fed to such a culture at a suitable rate, a **steady state** is achieved eventually. At steady state, the cell growth and substrate consumption takes place at a fixed rate **(Fig. 7)**. Growth rate of cells remains constant during steady state operation. This means that at steady state, the concentration of cells, metabolites and other nutrients inside the reactor remain constant. That is, formation of new biomass by the culture is balanced by the loss of the culture from the vessel. Continuous culture is most suitable for production of biomass or metabolites. Continuous culture is widely used for production of both microbial biomass as well as their metabolites.

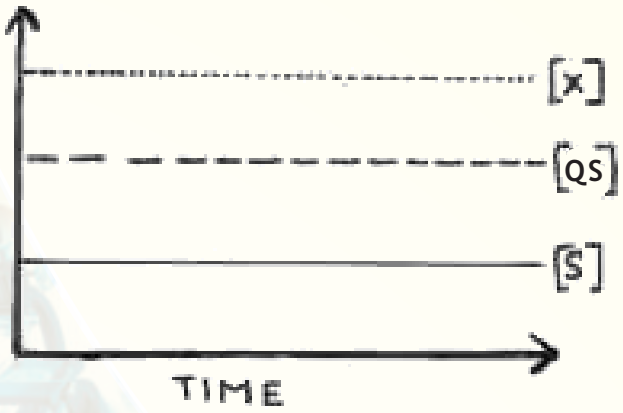


Fig. 7. Characteristic features of a continuous culture at steady state.

6.1.3. Measurement and kinetics of microbial growth

A clear understanding of microbial growth is necessary for proper utilization of biological processes for production of metabolites. In this section, the methods used for the measurement and quantitative evaluation of microbial growth will be discussed. There are four general patterns of microbial growth exemplified by bacteria, yeast, mold and viruses. All these microorganisms grow in different ways: bacteria grows by binary fission, yeast divide by budding, fungi divide by chain elongation and branching whereas viruses normally do not follow a regular growth pattern as they grow intracellularly in host cells.

Measurement of microbial growth

Microbial growth is defined as an orderly increase in all chemical components in the presence of suitable medium and environment. During the period of balanced growth, doubling of biomass is accompanied by doubling of other measurable properties of the microbe such as protein, DNA, RNA or intracellular water. In general, measurement of cell mass or cell number is used for

quantitative measurement of cell growth. The parameter that characterises microbial growth is the doubling time. It is the time required for the cell mass or number to double its original value during the balanced growth (i.e., log phase) of the organism.

Measurement of cell mass or number is one of the easiest ways to measure microbial growth.

(1) It is carried out by measuring the dry weight of the cell material in a fixed volume of the culture by removing the cells from the medium, and drying them till constant weight is obtained. For *E. coli*, the bacterium that is extensively used in laboratories, the dry cell weight of 1 billion cells is approximately 150 mg.

(2) Cell growth is also measured by measuring the absorbance of cell suspensions in a spectrophotometer. This principle is based on the fact that small molecules scatter light proportionate to their concentration. When light passes through a suspension of bacteria, there is a reduction in light transmitted as a consequence of scattering. Thus, with different cell concentrations, the absorbance at a particular wavelength will be proportional to the cell concentration. If you have a standard graph, plotted with absorbance versus cell concentration, the cell concentration of the unknown microbial sample can be calculated by measuring the absorbance at the same wavelength.

Apart from the methods described above, other methods are also used for measuring cell growth. Some of these are: measurement of wet weight of cells, turbidity measurements, ATP measurement, viable plate count **(colony forming units or cfu)** and use of Coulter counter. Coulter counter is an electronic instrument and is used for direct counting of microbial cells in suspension. In Coulter counter, the cells are made to pass in a single file through electrical field and the electrical impedance thus generated is recorded.

Growth kinetics and specific growth rate

In order to understand microbial growth kinetics, we shall take bacterial **binary fission** as the example where each cell division produces two identical daughter cells. The time taken for the cell to divide is called **generation time**. The generation time is also known as **doubling time (t_d)**, because the population of cells doubles over this time. For example, one cell of *E.coli* put in a nutritive medium will divide every 20 minutes. After one hour of growth (i.e. after three generations), one cell will have become eight (1 to 2, 2 to 4, 4 to 8). Because cell number doubles with each division, the increase in cell number over time is **exponential** or **logarithmic**.

A typical bacterial growth curve is depicted in **Fig. 8,** about which you have already learnt in Unit I (Chapter 1) of the textbook for class XI. In the lag phase, the microbial population remains almost constant. It is however a period of intense metabolic activity as the bacteria adapt to the culture conditions. Once the cells have adapted to the culture, cell division occurs with increasing frequency until the maximum growth is reached. This is called the log phase. At this

point **exponential growth** occurs and cell biomass or cell number increases at a constant rate.

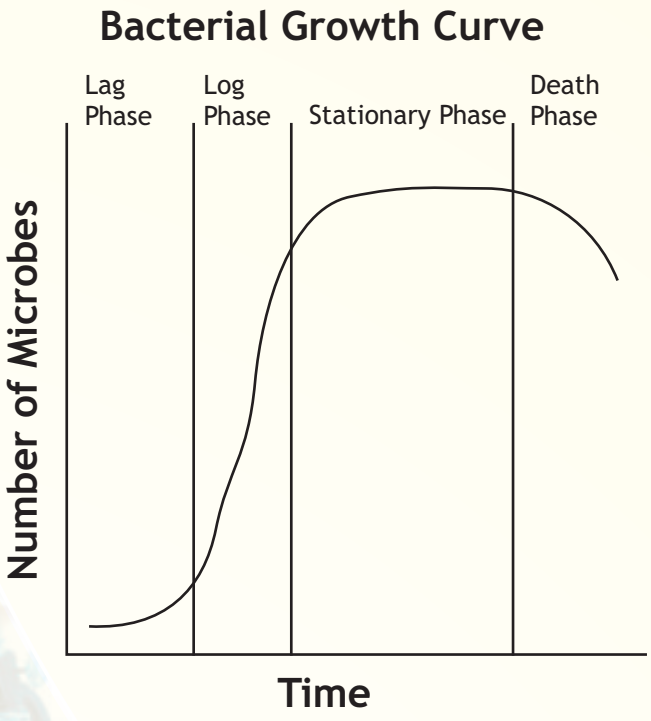


Fig. 8. A typical bacterial growth curve.

Mathematically, exponential growth can be described by two methods:

One related by biomass (X) and the other by cell numbers (N).

Let us first consider the exponential growth as related to biomass.

Balanced growth is like a chemical reaction where substrate is getting converted into product i.e. cell biomass in this case. Thus, growth behaves like an auto-catalytic reaction and the rate of growth will be proportional to the cell biomass present at that time. This can be described as follows:

rate of change of biomass $dX / dt = \mu X$ Eq.------- 1

where

X : concentration of biomass (g/L)

t : time (h)

μ : specific growth rate ($hour^{-1}$)

The equation 1 can be rearranged to estimate the specific growth rate (μ) as:

The specific growth rate $\mu = 1 / X \cdot dX / dt$ Eq.------- 2

Specific growth rate is an index of rate of growth of the cells in particular environment. In other words, specific growth rate is characteristic of the microorganism and is a function of the growth environment including temperature, pH, medium composition and levels of dissolved oxygen.

During the period of true exponential growth, equation 1 can be integrated to provide the following equation:

$$X_t = X_0 e^{\mu t} \qquad \text{Eq. ------- 3}$$

where

X_t: biomass conc. after time t

X_0: biomass conc. at the start of the exponential growth

e : base of the natural logarithm

Taking natural logarithm:

$$\ln X_t = \ln X_0 + \mu t \qquad \text{Eq. -------- 4}$$

This equation is of the form of straight-line y = mx + c.

Thus, for an exponentially growing culture, a plot of natural log of biomass concentration against time (i.e. a semilog plot) will give a straight line with the slope equal to μ.

or $\mu = (\ln X_t - \ln X_0) / t$

Please note that if $\log_{10}$ values were plotted instead of natural log the slope would be equal to μ/2.303.

$$\mu = 2.303 (\text{Log } X_t - \text{Log} X_0) / t \qquad \text{Eq. -------- 5}$$

If we consider a situation where at time zero, the cell biomass is X_0, then after exponential growth equivalent to one doubling time (t_d), the microbial biomass will become $2 X_0$.

i.e. $X_t = 2X_0$ when $t = t_d$

Substituting these parameters in Equation 3 gives:

$$2X_0 = X_0 e^{\mu t d} \qquad \text{Eq. ---------- 6}$$

Taking natural logarithm gives:

$$\ln 2X_0 = \ln X_0 + \mu t_d$$

or $$\mu t_d = \ln 2$$

Thus,

$$t_d = 0.693 / \mu \qquad \text{Eq. ---------- 7}$$

Where t_d is the doubling time of the culture.

By calculating μ from the graph or from Equation 5, and substituting it in Equation 7, we can calculate the doubling time of the culture. From the Equation 7, we can also say that doubling time and specific growth rates are inversely related. Higher the doubling time, lower will be the specific growth rate and vice versa. Thus, the microbial culture having high specific growth rate will have low doubling time.

Specific growth rate of microbial cells, in general, is calculated during the balanced growth of cells particularly during the exponential phase of growth. It signifies growth capacity of the culture in the particular environment. Specific growth rates of microbial cells change during different phases of batch growth, having a maximum value during the exponential phase. It is also affected by growth temperature, medium composition and other environmental parameters that affect the microbial growth.

Let us now study the second approach, which examines growth in relation to cell number, where number of the cells at the start of the exponential growth is N_0.

Let us take an example where $N_0 = 1$.

As the cell divides, we shall have:

No. of cell division	0	1	2	3	n
No. of cells	1	2	4	8	2^n
Mathematically	N_o	$N_o \times 2$ $N_o 2^1$	$N_o \times 2 \times 2$ $N_o 2^2$	$N_o \times 2 \times 2 \times 2$ $N_o 2^3$	- $N_o 2^n$

Thus, after a period of exponential growth time (t), the number of cells (Nt) will be given by:

$$N_t = N_0 2^n \qquad \text{Eq.-------- 8}$$

Where,

n = the number of the divisions

N_0 = the initial cell number

Taking logarithm

$$\text{Log } N_t = \text{Log } N_0 + n \text{ Log } 2$$

Thus, the number of divisions (n) that have taken place is given by:

$$n = (\text{Log } N_t - \text{Log } N_0) / \text{Log } 2$$

$$= (\text{Log } N_t - \text{Log } N_0) / 0.301$$

$$= 3.3 (\text{Log } N_t - \text{Log } N_0) \qquad \text{Eq.-------- --- 9}$$

The mean generation time or doubling time (t_d) i.e. the time taken to undergo single generation that doubles the population is:

t_d = Total growth time / number of divisions

$$= t / n \qquad \text{Eq.---------- 10}$$

The specific growth rate and doubling times of the organism decide the medium requirements and fermentation batch time for the production of biochemical molecules, and thus, are important parameters for large-scale production process.

Example: Calculate the generation time (doubling time or t_d) of a bacterial population in which the number of bacteria increases from 10^4/ml to 10^7/ml during four hours of exponential growth.

Answer: First calculate the number of divisions the population must have undergone to increase from 10^4 to 10^7 in 4 hours.

Using the equation 9: $n = 3.3 (\text{Log } 10^7 - \text{Log } 10^4)$

$= 3.3 (7 - 4)$

$= 10$

Using the equation 10: $t_d = 240$ minutes / 10

= 24 minutes

6.1.4. Scale-up of microbial processes

In your class XI text book (Chapter I: Introduction to Biotechnology), you were introduced to the early attempts on large-scale production of acetone from the bacterium *Clostridium acetobutylicum*. This chapter also brings out a number of products, which are produced from microbes for which a large-scale cultivation of microbes is necessary. You were also introduced to

the quantitative approach of producing just 1 liter of curd at home. Imagine the issues involved when one has to produce thousands of liters of curd in dairy industry.

Direct production of microbes on a large or commercial scale has the risk of not only large investments, but also producing products, which may not be of appropriate quality so that there are problems in their commercialisation. To avoid these risks, the manufacturers try to validate laboratory process on an intermediate scale before attempting commercial production. This step is carried out in a **pilot plant**, which is a mini version of the commercial plant.

Before setting up a commercial scale plant the results, which are derived from pilot plant, are theoretically extrapolated to the commercial scale in terms of equipment size (fermentor / bioreactor), utilities like steam, water, electricity, fittings etc, labour (technical and non-technical) and market surveys. All these data are compiled into a techno-economic report to assess the feasibility of the project. Such a report is also important for seeking financial assistance for the project from banks or other financial institutions.

The following example will give you an idea about the importance of scale-up calculations for the bioprocess industry in terms of reactor size:

Recombinant insulin is produced at 100 mg/L by *E. coli* at a cell concentration of 1 g/L. Calculate the volume of reactor (size of the fermentor) needed to produce 1 Kilogram of insulin in the following conditions:

(a) When the cell concentration is 1 g/L and insulin production is 100 mg /L.

(b) When the cell concentration is 50 g/L and insulin production is 100 mg /L.

(c) When the cell concentration is 50 g/L and insulin production is 500 mg /L.

Answer:

(a) Insulin production is 100 mg/L; so fermentor volume needed for 1 Kg of insulin is 1 Kg / 100mg = 1000, 000mg/100,g = 10,000mg = 10,000L.

So we need 10,000-litre fermentor to produce 1 Kilogram of insulin in one batch.

(b) In this case the cell concentration is increased to 50 g/L; so insulin production per liter will be 50 X 100 = 5000 mg = 5 g / L;

Thus, to produce 1 Kilogram of insulin we need 1 Kilogram / 5 g = 1000 g / 5g = 200 g. So, if the cell concentration is increased 50 times, we need 200-litre reactor to produce 1 Kilogram of insulin.

(c) In this case cell concentration is high (50 g/L) and the insulin concentration is also very high (500 mg/L). Thus insulin yield per liter of culture is 500 X 50 = 25, 000 mg / L which is 25 gram/L. Thus, to produce 1 Kilogram of insulin we need 1 Kilogram / 25

gram1000g/25g = 40g thus, we need a 40 liter reactor to produce 1 kilogram of insulin.

Please notice that the reactor volume decreases when we have high cell density or high concentration of insulin production as in case of (c). Such calculations are useful in deciding the reactor size required to produce a desired amount of the metabolite. Usually 30% extra space is recommended in the fermenter vessel for air, froth etc.

6.1.5. Isolation of microbial products

Once the fermentation is complete, it is necessary to recover the desired metabolite. Minimally, this will involve separation of the cells from the fermentation broth. But it may also include, purification of the metabolite with or without cell disruption; cell disruption will be necessary if the metabolite is intracellular. Such operations are referred to as **downstream processing**. The steps involved in isolation of the desired microbial product are: (1) separation of cells from the fermented broth, (2) cell disruption if the product is intracellular or concentration of the broth if the product is extracellular (3) initial purification of the metabolite, (4) metabolite-specific purification in which the metabolite of interest is purified to a high degree, and (5) polishing of the metabolite (bringing it to 98 -100% purity) where it is further concentrated and formulated for use.

Fig. 9 and 10 depict the steps involved in isolation of microbial products or metabolites of extracellular and intracellular origin respectively. The important steps are: separation of microbial cells (biomass / pellet) from the fermentation broth, concentration, metabolite-specific purification and final purification. Isolation of cells from the fermented broth is, in general, carried out by either centrifugation or ultra filtration. Some cells rapidly settle out of suspension once aeration and agitation of the fermented broth ceases. The settling of cells may also be assisted by the addition of certain flocculating agents. Where cell settling does not occur, cell removal can be effected by centrifugation. An alternative to centrifugation is ultra filtration. The term ultra filtration describes processes in which particles significantly greater in size than the solvent are retained when the solution is forced through a membrane of very fine pore size, usually less than 0.5 mm. Microbial cells can be concentrated using ultra filtration so that the fermented broth is separated from cells.

The clarified fermentation liquor will contain microbial metabolites and extra cellular enzymes. Several methods are available for recovery of metabolites such as precipitation, solvent extraction and ion exchange chromatography.

Different downstream operations are available for concentration as well as purification of the metabolite. But it is always advisable to use lesser number of steps to achieve desired purity of the metabolite or product. This is because, more the number of steps involved, more will be the cost of the production and lower would be the yield.

Most of the antibiotics are secreted into the medium, so their isolation mainly involves steps

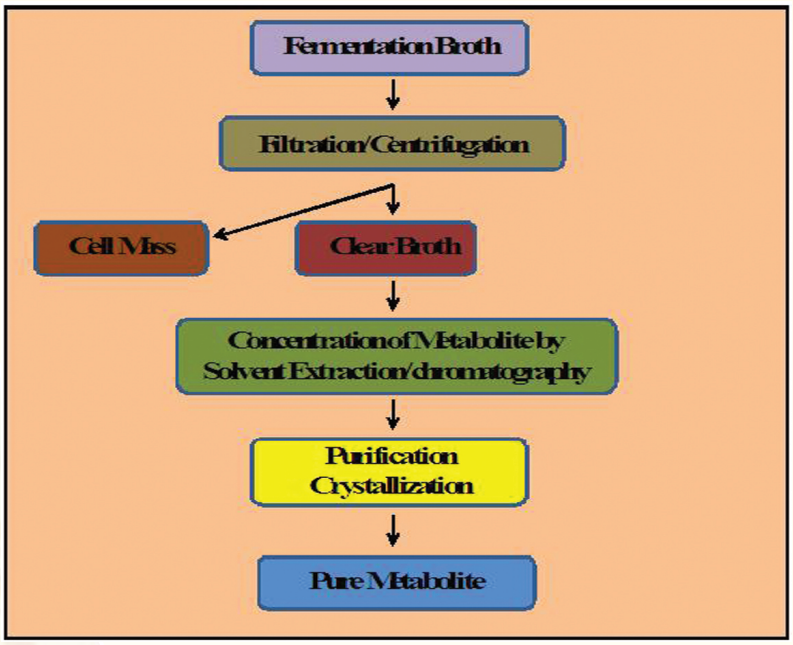


Fig. 9. Isolation of an extracellular microbial product (cell mass is discarded). Example: Streptomycin from *Streptomyces gresius*.

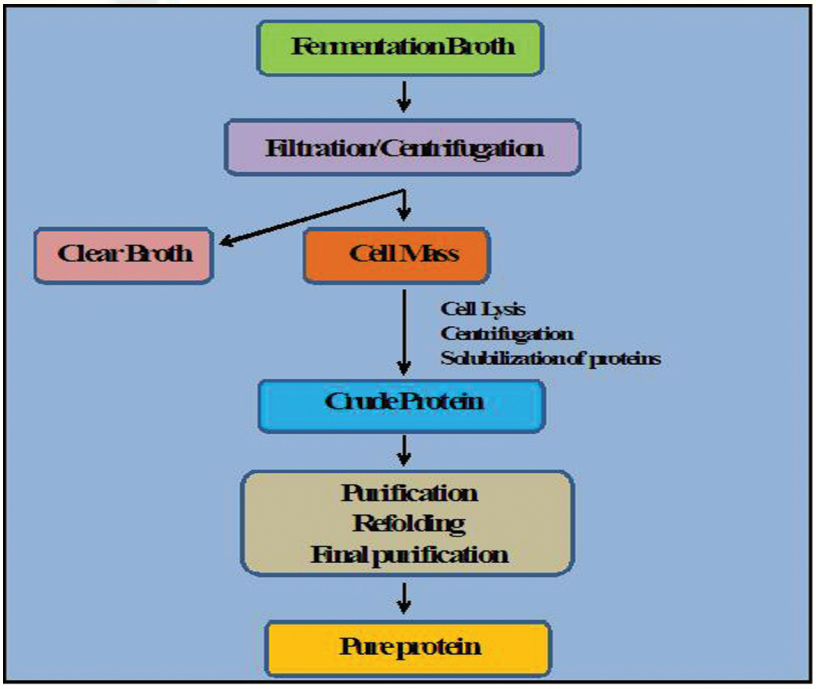


Fig. 10. Isolation of an intracellular microbial product (clear broth is discarded). Example: Recombinant insulin (Humulin®) from *E. coli*.

depicted in **Fig. 9**.

Most of the recombinant proteins expressed in *E. coli* accumulate intracellularly as protein aggregates. So their isolation and purification involve steps as described in **Fig. 10**. Once the pure metabolite is obtained, a stabilized formulation is made using several ingredients known as excipients.

6.1.6. Strain isolation, improvement and preservation

By this time, you must have realized that one of our major purposes to culture microbes is to produce useful products. But, do you think that all microbes produce useful or novel products? The answer is no. So, we have to search for those microbes, which produce substances of our interest. Where do we get these microorganisms?

As you know, microbes are wide spread in natural habitats especially in soil and in aquatic environments. They are also found in extreme environments namely arctic waters and hot springs. These natural habitats are our source of microbes. Once we have isolated microbes of desired interest, we can further improve their desired traits using a variety of methods.

Strain isolation:

The sample containing the microbes (e.g. soil) is put in a nutritive medium and allowed to grow in shake cultures. The growth conditions (e.g. temperature) or nutrients in the medium are provided such that these favour the growth of microbes of our interest. This is called **enrichment technique**. The enriched culture can further be sub-cultured by taking a small inoculum and putting it into fresh medium. In this way, the growth of the desired organisms improves successively. Further screening is done using a method where the organism will show its desired properties. For example, if we are looking for a microorganism, which produces an antibiotic, we may detect it by growing the culture on an agar plate in the presence of that bacterium against which antimicrobial activity is desired. Immunological methods are also available in which the microbes producing products are detected using **specific antibodies**. Molecular biology has made available a variety of **probes**, which enable the detection of organisms capable of producing specific products. Recently some of these methods have been adapted to robotic automation resulting in enormous throughput screening of microbes for newer / novel molecules.

Strain improvement:

Strain isolation procedure described above only identifies a strain, which has the capability or potential to produce a desired molecule. It does not ensure that it produces molecule in sufficient quantities to be economically viable. Techniques of classical genetics and genetic engineering are used to improve the desirable characteristics of the strain.

Mutation Selection: This is one of the oldest methods of strain improvement. The strain is exposed to chemical (e.g. nitrosoguanidine or NTG) or physical (e.g. UV rays) mutagens and the mutants having improved characteristics are selected. It is often necessary to carry out multiple successive mutations before we get the desired results. One of the classical examples of strain improvement using this methodology is the production of antibiotic penicillin. Several successive mutations were necessary to develop a strain of *Penicillium chrysogenum* capable of producing nearly 100 times the concentration of penicillin produced by the original strain *(Penicillium notatum)*, thus making production of penicillin commercially feasible.

Genetic Engineering Techniques: Until the recent breakthroughs in the techniques of genetic engineering, a bacterium could produce only substances coded for in its genome. Genetic engineering techniques about which you have learnt in the unit V (Chapter 1) allow totally new properties or capabilities to be added to the microorganisms giving rise to **recombinant strains.** Using these techniques, microorganisms may be manipulated to, synthesize or secrete enhanced quantities of biomolecules, facilitate production of novel compounds or allow utilization of cheaper substrates. Using these techniques, the microorganisms may also be utilized to produce plant, animal or human proteins. Some of the valuable human proteins which are being produced in microorganisms using this technology include recombinant human insulin (Humulin), hepatitis B surface antigen, human growth hormone and interferons. These proteins can now be produced in large quantities. Consequently the cost of the therapies which make use of these proteins viz. insulin (diabetes), hepatitis B surface antigen (vaccination against hepatitis B virus), human growth hormone (growth retardation) and interferons (immunotherapy) has been reduced considerably.

The tools, which are used for genetic engineering viz. the restriction enzymes, cloning and expression vectors and introduction of recombinant DNA into host cells, have been discussed fairly in detail in the unit V (Chapter 1). However there are many practical problems, which must be taken care of before a foreign **(heterologous)** gene may be expressed in a microorganism to make it commercially viable. For example, when a foreign gene is introduced into a host bacterium, it may not be expressed there. This problem is overcome by placing foreign gene under regulatory controls recognized by the host microorganism. To maximize production of foreign protein, the **expression vector** used is such that it replicates to **high copy number** and is stable. The foreign gene should ideally be linked to a strong promoter that has high affinity for RNA polymerase. The foreign gene may also be put under the control of a regulatory switch such that production of recombinant protein does not occur until required.

When a eukaryotic gene (e.g., plant, animal, human) is expressed in prokaryotic (bacterial) host, there are additional problems to be tackled. The non-coding region of eukaryotic gene must be excised. This requires use of reverse transcription of mRNA into cDNA. Additionally, the recombinant protein may not be secreted into the medium or its incorrect folding and accumulation intracellularly may generate **inclusion bodies**. All these problems make

downstream processing difficult and costly. Thus, an alternative would be to use a eukaryotic expression host. For this purpose, *Saccharomyces cerevisiae* has been quite popular because it is safe and scientists have long experience of using this yeast in industrial fermentations. Detailed information on biochemistry, physiology and genetics of this yeast is also known. Moreover, this yeast can be manipulated genetically rather easily. However, product yields are relatively low at 1-5% of the total protein. Other yeasts like *Pichia pastoris* has a number of advantages: it has strong inducible promoters; it is capable of making post-translational modifications similar to those performed by human cells; downstream processing is simpler as *Pichia* does not secrete its own proteins into the fermentation medium.

Metagenomics

In the last few years, another approach has been developed to identify and select microbial genes synthesizing novel molecules. This approach directly utilizes the large number of microbial genomes present in an environmental niche, for example in soil, in water such as ocean or in human gut. These genomes are contributed by both the culturable and the non-culturable variety of microbes and together constitute what has been termed as **metagenome**. The collective DNA is extracted from a sample of soil, water or any other environmental niche. It is subjected to restriction digestion using restriction endonucleses and the fragments are cloned as described in unit V (Chapter 1) of this book. The clones are then screened for presence of a variety of molecules. The clones expressing novel molecules or molecules with improved characteristics are used for large-scale production by fermentation techniques described in this chapter.

The metagenomic approach not only give the scientists an opportunity to cast a wider net on microbial resource present in the environment to fish out genes of their interest, it also gives them the opportunity to analyze the genomes of the microbes without culturing these in the laboratory. Thus, it is really a very useful approach to study those microbes, which are difficult to culture in the laboratory or have never been cultured in the laboratory as yet, and analyze these to see if they carry any genes, which may be exploited for human use. A typical procedure depicting metagenomic approach is shown in **Fig. 11**.

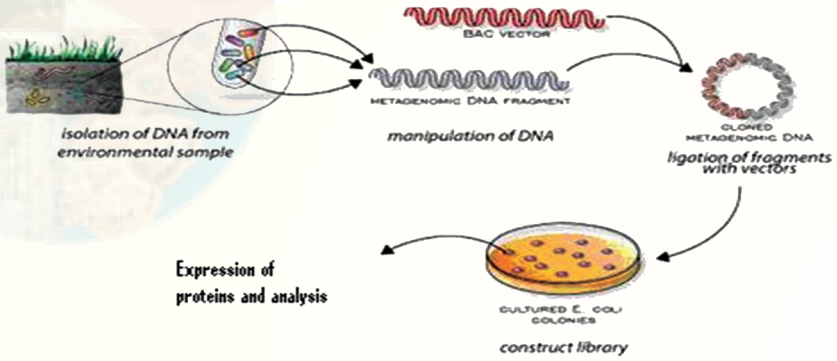


Fig. 11. A typical procedure depicting metagenomics approach .

Strain preservation:

Once a strain producing a novel or desired product has been obtained, it must be appropriately preserved for future use. If not done properly, the strain may be lost through loss of viability or even show decline in the production of the product for which it was isolated.

Storage on agar: Cultures are grown on agar slants or stabs & stored at 5 to -20 °C. These must be sub-cultured at approximately 6-month interval. The time of sub-culture may be extended to 1 year if cultures are covered with sterile mineral oil.

Storage in liquid nitrogen: The culture is grown and a cryoprotective agent like glycerol (10-30%) is added. These are dispensed in sealed ampoules & frozen in liquid nitrogen. (-176 to -196 °C).

Lyophilisation: Lyophilization or freeze-drying involves freezing of a culture followed by drying under vacuum. This results in sublimation of cell water. Lyophilised culture may remain viable for 5-10 years or more.

Culture Collections Centers

Cultures may be deposited to culture collection centers. These centers safely maintain cultures for years. The cultures are also made available to prospective investigators. With the advent of the modern biotechnology and the associated commercial and financial implications, the culture collection centers are governed by stringent rules & regulations to protect the intellectual property rights of the depositors. Some of the well-known culture collection centers of international repute are ATCC (American Type Culture Collection, USA), NCIB (National Collection of Industrial Bacteria, UK) and DSM (Deutsche Sammlung von Mikroorganismen and Zelkulturen, Germany). The National culture collection of India is called **MTCC (Microbial Type Culture a Collection and Gene Bank)** and is located at Institute of Microbial Technology, Chandigarh. Recently another National culture collection centre named **NBAIM (National Bureau of Agriculturally Important Microorganisms)** has been established in India at Mau in Uttar Pradesh (U.P.)

6.1.7. Applications of microbial culture technology

Microbial cultures have immense potential for production of very useful compounds. Once the microbial culture is established, depending on its metabolic activity it can be used for the production of numerous compounds. In general, microbial cultures can be exploited primarily in six different ways for the production of metabolites. They are listed below:

1. Production of whole microbial cells (for food, vaccines).
2. Production of primary metabolites (acids, alcohol).
3. Production of secondary metabolites (antibiotics).
4. Biotransformation reactions (enzymatic, steroid).
5. Exploitation of metabolism (microbial leaching, biodegradable waste treatment).
6. Synthesis of recombinant proteins (therapeutic proteins).

The most ancient use of microbial cultures is for the production of fermented foods such as curd and cheese where the whole bacteria are used as starter cultures. A complete list of such foods has been given in Unit I (Chapter 1) of the textbook for the class XI. The whole microorganisms are also used for preparations such as bacterial vaccines, e.g. vaccines for typhoid and tuberculosis. Single cell protein (SCP) is another example where the whole microorganisms are used a source of protein. Production of alcohol and acids are examples of primary metabolic products, whereas antibiotics are examples of secondary metabolites produced by different microorganisms. Microbial metabolism has also been exploited for the microbial production of vitamins. Extraction of metals from ores and treatment of liquid waste are also examples where microbial metabolism is used to convert unsuitable substrates to useful products.

Table 1. Some microbial species used for producing commercial products

Microorganisms	Products
Saccharomyces cerevisiae	Ethanol
Aspergillus niger	Citric Acid
Penicillium chrysogenum	Penicillin
Streptomyces griseus	Streptomycin
Corynebacterium glutamicum	L-Lysine
Propionibacterium shermanii	Vitamin B12
Aspergillus oryzae	Amylases
Leuconostoc mesenteroides	Dextran
Escherichia coli (via recombinant technology)	Insulin, growth hormones and interferons
Saccharomyces cerevisiae (via recombinant technology)	Hepatitis B surface antigen
Alcaligenes eutrophus	Poly 3-hydroxybutyrate (PHB)

One of the most important uses of microorganisms in recent years has been as hosts for production of recombinant proteins using genetic engineering techniques. Expression of human insulin in *Escherichia coli* and hepatitis B surface antigen in Yeast for making hepatitis B vaccine are the most notable examples of applications of microbes for human use. Some specific examples of products derived from microbial cultures have been given in **Table 1**.

6.1.8. Biosafety issues in Microbial Technology

Till date, all the products of the biotechnology industry have been found to be safe. Nevertheless, for all biotechnology processes, safety is of paramount importance. The main areas of consideration for safety aspects specific to biotechnology are:

- Potential of genetically modified organisms (GMO) or recombinant strains to infect humans, animals and plants to cause diseases.
- Toxicity and allergy associated with the use of microbially produced biomolecules especially the recombinant molecules.
- Other medically relevant implications like increasing the environmental pool of antibiotic resistant microorganisms or transfer of antibiotic resistant genes.
- Problems associated with the disposal of spent microbial biomass.
- Safety aspects associated with contamination, infection or mutation of process strains.

When the microbes are to be used in industrial processes, it is preferred to use only those microorganisms, which are included in the **'GRAS' (generally regarded as safe)** category. Also, when microbes are to be used as hosts for production of recombinant molecules, this should be carried out using GRAS organisms.

The main fear associated with the genetically engineered microorganisms (GMO) or recombinant strains is that they could escape from the laboratory into the environment with unpredictable and perhaps catastrophic consequences. It was believed that such released microorganisms could upset the balance of nature or that foreign DNA in the new microorganism could alter the metabolic activity of the host microbe in unpredictable and undesirable ways. Guidelines have been established to ensure safe working practices and levels of containment based on potential hazards. Many of the GRAS organisms found in nature have been genetically modified so that these may be used as host for production of recombinant biomolecules. At present, all new releases of genetically modified organisms are evaluated by expert committees on a case by case basis for various safety aspects before approval for their use is granted.

Besides these, whenever we are required to culture microbes whether for research or for industrial applications, we must follow **good microbiological practices (GMP)**. These include:

i. Persons must wash their hands with germicidal soap after handling viable microorganisms.

i. Eating, drinking, smoking etc. are not permitted in the working area.

ii. Food is to be stored outside the work area in cabinets / refrigerators designated for this purpose only.

iii. Mouth pipetting is strictly prohibited.

iv. All procedures are carried out in a way, so as to minimize splashes and generation of aerosols.

v. After work with viable microorganisms, work surfaces are decontaminated thoroughly.

vi. It is recommended that laboratory coats / aprons should be worn while working.

vii. All cultures, stocks or other waste are decontaminated and autoclaved before disposal.

Review Questions

1. While culturing microbes in the laboratory in a flask, what measures do you suggest to enhance their growth?
2. Differentiate between:
   a. Batch and Fed batch culture.
   b. Chemostat and turbidostat.
3. How is a continuous culture better than batch or fed batch cultures?
4. Explain what is meant by steady state in relation to the growth of microbial cultures?
5. How is the large quantity of air, required in industrial fermentors, sterilized?
6. What problems must be tackled while expressing a eukaryotic gene in a prokaryotic host?
7. How would you ensure that production of a recombinant molecule does not occur until required?
8. What problems make the downstream processing of recombinant proteins difficult and costly?
9. How is *Pichia pastoris* better expression host compared to *Saccharomyces cerevisiae*?

10. How metagenomic approach helps to identify newer/novel genes?
11. Suggest two methods of preserving microbial strains.
12. What is lyophilisation?
13. Explain what is meant by GMP, GRAS?
14. What are the functions of the microbial culture collections?
15. Enlist five good laboratory practices, which need to be followed while working with microbes.
16. Why lyophilised cultures of microbes remain viable for several years?

References

1. Introduction to Biotechnology, By C.M. Brown, I. Campbell and EG. Priest, Panima Publishing Corporation (2002).
2. Industrial Microbiology, An introduction By M.W. Waites, N.L. Morgan, J.S. Rockey and G. Higton, Blackwell Publishing (2007).
3. Microbial Biotechnology, Principles and Applications By Lee Yuan Kun (Ed), World Scientific (2004).
4. Process Engineering in Biotechnology By A. T. Jackson, Open University Press (1990).
5. Microbial Biotechnology, Fundamentals of Applied Microbiology (2ND Ed.) By A. N. Glazer and H. Nikaido, Cambridge University Press (2007).

CHAPTER

2

PLANT CELL CULTURE AND APPLICATIONS

6.2.1. Introduction

The plant cell culture is based on a unique property of the cell, i.e., totipotency which may be defined as the ability of a plant cell to regenerate into whole plant on a defined artificial nutrient medium under the suitable physical conditions in the laboratory. In fact, it was Schwann who first drew attention to the fact that a single cell has the capacity to both grow and divide in a self regulatory fashion and that a single cell is also totipotent. This was stated in the famous 'cell theory' which was proposed in the beginning of 19th century by Schleiden and Schwann. **Gottlieb Haberlandt,** who is regarded as **'Father of Plant Tissue Culture',** first attempted in 1902 to cultivate the mechanically isolated plant leaf cells on a simple nutrient medium. Though unsuccessful in achieving the growth and differentiation of the cultured cells, he has made several predictions in plant tissue culture like the concept of growth hormones, the use of embryo sac fluids, the cultivation of artificial embryos from somatic cells, etc. and indeed all of his predictions were found to be true as demonstrated by later researchers. In the first phase during the period 1902 - 1930s, attempts were made by several scientists to culture the isolated plant organs such as roots and shoot apices (organ culture). In the second phase during 1940s - 1970s, the extensive studies were undertaken to develop the suitable nutrient media to culture plant tissues, embryos, anthers, pollen, cells and protoplasts, and the regeneration of complete plants (in vitro morphogenesis) from cultured tissues and cells. Since 1980s, a new era has started involving the introduction of foreign genes into crop plants using cell and tissue culture systems to develop genetically modified (GM) or transgenic crops with improved characteristics, which may be responsible for the 'second green revolution.

6.2.2. Cell and Tissue Culture Techniques

Basic technique

The whole plants can be regenerated virtually from any plant part (referred to as explant) or **cells**. The basic technique of plant tissue culture **(Fig. 1)** involves the following steps:

1. Selection of suitable explants like shoot tip, leaf, cotyledon and hypocotyls.
2. Surface sterilization of the explants by disinfectants (e.g. sodium hypochlorite) and then washing the explants with sterile distilled water.
3. Inoculation (transfer) of the explants onto the suitable nutrient medium (shoot regeneration medium, which is sterilized by autoclaving or filter-sterilized to avoid

microbial contamination) in culture vessels under sterile conditions (i.e., in laminar flow cabinet).

4. Growing the cultures in the growth chamber or plant tissue culture room **(Fig. 2A)**, having the appropriate physical conditions [i.e., artificial light (16 h photoperiod), temperature (~26°C) and relative humidity (50-60%)].
5. Regeneration of shoots from cultured plant tissues and their elongation.
6. Rooting of regenerated shoots on rooting medium.
7. Transfer of plants to the transgenic green-house **(Fig. 2B)** or field conditions following the acclimatization (tissue hardening) of the regenerated plants.

Nutrient media

The *in vitro* culture of plant parts or cells require a variety of nutrients and suitable physical conditions, unlike the intact plants which can synthesize their own food and many other essential compounds needed for their growth and development using light, CO_2, water and minerals. The composition of plant tissue culture medium can vary, depending upon the type of plant tissues or cells that are used for culture.

The typical plant tissue culture nutrient medium consists of inorganic salts (both micro- and macro-elements), a carbon source (usually sucrose), vitamins (e.g. nicotinic acid, thiamine, pyridoxine and myo-inositol), amino acids (e.g., arginine) and growth regulators (e.g. auxins, cytokinins and gibbrellins). Other compounds like casein hydrolysate, coconut milk, malt extract, yeast extract, tomato juice, etc. may be added

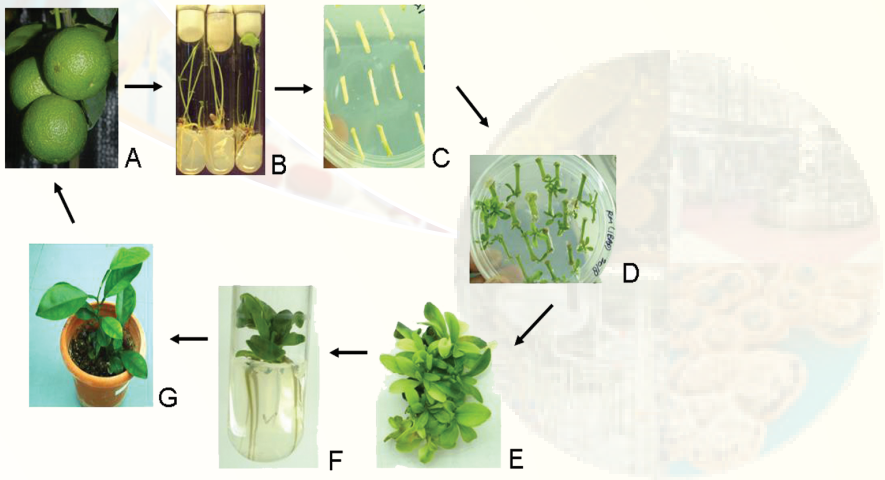


Fig. 1. Various steps involved in the regeneration of whole plants using tissue culture techniques. (A) Citrus plant with fruits, (B) Seedlings raised by germinating Citrus seeds on nutrient medium, (C) Transfer of epicotyl explants onto shoot regeneration medium, (D) Induction of shoots from explants, (E) Elongation of shoots, (F) Rooting of *in vitro* shoot, (G) Regenerated plant in a pot.

Fig. 2. Plant tissue culture room (A) and transgenic green-house (B).

for specific purposes. Plant hormones play a pivotal role in growth and differentiation of cultured cells and tissues. An optimum pH (usually 5.7) is also very important.

There are two kinds of media - **liquid and solid media**, and in the later case a gelling agent such as **agar** (a polysaccharide obtained from a red algae, *Gelidium amansii)* is added for solidification of the medium. There are several types of media like MS, LS, B5 and Nitsch's medium, and the choice of media is dictated by the plant species, variety or plant part. However, the most extensively used nutrient medium is MS medium, which was developed by Murashige and Skoog in 1962. The success of plant regeneration from cultured cells and tissues are largely governed by the composition of the culture medium. The sterilization of nutrient media by autoclaving or membrane filter-sterilization is utmost important to avoid microbial contamination.

Types of cultures

Organ culture: It deals with the culture of the isolated organs (like roots) under laboratory conditions *(in vitro)*, and different names are given depending upon the organ utilized for the culture. For instance, the culture of roots, endosperm, ovary, ovule and anther are called as i) Root culture, ii) Endosperm culture, iii) Ovary culture, iv) Ovule culture and v) Anther culture respectively.

Explant culture: The culture of plant parts **(explants)** is known as explant culture. The explants can be any part of the plant like the piece of stem, leaf, cotyledon, hypocotyls, etc. The explant cultures are generally used to induce callus or plant regeneration.

Callus culture: Callus refers to an unorganised mass of cells, which are generally parenchymatous in nature **(Fig. 3A)**. A variety of plant parts can be induced to show callusing response, but the response varies with the composition of culture media. Generally, auxins are added to culture medium for callus induction but the nature and quantity of auxin added depends on the nature and source of explant and its genotype besides other factors. Callus cultures can be maintained for prolonged period of time by repeated sub-culture. Callus cultures are used for

- plant regeneration.
- preparation of the single cell suspensions and protoplasts.
- genetic transformation studies.

Cell suspension cultures: Single cells can be isolated from either callus or any other part of the plant (e.g. leaf) and cultured in liquid medium. Both mechanical and enzymatic methods can be used for isolation of plant cells. Mechanical method involves grinding of the tissue to a fine suspension in a buffered medium followed by filtration/centrifugation to get rid of cell debris. The enzymatic method is based on the usage of enzymes (pectinase/macerozyme), which dissolve the middle lamella between the cells, i.e., the inter-cellular cement, to release single cells. Once the cells have been isolated, they may be cultured by **batch cultures** or **continuous cultures**, which have been discussed in detail (see the previous chapter). The cell suspension cultures can be used for:

- induction of somatic embryos/shoots.
- *in vitro* mutagenesis and mutant selection.
- genetic transformation.
- production of secondary metabolites.

Mass cell culture: It involves the large-scale culture of cells in specially designed 'plant bioreactors', which essentially do not have a stirrer (as plant cells are shear sensitive). In place of stirrer, gas is gently bubbled, which provides stirring as well as meet the demand of a higher oxygen supply. Mass cell cultures are ideal for producing pharmaceutically important secondary metabolites.

Protoplast culture: Protoplasts are plant cells without cell wall **(Fig. 3B)** and can be isolated from a variety of plant tissues (usually leaves, callus pieces, single cells or pollen grains) by enzymatic method using cell wall digesting enzymes (cellulases, hemicellulases and pectinases). Protoplasts are usually cultured by suspension culture in petri plates. As the protoplasts lack cell wall, they can be utilized for many purposes such as:

- Various biochemical and metabolic studies.
- Fusion of two somatic cells to produce somatic hybrids.
- Fusion of enucleated and nucleated protoplasts to produce cytoplasmic hybrids (Cybrids).
- Genetic transformation.

Protoplast fusion: The protoplasts from two different plant genotypes can be fused in the presence of fusogenic agents like polyethylene glycol (PEG - most widely used and most successful method for protoplast fusion) or by electro-fusion. The hybrid cells **(heterokaryons)** can be selected by various methods such as the use of different antibiotic markers or fluorescent dyes for two different protoplasts. Then, regenerated somatic hybrids from hybrid cells are characterized by morphology and molecular analysis (e.g. RAPD). This technique offers a unique method to raise new genotypes (i.e. somatic hybrids between distantly related plants or cybrids) with useful agronomic traits.

Plant regeneration pathways

The plants can be regenerated by either **organogenesis** and **somatic embryogenesis**.

Organogenesis means formation of organs like shoots from the cultured explants **(Fig. 3C)**. Miller and Skoog experimentally proved that formation of shoot or root first on the cultured tissue depends on the relative concentration of auxin and cytokinin. If auxins are high in the medium, then it promotes rooting while if cytokinins are high, shoot formation is promoted.

In **somatic embryogenesis**, the totipotent cells may undergo embryogenic pathway to form somatic embryos **(Fig. 3D-G)**, which can be grown to regenerate into complete plants. Generally, somatic embryos resemble the zygotic embryos (seed embryos) except in their place of origin and larger size. For the first time, Steward in 1958 and Reinert in 1959 independently reported the somatic embryogenesis from carrot cultures.

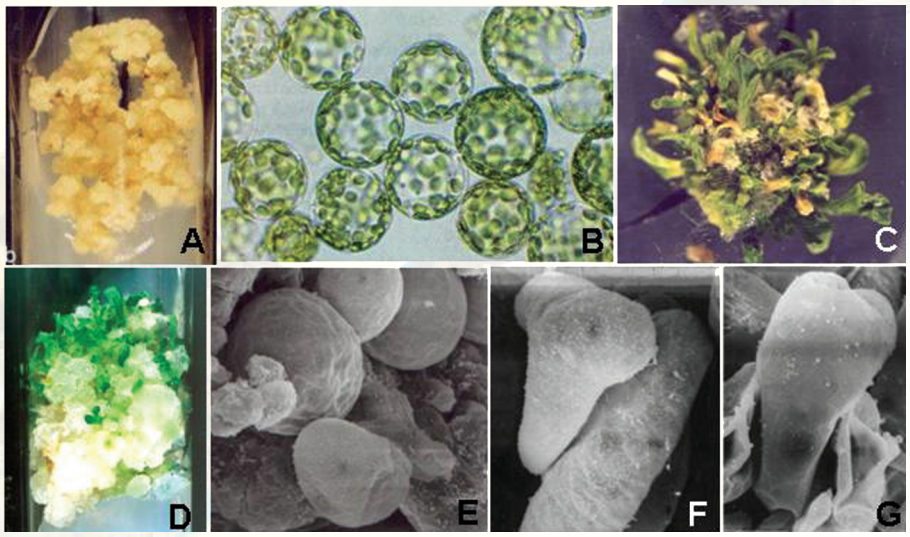


Fig. 3. Callus culture of indica rice (A), protoplasts isolated from leaf mesophyll cells of tobacco (B), multiple shoot regeneration from leaf explants of brinjal (C), regeneration of somatic embryos from leaf explants of brinjal (D) and scanning electron microscopy photographs of the various stages of somatic embryogenesis - globular (E), heart-shape (F) and torpedo (G) stage embryos from the brinjal culture (D).

6.2.3. Applications of Cell and Tissue Culture

The plant cell culture offers many potential applications in agriculture and health-care. Some of the important applications of plant cell and tissue culture are given below.

Micropropagation

Vegetative propagation of plants is of considerable importance in agriculture, horticulture and forestry as it provides the multiplication of uniform material for crop planting **(Clones)**. Traditionally, it is done by using cuttings, budding, grafting, corms, tubers and other vegetative propagules. The main problem with this method is that it is labour-intensive, low productivity and seasonal. Thus, tissue culture method of plant propagation, known as '**micropropagation**' can be used to overcome the problems mentioned above. This technique utilizes the culture of apical shoots, axillary buds and meristems on suitable nutrient medium, and has been adopted for commercialisation of important plants such as banana, apple, pears, strawberry, cardamom, many ornamentals (e.g. Orchids) and other plants.

There are four defined steps in micropropagation method. These are:

- **Initiation of culture -** from an explant like shoot tip on a suitable nutrient medium.
- **Shoot multiplication -** multiple shoots formation **(Fig. 4A)** from the cultured explant.
- **Rooting of shoots -** rooting of *in vitro* developed shoots on rooting medium.
- **Transplantation -** the hardening of tissue culture raised plants **(Fig. 4B)** as they are tender and subsequent transplantation to the green-houses or field.

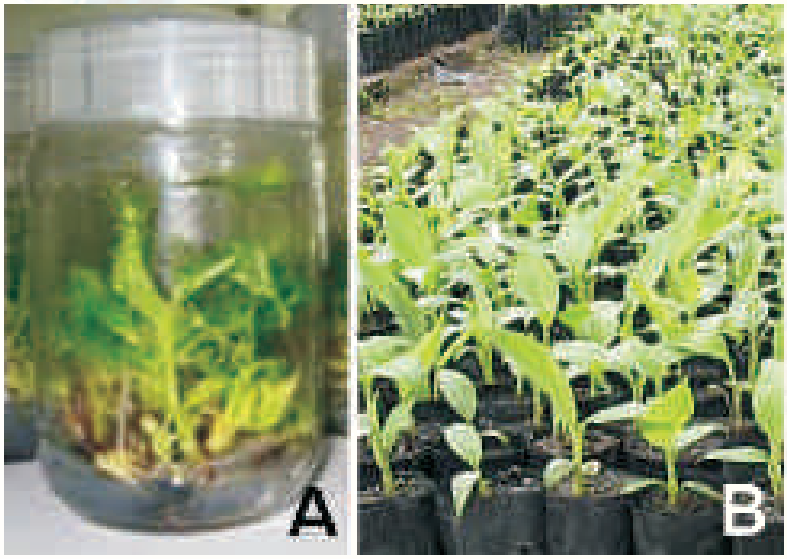


Fig. 4. Micropropagation of banana using tissue culture techniques. Multiple shoot regeneration (A) and acclimatized regenerated banana plants in green-house (B).

In recent years, much effort has been devoted to develop automated, robotised and more efficient transplant production methods. In India, the Department of Biotechnology (DBT) initiated a major network programme in 1989, and so far propagation protocols have been developed for more than 60 elite plant species.

Virus-free plants

Most of the crops plants, particularly vegetatively propagated plants are systemically infected with viruses. If the stock of a cultivar (vegetative propagules) are infected with a virus, the entire clonal population raised with such stock will also be infected and reduce the yield and quality significantly. Therefore, the production of virus-free plants is important to increase yield and quality. Interestingly, the distribution of viruses in plants is uneven, and the apical or axillary meristems are generally free from viral particles. This has enabled scientists to produce virus-free plants by culturing small meristems (usually less than 1 mm long) collected from virus-infected plants. Meristem culture technique is now successfully used to produce virus-free plants (but not virus-resistant plants) in a number of important clonal plants like potato, sugarcane, banana and apple. The scheme used for the production of virus-free plants by meristem culture is essentially same as micropropagation technique, except that the starting material (explant) is meristem which is dissected out from shoot tips of infected plants.

Artificial seeds

The artificial seeds (also called as synthetic seeds or somatic seeds) can be utilized for the rapid and mass propagation of elite plant species as well as hybrid varieties. Artificial seeds are produced either by encapsulating the somatic embryos in a protective coating, i.e., calcium alginate beads or by desiccating the somatic embryos with or without coating. A typical artificial seed is shown in **Fig. 5**.

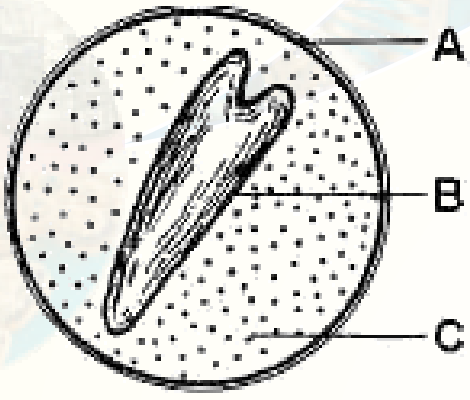


Fig. 5. The diagrammatic representation of an artificial seed. The artificial seed coat (A), somatic embryo at torpedo stage (B) and artificial endosperm (C).

Embryo rescue

It is very difficult to produce hybrids in case of inter-specific and inter-generic crosses (crosses between distantly related plants) because of abnormal development of the endosperm, which causes premature death of the hybrid embryo and leads to sterile seeds. The embryo from such sterile hybrid seeds can be excised at an appropriate time and cultured on a suitable nutrient medium to produce novel hybrids which is otherwise not possible. This is known as embryo rescue. Several useful hybrids were produced in a variety of crops using this technique.

Haploids and triploids

The technique of haploid production through anther and pollen culture as well ovary culture is of immense use in plant breeding to improve crop plants. It enables raising plants expressing traits that are otherwise recessive. The genetically homozygous diploid plants, which serve as parents in cross breeding can also be produced by diploidisation of haploid plants using colchicine chemical.

The endosperm is a triploid tissue and an excellent material to produce triploid plants by culturing endosperm tissue. Triploid plants usually show seed sterility or seedlessness, which is desirable in crops like citrus, apple and pear.

Somatic hybrids and cybrids

For many years, sexual hybridisation has been successfully used for crop improvement, however this method is limited in most cases, especially inter-specific and inter-generic crosses. The **somatic cell hybridisation** (also known as **parasexual hybridisation**) offers an excellent alternative for obtaining distant hybrids with useful agronomic traits (known as **somatic hybrids** or **parasexual hybrids**), which would never be formed in nature through sexual fertilization. Essentially, the protoplasts are isolated from the two different plants and are allowed to fuse with each other and the required fusion products (hybrid cells) are selected as discussed earlier in this chapter (see protoplast isolation and fusion).

The first inter-specific somatic hybrid between *Nicotiana glauca* and *N. langsdorffii* was produced by Carlson and his associates in 1972. Later, Melchers and his team in 1978 developed the first inter-genetic somatic hybrids between *Solanum tuberosum* (potato) and *Lycopersicon esculentum* (tomato) and the hybrids are known as '**Pomatoes** or **Topatoes**'. Several inter-specific somatic hybrids have been produced in number of crops like brinjal **(Fig. 6)**.

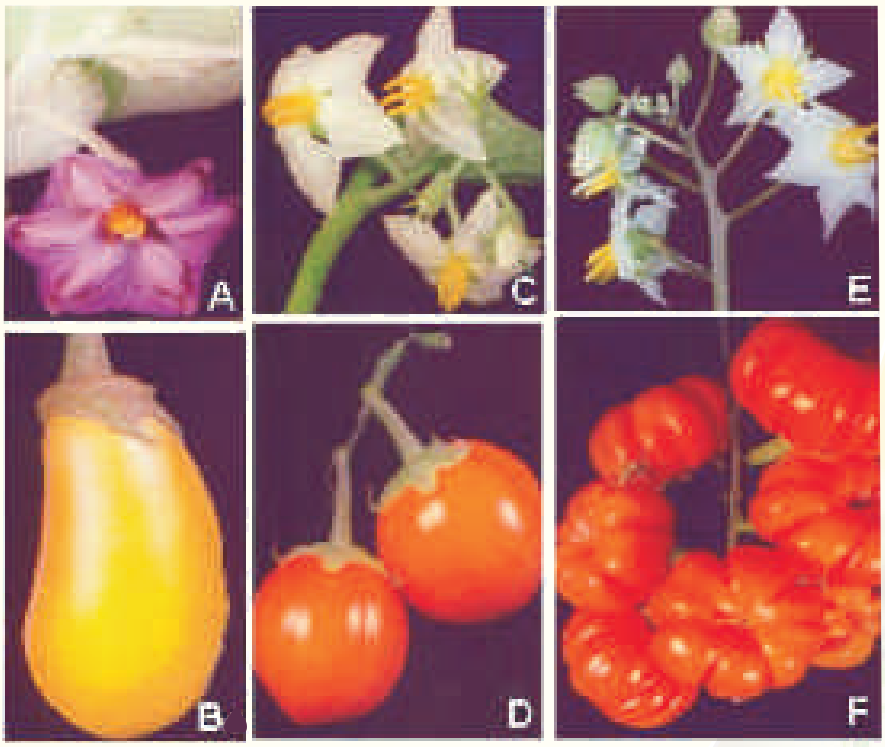


Fig. 6. Development of somatic hybrids in brinjal for transferring disease resistance from wild species to the cultivated variety. Flowers and fruits of the cultivated brinjal **(A & B)**, somatic hybrids **(C & D)** and wild-type brinjal **(E & F)**.

It is also possible to produce cytoplasmic hybrids **(cybrids)** through protoplast fusion in which the genomes of one of the partners is lost. Alternatively, the isolated and purified organelles - chloroplasts or mitochondria from one species can be fused with the recipient protoplasts from a different plant species (known as **organelle transfer** or **organelle uptake**) to transfer useful cytoplasmic traits like herbicide tolerance and cytoplasmic male sterility.

Production of secondary metabolites

Plants produce throusands of sophisticated chemical molecules. These include the chemicals required for a plant's basic metabolic processes such as sugars, lipids, amino acids and nucleic acids **(primary metabolites)** and also some other additional products such as alkaloids, resins, tannins, latex, etc. **(secondary metabolites)**.

The function of secondary metabolites in plants is not clearly understood, although they have been implicated in defence mechanism of the plant against pests and pathogens as well as

feeding by animals. However, many of the secondary products, especially various alkaloids are of immense use in medicine **(Table 1)**. Such compounds are produced in plants only in small amounts, and therefore, they are quite expensive.

Table 1. Few examples of industrially important plant secondary metabolites produced through cell and tissue cultures

Product	Plant source	Uses
Artemisin	*Artemisia* spp.	Antimalarial
Azadirachtin	*Azadirachta indica* (Neem)	Insecticidal
Berberine	*Coptis japonica*	Antibacterial, Antiinflammatory
Capsaicin	*Capsicum annuum* (chilli)	Reumatic pain treatment
Codeine	*Papaver* spp.	Analgesic
Digoxin	*Digitalis lanata*	Cardiac tonic
Diosgenin	*Dioscorea deltoidea*	Antifertility
Scopolamine	*Datura stramonium*	Antihypertensive
Quinine	*Cinchona officinalis*	Antimalarial
Shikonin	*Lithospermum erythrorhizon*	Antimicrobial; Red pigment used in lipstics & dye for silk
Taxol	*Taxus* spp.	Anticarcinogenic
Vincristine	*Cathranthus roseus*	Anticarcinogenic

The indiscriminative use of such valuable medicinal and other plants has brought them to near extinction. A possible solution in this direction is provided by cell and root cultures. Certain plant products such as shikonin are being produced commercially in large-scale using bioreactors.

Somaclonal variation

It has been observed that the long-term callus and cell suspension culture and plants regenerated from such cultures are often associated with chromosomal variations **(somaclonal variation)**. It is this property of cultured cells that finds potential application in the crop improvement and in the production of mutants (e.g., disease resistance in potato).

Larkin and Scowcroft (1981) proposed the term **'somaclones'** for plant variants obtained from tissue cultures of somatic tissues. Similarly, if the tissue from which the variants have been obtained is having gametophytic origin such as pollen or egg cell, it is known as **'gametoclonal variation.**

In vitro plant germplasm conservation

Germplasm refers to the sum total of all the genes present in a crop and its related species. The availability of a wide diversity (which includes land races, cultivars and primitive wild species of crop plants) and its conservation is utmost important as they are invaluable for future breeding programmes. However, the existence of primitive cultivars and wild relatives of crop plants has been endangered by the extensive use of newly introduced cultivars and hybrids, and this has posed a great concern to avoid the erosion of plant genetic resources. Therefore, attempts have been made to preserve the genetic resources by conventional methods (e.g. seeds, vegetative propagules, etc. and it is known as ***in vivo* gene banks**) as well as non-conventional methods, i.e. cell and tissue culture methods (known as ***in vitro* gene banks**). This will ensure the availability of valuable germplasm to breeder, to develop new and improved varieties.

The germplasm conservation through the conventional methods has several limitations such as, the short-lived seeds, seed dormancy, seed-borne diseases and high inputs of cost and labour. On the other hand, the biotechnological approaches for the conservation of plant germplasm can surpass many of these problems, and it can be done by the following approaches.

- **Freezing storage or Cryopreservation** - this utilizes the long-term preservation of cells and tissues (e.g. shoot tips, axillary buds, meristems, somatic embryos, etc.) at ultra-low temperature (-196°C, i.e. in liquid nitrogen) for indefinite time by using cryoprotectants (e.g. dimethylsulfoxide, glycerol, proline and mannitol). The cells and tissues can be recovered after thawing, and can be used for regeneration of plants.
- **Cold storage** - this uses the short- or medium-term storage of germplasm by using shoot tip, nodal or meristem explant cultures. The storage in done under conditions that impose slow growth such as low temperature (4°C or 15°C), nutrient limitation or the addition of growth retardants (e.g. absisic acid) in medium.

6.2.4. Gene transfer methods in plants

For achieving genetic transformation in plants, the basic pre-requisite is the construction of a vector (genetic vehicle) which carrys the genes of interest flanked by the necessary controlling sequences, i.e. promoter and terminator, and deliver the genes into the host plant. There are two kinds of gene transfer methods in plants as discussed below.

Vector-mediated or indirect gene transfer

Among the various vectors used in plant transformation, the Ti-plasmid of *Agrobacterium tumefaciens* has been used extensively. This bacterium contains large size plasmid, known as Ti-plasmid (tumour-inducing plasmid) and portion of this plasmid referred as **T-DNA** (transferred DNA) is transferred to plant genome in the infected cells and cause plant tumours (**crown galls - Fig. 7A**). This means that *A. tumefaciens* has natural ability to transfer T-DNA of its plasmid into plant genome. (plant chromosomes) upon infection of cells at the wound site, and therefore this bacterium is known as '**natural genetic engineer of plants**'. Because of this unique property, Ti-plasmid can be used as gene vectors for delivering useful foreign genes into plant cells and tissues. The foreign genes (**transgenes**), i.e., the gene of interest (e.g. Bt gene for insect resistance) and plant selection marker gene, usually an antibiotic gene like *nptII* which confer resistance to kanamycin are cloned in the T-DNA region of Ti-plasmid in place of unwanted DNA sequences **(Fig. 7B)**.

The general strategy for transforming plants is to collect leaf discs (in case of dicots) or embryogenic callus (in case of monocots like cereals) and then infect the tissue with Agrobacterium carrying recombinant disarmed Ti-plasmid vector **(Fig. 7B)**. The infected tissue can then be cultured (**co-cultivation**) on shoot regeneration medium for 2-3 days during which time the transfer of T-DNA along with foreign genes takes place. Later, the transformed tissues (leaf discs/calli) are transferred onto selection cum plant regeneration medium supplemented with usually lethal concentration of an antibiotic (e.g. kanamycin) to selectively eliminate non-transformed tissues. This **selection medium** also contain a bacteriostatic agent, i.e. an antibiotic like cefotaxime which suppresses or kills the Agrobacterium present with the transformed tissues and is no longer needed as the transfer of foreign genes has already taken place during co-cultivation. After 3-5 weeks, the regenerated shoots (from leaf discs) are transferred to root-inducing medium, and after another 3-4 weeks, complete plants can be transferred to soil following the hardening (acclimatisation) of regenerated plants. The presence of foreign genes in the putative transgenic plants can be confirmed by molecular techniques like PCR and Southern blot hybridisation.

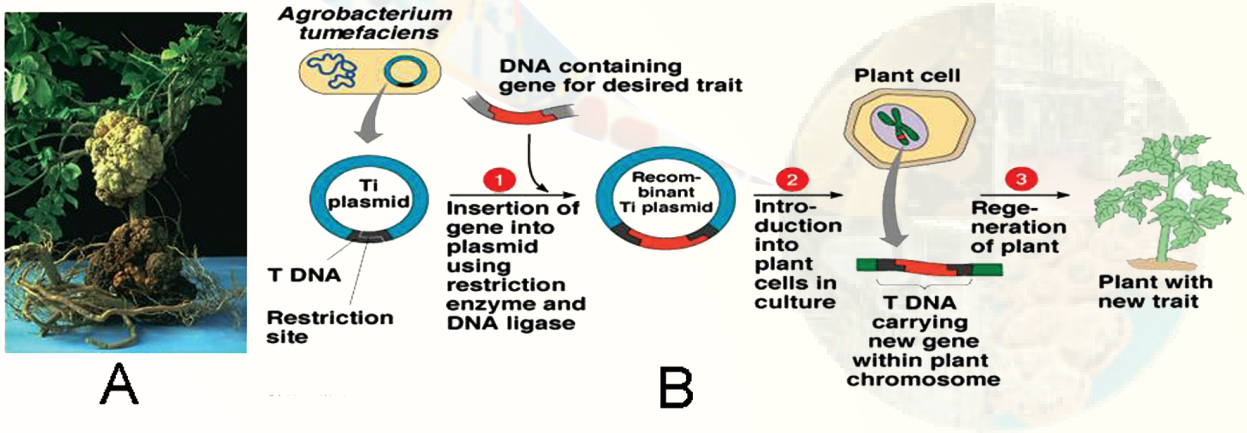


Fig. 7. Induction of crown gall on the wounded stem induced by wild-type virulent *Agrobacterium tumefaciens* (A), and the schematic diagram showing the cloning of the gene of interest in Ti-plasmid of Agrobacterium and its transfer to plant cells in culture to produce transgenic plants with desirable traits (B).

Vectorless or direct gene transfer

In the direct gene transfer methods, the foreign gene of interest is delivered into the host plant cell without the help of a vector. The following are some of the common methods of direct gene transfer in plants.

Chemical mediated gene transfer: Certain chemicals like polyethylene glycol (PEG) and dextran sulphate induce DNA uptake into plant protoplasts.

Microinjection: Here, the DNA is directly injected into plant protoplasts (specifically into the nucleus) using fine tipped (0.5-10 micrometer diameter) glass needle or micropipette to transform plant cells. The protoplasts are immobilized on a solid support (e.g. agarose on a microscopic slide) or held with a holding pipette under suction, and the DNA is injected into the protoplasts with the help of a fine microscope.

Electroporation: In this case, a pulse of high voltage is applied to protoplasts/cells/tissues, which makes transient (temporary) pores in the plasma-membrane which facilitates the uptake of foreign DNA. The DNA is mixed with the plant protoplasts in a specially made cuvettes before the application of the electric field.

Particle gun: A popular and widely used direct gene transfer method for delivering foreign genes into virtually any tissues and cells or even intact seedlings. In this method, the foreign DNA (containing the genes of interest) is coated (precipitated) onto the surface of minute gold or tungsten particles (1-3 micrometers) and bombarded (shot) onto the target tissue or cells using a particle gun (also called as gene gun/shotgun/microprojectile gun). Then, the bombarded cells or tissues are cultured on selection medium to regenerate plants from the transformed cells **(Fig. 8)**.

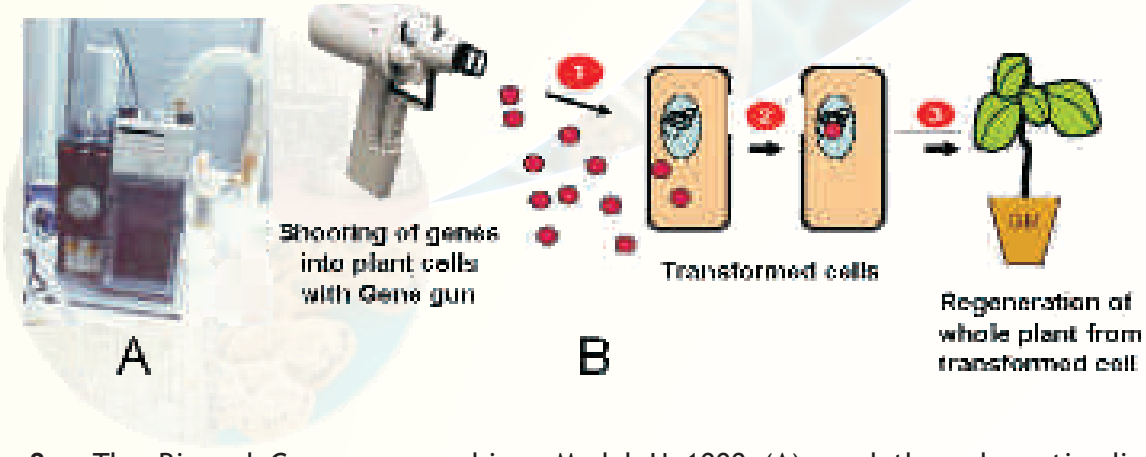


Fig. 8. The Bio-rad Gene gun machine, Model He1000 (A), and the schematic diagram showing the bombardment of gold particles containing the DNA (gene of interest) on their surface onto the plant cells in culture to produce transgenic plants with desirable traits (B).

Transgene analysis

The selection of transformed plant cells from untransformed cells is an important step in the plant genetic engineering. For selection, the gene of interest is transferred along with a genetic marker gene (usually antibiotic resistant genes such as *npt*II that confer resistance to kanamycin), so that only the transformed cells can be selected and regenerated on the selection medium containing the selection agent (e.g. kanamycin). The untransformed cells are eliminated on the selection medium as they are susceptible to the antibiotic (due to the absence of antibiotic resistance gene). The segregation and stability of the transgene integration and expression in the subsequent generations can be studied by genetic and molecular analyses (PCR, Southern, Northern and Western blot analysis). (Refer to earlier chapters for details of these techniques).

6.2.5. Transgenic plants with beneficial traits

During the last couple of decades, considerable progress has been made to understand the function of genes, isolation of novel genes and promoters as well as the utilization of these genes for the development of transgenic or genetically modified (GM) crops with improved and new characters. There are many potential applications of plant genetic engineering and some important examples of transgenic crops, approved by the U.S Food and Drug administration are summarised in **Table 2.**

Table 2. Few Examples of transgenic crops approved by the U.S Food and Drug administration

Gene(s) introduced	New/Improved Character	Developer
Thioesterase	High laurate oil Weed	Calgene
EPSP synthase PAT & Barnase/	Control Weed Control &	Monsanto
Barstar	Hybrid production	AgrEvo
EPSP synthase	Weed control	Monsanto Monsanto,
Bt CrylA(b)	Insect resistance	Ciba-Geigy Northrup King
Bt CrylA(c)	Insect Resistance	Dekalb Genetics
Acetolactate synthase	Weed Control	DuPont
Nitrilase	Weed Control	Calgene
EPSP synthase	Weed Control	Monsanto
Bt CrylA(c)	Insect Control	Monsanto
Coat protein	Virus Resistance	Univ. Hawaii & Cornell Univ.
Bt CryIIIA & Coat protein	Insect & virus control	Monsanto
EPSP synthase	Weed control	Monsanto
Gmfad2-1	Improved Oil	DuPont
Bt CrylA(c)	Insect control	Monsanto Calgene
Antisense PG	Delayed ripening	

Stress tolerance

Crop plants are very productive under ideal cultural conditions, but ideal growing conditions rarely occur. Moreover, plants encounter both biotic (viral, bacterial, fungal pathogens, nematodes, insect pests and weeds) and abiotic (salinity, drought, extreme temperatures, nutrient deficiency, etc.) stresses and these stresses cause a colossal loss of crop yield and quality. The application of chemical and biological pesticides as well as the use of resistant varieties is only partial success and these have certain limitations. Therefore, newer and effective technologies are essential to meet the demand. In this context, biotechnological strategies can be used to create transgenic plants with increased resistance to diseases and pests as well as abiotic stresses.

Biotic stress tolerance

Herbicide tolerance: Weeds (plants growing where they are not wanted, e.g. Striga) decrease crop yields and quality primarily by competing with crop plants for light, water and nutrients. Farmers apply herbicides/weedicides (e.g. glyphosate) for the eradication of weeds in the fields, but the main problem with this is the development of herbicide tolerance by weeds. Newer techniques, based on biotechnological tools, have been developed which are quite effective for weed management as well as in increasing the yields and income. There are several biotechnological strategies for weed control, but the most commonly employed approach is the over-production of herbicide target enzyme (usually in the chloroplast) in the plant, so that it becomes insensitive to the herbicide. The popular example for such an approach is the introduction of a modified gene from an Agrobacterium species that encodes for a resistant form of the herbicide target enzyme into crop plants for tolerance against the most extensively used herbicide glyphosate (sold as Roundup) and is effective against many weeds. Roundup Ready GM crop plants such as canola **(Fig. 9A)**, soybean, corn and cotton tolerant to glyphosate has already been commercialised.

Pest resistance: All crop plants are affected by a variety of insects and nematodes, and significantly reduce their yield and quality. To minimize these losses (both food and money), farmers use the synthetic pesticides extensively which cause severe effects on human health and environment. The transgenic technology provides an alternative and innovative method to improve pest control management which are eco-friendly, effective, sustainable and beneficial in terms of yield. The first genes available for genetic engineering of crop plants for pest resistance were *Cry* genes (popularly known as Bt genes) from a bacterium *Bacillus thuringiensis*.

These are specific to particular group of insect pests, and are not harmful to useful insects like butter flies and silk warms. Transgenic crops (e.g. cotton, rice, maize, potato, tomato, brinjal, cauliflower, cabbage, etc.) with Bt genes have been developed for and such transgenic varieties proved effective in controlling the insect pests and it has been claimed worldwide that it has led to significant increase in yields, and dramatic reduction in pesticides' use. Bt crops have already been commercialised in several countries. The most notable example is Bt cotton (which contains *CrylAc* gene) that is resistant to a notorious insect pest Bollworm and in the year 2002, Bt cotton was adopted in India **(Fig. 9B)**.

Disease resistance: Pathogens (viruses, fungi and bacteria) infect crop plants and drastically reduce their yield and quality. Globally, several diseases have caused havoc at several points of time in several countries. For instance, the great Irish famine resulted in great damage of potato crop due to the late blight disease caused by fungal pathogen. India has also experienced a famine in the form of Bengal famine due to destruction of rice caused by fungal pathogen. Thus, there is a great concern about the control of plant diseases. Traditionally, farmers apply chemical pesticides or use resistant crop varieties developed by the breeders, but these methods have certain limitations.The alternative and useful strategy is the creation of disease resistant transgenic crop plants by transfer of resistance genes from varied sources.

Virus resistance: There are several strategies for engineering plants for viral resistance, and these utilizes the genes from virus itself (e.g. the viral coat protein gene). The virus-derived resistance has given promising results in a number of crop plants such as tobacco, tomato, potato, alfalfa and papaya. Some viral resistance transgenic plants like papaya resistant to papaya ring spot virus **(Fig. 9C)** have been commercialised in some countries.

Fungi and bacteria: Plants respond to pathogens by inducing a variety of defence responses like pathogenesis-related proteins (PR proteins), enzymes that degrade/destroy fungal cell wall, antifungal proteins and compounds, phytoalexins, etc. Several transgenic crop plants showing increased resistance to fungal pathogens are being raised with genes coding for the different compounds mentioned above.

Some of these strategies, particularly the genes that encode for phytoalexins and cell wall degrading enzymes were also used for producing bacterial resistant transgenic plants.

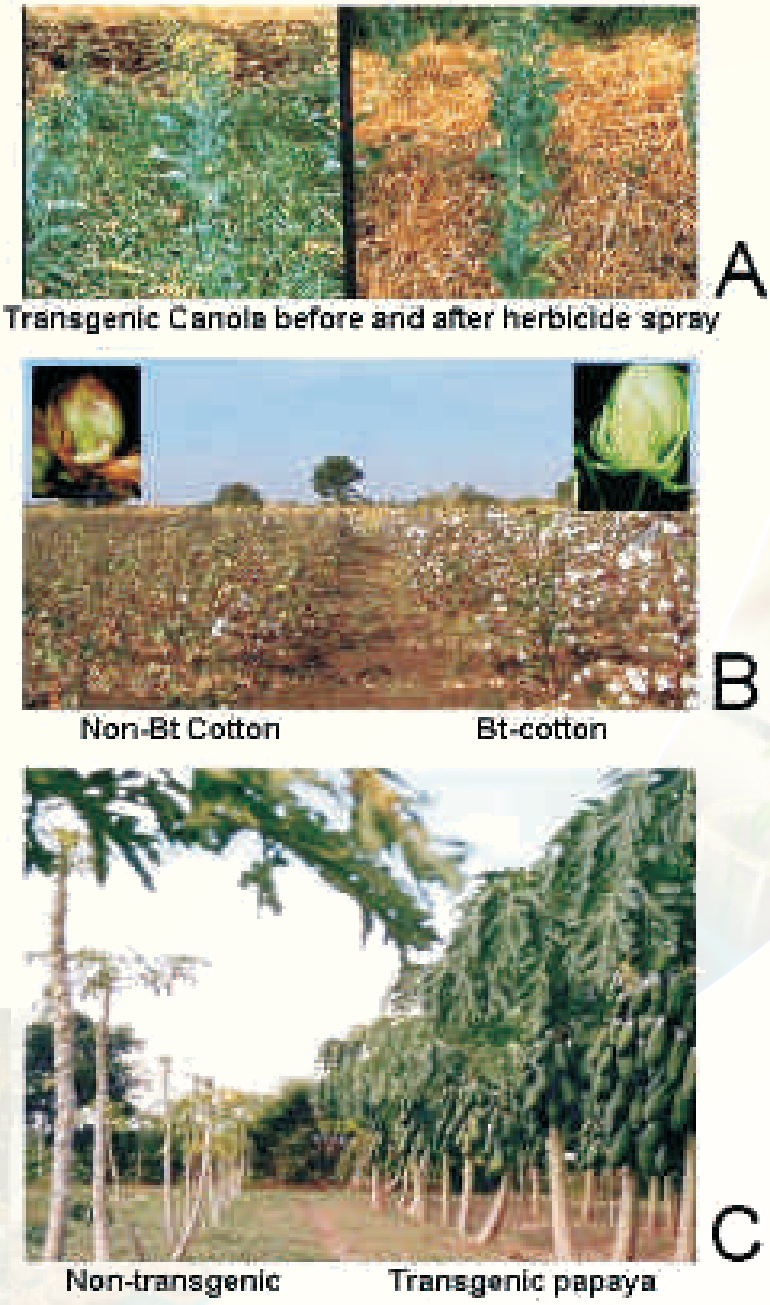


Fig.9. Genetically engineered crops with desirable traits. Transgenic canola for herbicide (glyphosate) tolerance before and after herbicide spray in the field (A); Bt cotton resistant to bollworm *(Helicoverpa armigera)* with healthy bolls (inserted photo) and non-Bt cotton with infested insect larvae (inserted photo) (B); Transgenic papaya resistant to papaya ringspot virus (c).

Abiotic stress tolerance

Plant growth and productivity are greatly affected by various environmental stresses/ abiotic stresses like high salinity and drought. Plant breeding efforts to produce abiotic stress tolerant plants while retaining high production is not very successful.

Plants have evolved many types of adaptations to cope with abiotic stress conditions like the production of the stress-related osmolytes like sugars (e.g. trihalose and fructans), sugar alcohols (e.g. mannitol) and amino acids (e.g. proline), glycine betaine, and certain proteins (e.g. antifreeze proteins). Transgenic plants have been developed which over-express the genes for one or more of the above mentioned compounds. Such plants have shown increased tolerance to environmental stresses.

Delayed fruit ripening

The gas hormone, ethylene is involved in the regulation of fruit ripening. Therefore, ripening can be slowed down by blocking or reducing ethylene production. This can be achieved by introducing ethylene forming gene(s) in a way that will suppress its own expression in the crop plant. Fruits from such plants ripen very slowly (however, they can be ripen by ethylene application) and are very important for export to longer distances without spoilage as they show longer-shelf life due to slow ripening. The notable example of this kind is the ***'Flavr Savr'*** transgenic tomatoes which were commercialized in U.S.

Male sterility

Male sterile plants are very important to prevent unnecessary pollination and to eliminate the process of emasculation during the production of hybrid plants. These are created by introducing a bacterial gene from *Bacillus amyloliquefaciens* that encode an enzyme **barnase**, which is an RNA hydrolyzing enzyme that inhibits pollen formation if, the expression of this gene specifically in the tapetal cells of anther using **tapetal-specific promoter** (e.g. TA29) to restrict its activity only to the cells involved in pollen production. Male fertility can be restored by introducing another gene from the same bacterium under the control of TA29, whose product **barstar** (protein) tightly bind with RNase, so that the normal pollen are formed. This **barnase/barstar** system was successfully utilized in producing male sterile/restorer lines in number of crops, particularly mustard for hybrid production.

Transgenic plants as bioreactors (Molecular farming)

Plants are amazing and cheap chemical factories that need only water, minerals, sun light and carbon dioxide to produce thousands of sophisticated chemical molecules with different structures. Given the right genes, plants can serve as bioreactors to modified or new compounds such as amino acids, proteins, vitamins, plastics, pharmaceuticals (peptides and proteins), drugs, enzymes for food industry and so on. Some of the potential and remarkable examples of this kind are described here.

Nutrient quality

Plants are the principle source of human nutrition. A small number of crop plants such as cereals, legumes, vegetables and root crops supply most of the energy and nutrients (in the form of fats, carbohydrates, proteins, vitamins and micronutrients) in the human diet. It is estimated that about 800 million people, mostly in developing countries are malnourished (particularly children) and suffering with several diseases due to the deficiency of nutrients. Therefore, the improvement of nutritional quality of crop plants is extremely important. Transgenic crops with improved nutritional quality have already been produced by introducing genes involved in the metabolism of vitamins, minerals and amino acids. Few examples of genetic modification of nutrition quality are described below.

Vitamin A: The source of vitamin A is either directly from animal food (e.g. milk, cheese and meat) or indirectly from green leafy vegetables (e.g. carrots, spinach, tomatoes, and chillies) and fruits (e.g. mango, melon and papaya), which contain carotenoids (beta-carotene - the provitamin A) that are converted to vitamin A in the body. Vitamin A deficiency can lead to night blindness, permanent blindness and skin disorder, among others. About 124 million children worldwide are deficient in vitamin A and a quarter of a million go blind each year due to vitamin A deficiency.

The staple food rice is extremely low in vitamin A, and therefore the improvement of vitamin A content is very important. In a remarkable example of genetic engineering, Prof. Ingo Potrykus and Dr. Peter Beyer developed genetically engineered rice (popularly known as **'Golden Rice'**), which is enriched in pro-vitamin A (beta-carotenoids) by introducing three genes involved in the biosynthetic pathway for carotenoid under the control of endosperm-specific promoter, so that gene products (enzymes) are synthesized in the rice endosperm. The seeds of Golden Rice are yellow in colour because of pro-vitamin A is produced in the entire grain **(Fig. 10A)**. Interestingly, they have also further engineered this Golden Rice by introducing three more genes from different organisms for iron source and its absorption. Similarly, transgenic crop plants are also being produced to raise the level of other vitamins, particularly vitamin E (which is an essential antioxidant in humans) and vitamin K (which is involved in bone formation).

Seed protein quality: The nutritional quality of cereals and legumes are limited because of deficiency of the essential amino acids, i.e. lysine in cereals, and methionine and tryptophan in pulses. Two genetic engineering approaches have been used to improve the seed protein quality. In the first case, a transgene (e.g. gene for protein containing sulphur rich amino acids was introduced into pea plant (which is deficient in methionine and cysteine, but rich in lysine) under the control of seed-specific promoter. In the second approach, the endogenous genes are modified, so as to increase the essential amino acids like lysine in the seed proteins of cereals.

Diagnostic and therapeutic proteins

Transgenic plants can also produce a variety of proteins used in diagnostics for detecting human diseases and therapeutics for curing human and animal diseases in large-scale with low cost. The monoclonal antibodies, blood plasma proteins, peptide hormones and cytokinins are being produced in transgenic plants and their parts such as tobacco (in leaves), potato (in tubers), sugarcane (in stems) and maize (in seed endosperm).

Edible vaccines

Crop plants offer a cost-effective bioreactor to express antigens which can be used as edible vaccines. Some of the antigenic determinants on the surface of viruses and bacteria are proteins made by the pathogen. The genes encoding these antigenic proteins can be isolated from the pathogens and expressed in plants, and such transgenic plants or their tissues producing antigens can be eaten for vaccination/immunisation **(edible vaccines)**. The expression of such antigenic proteins in crops like banana and tomato are useful for immunisation of humans since banana and tomato fruits can be eaten raw. In case of animals, such genes can be expressed in crops like alfalfa and other forage/fodder crops, which are suitable for vaccination. The edible vaccines that are produced in transgenic plants have great advantages like the alleviation of storage problems, easy delivery system by feeding and low cost as compared to the recombinant vaccines produced by bacterial fermentation. Vaccinating people against dreadful diseases like cholera and hepatitis B by feeding them banana/tomato, and vaccinating animals against important diseases such as foot and mouth disease by feeding them sugar beets could be a reality in the near future.

Biodegradable plastics

The biodegradable plastic, e.g. polyhydroxybutyrate (PHB) is being produced commercially by fermentation with the bacterium *Alcaligenes eutrophus*. The main drawback of bacterial PHB is its high production cost, making it substantially very expensive than synthetic plastics. Alternatively, transgenic plants can be used as factories to produce PHB. The genetically engineered Arabidopisis plants with the three genes involved in PHB synthesis from *A. eutrophus* produced PHB globules exclusively in their chloroplasts without effecting plant growth and development **(Fig. 10B)**. The large-scale production of PHB may be easily achieved in tree plants like populus, where PHB can be extracted from leaves. Industry has already begun to explore the production of biodegradable plastics from transgenic plants.

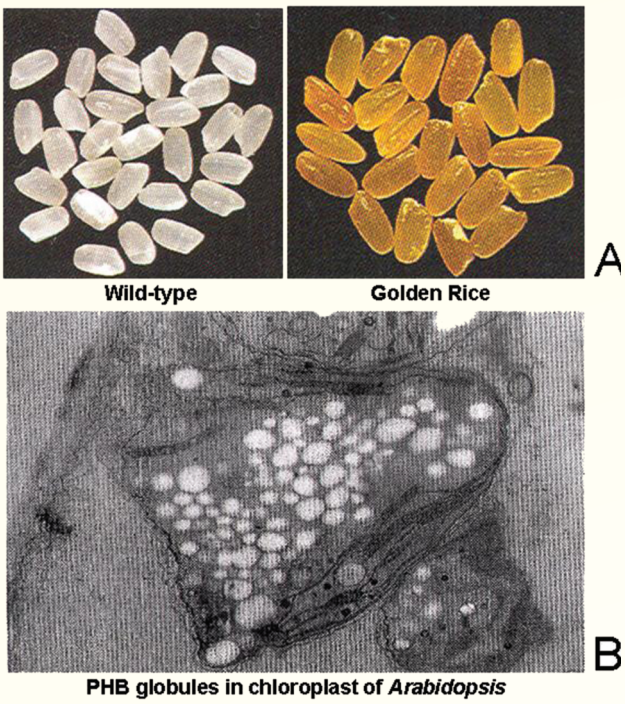


Fig. 10. Golden rice (A) and electron microscopic picture of chloroplast of transgenic Arabidopsis showing PHB globules (B).

Metabolic engineering and secondary products

Plant biotechnology will lead to improved plant sources for the production of valuable secondary metabolites mentioned in previous section on cell culture products. Biosynthetic pathway which lead to the production of such valuable products are being engineered for their over-production to reduce the cost. The over-expression of the gene, which encode for the first enzyme in a pathway generally results in higher levels of the desired end product, and this has been successfully done in the enhancement of taxol production from the transformed tissue cultures of *Taxus* sp. Another strategy involves the use of *Agrobacterium rhizogenes* to induce the excessive formation of secondary roots **(hairy roots)** in plants that normally produce useful secondary metabolites in this region.

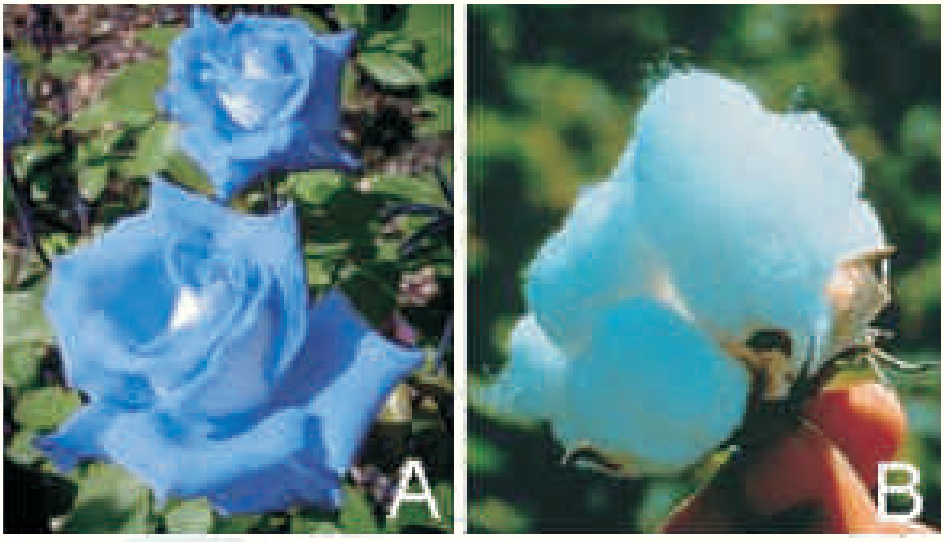
Other applications

There are many other applications of plant genetic engineering, which are listed below:

- Production of healthy oils with altered fatty acid profiles.
- Modification of starch properties for specific uses.
- Favourable change of grain storage products and their chemical composition to improve the processing of bread making with wheat flour, malting of barley and brewing of beer.

- Removal of undesirable toxic compounds in certain plants.
- Development of blue roses **(Fig. 11A)**, which is otherwise not possible by conventional plant breeding because of the absence of blue pigment in roses.
- Development of blue coloured cotton **(Fig. 11B)** and also with other colours.
- Development of tear-less onions, caffeine-free coffee and low nicotine tobacco.

Undoubtedly, there will be many more exciting applications of plant genetic engineering in the future.



Blue Rose **Blue Colore Cotton**

Fig. 12. Genetically modified blue rose (A) and blue coloured cotton (B).

6.2.6. Biosafety in Plant Genetic Engineering

The GM crops are fast becoming a part of agriculture throughout the world because of their contribution to the increased crop productivity and to global food, feed and fibre security, besides their use in health-care and industry. However, the constraints associated with public acceptance of transgenic crops continue to be important challenges facing the global community. The following are the major concerns about GM crops and GM foods:

- The safety of GM food for human and animal consumption (e.g. GM food may cause allergenicity).
- The effect of GM crops on biodiversity and environment.
- The effect of GM crops on non-target and beneficial insects/microbes.
- Transgenes may escape through pollen to related plant species (gene pollution) and may lead to the development of super weeds.

- The GM crops may change the fundamental vegetable nature of plants as the genes from animals (e.g. fish or mouse) are being introduced into crop plants.
- The antibiotic resistance marker genes used to produce transgenic crops may horizontally transfer into microbes and thus exacerbate problem of antibiotic resistance in human and animal pathogens (i.e. transgenes may move from plants to gut microflora of humans and animals).
- The GM crops may lead to the change in the evolutionary pattern.

Unfortunately, the public debate over the benefits and hazards of plant gene technology suffers from an astounding array of misinformation, misunderstanding and manipulation. In fact, with the continuing accumulation of evidence of safety and efficiency, and the complete absence of any evidence of harm to the humans and animals as well as the environment, more and more consumers are becoming comfortable because of the increased awareness about the potential of the plant genetic engineering. Further, the transgenic crops (e.g. cotton, tomato, corn and soybean) which have already entered the market place (which carry a label, i.e., GM crop or GM food in several countries were subjected for the extensive field trials for environmental safety related to wild species and competitive performance of transgenic and toxic effects as per the standard regulatory policies before they were approved for commercialisation. In fact, in 2009, a record of 14 million farmers from 25 countries cultivated 134 million hectares (330 million acres) with the crops that were genetically engineered for herbicide, insect resistance, delayed fruit ripening and improved oil quality. The global area of biotech crops planted from 1996 to 2009 is shown in **Fig. 12**.

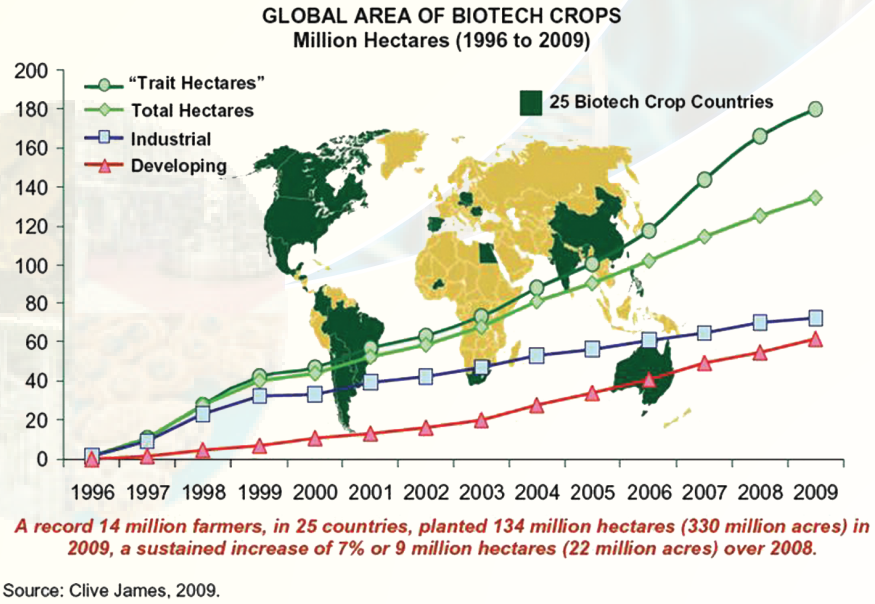


Fig. 12. Global area of biotech crops from 1996 to 2009 (in million hectares).

Nevertheless, the thorough assessment of the risks associated with transgenics for plants, animals and humans is important before they are released. Foods or food ingredients derived from GMOs must be shown to be as safe as or safer than their traditional counterpart before they can be recommended for public use. In general, many scientists believe that plant genetic engineering is the 'technology' that can solve various human problems, especially hunger and disease.

Review Questions

1. What is meant by plant tissue culture? List the various steps involved in this technique?
2. What is plant regeneration? Give the different pathways of plant regeneration.
3. Name a few useful secondary metabolites that can be produced through plant cell culture technology.
4. List the various applications of plant cell and tissue culture technology.
5. Differentiate:
   i) Cell culture and protoplast culture
   ii) Direct gene transfer and indirect gene transfer
6. Explain how embryo rescue can be used to produce novel hybrids.
7. Write short note on artificial seeds.
8. How are plant cells converted to protoplasts? Why are protoplasts more amenable to produce somatic cell hybrids and cybrids?
9. What is genetic engineering? Add a note on the various steps involved in this technology.
10. Why is *Agrobacterium* described as 'natural genetic engineer of plants'?
11. Explain how gene gun functions to deliver genes into plant cells?
12. What are the proposed benefits of genetic engineering in crop improvement?
13. What are the genetic engineering strategies to create the following traits in transgenic crops:
    a) Herbicide tolerance
    b) Abiotic stress tolerance
    c) Insect resistance
    d) Virus resistance
14. What is the common strategy to produce transgenic crops with delayed ripening and longer shelf life of fruits?

15. What is meant by 'Golden Rice'? In what way it is different from the normal rice?

16. Write briefly the benefits of biodegradable plastics that are produced from GM plants.

17. What are the transgenic crops that are commercialised globally?

18. Explain the social, economical and environmental implications of genetic engineering techniques.

19. Define:

Explant	Hairy roots
Clone	Roundup
Callus	PR proteins
Micropropagation	Bt genes
Encapsulation	Male sterility
Embryo rescue	*Flavr Savr*
Somatic hybrids	Molecular farming
Cybrids	PHB
In vitro gene bank	Edible vaccines
T-DNA	Metabolic engineering
Crown galls	

20. Fill in the blanks:

I) Rapid multiplication of plants by tissue culture techniques is referred to as _____.

ii) Gottlieb Haberlandt is known as father of _____.

iii) The most commonly employed gene transfer method in plants is _____.

References

1. Bhojwani SS & Razdan MK. 1996. Plant Tissue Culture: Theory and Practice (Revised Edition), Elsevier, Amsterdam.
2. Chrispeels MJ & Sadava DE. 2003. Plants, Genes and Crop Biotechnology (Second Edition), Jones and Bartlett Publishers, Inc., Boston.
3. Davey MR & Anthony P. 2010. Plant Cell Culture: Essential Methods, Wiley-Blackwell, Oxford.
4. Gamborg OL & Phillips GC.1996. Plant Cell, Tissue and Organ Culture, Narosa Publishing House, New Delhi.
5. Glick BR & Pasternak JJ. 1998. Molecular Biotechnology: Principles and Applications of Recombinant DNA (Second Edition), ASM Press, Washington, D.C.
6. Oksman-Caldentey K-M & Barz WH. 2002. Plant Biotechnology and Transgenic Plants, Marcel Dekker, Inc., New York.
7. Singh BD. 2006. Plant Biotechnology, Kalyani Publishers, Ludhiana.
8. Slater A, Scott N & Fowler M. 2003. Plant Biotechnology: The Genetic Manipulation of Plants, Oxford University Press, Oxford.

CHAPTER 3

ANIMAL CELL CULTURE AND APPLICATIONS

6.3.1. Introduction

Animal cells, just as plant cells, will continue to grow outside the living organism if supplied with the appropriate nutrients and growth conditions. The process of growing of cells under laboratory conditions is called **Cell Culture**. It is carried out ***in vitro*** ('within glass') as opposed to ***in vivo*** ('within the living'). The advantage and limitations of animal cell culture are given in **Table 1**. A homogenous population of cells derived from a single parental cell is called a **clone**. Therefore, all cells within a clonal population are genetically identical. The growth rate of animal cells is relatively slow and usually require 18 to 24 hour to divide. This makes the animal cell culture vulnerable to contamination, as a small number of bacteria would soon outgrow a larger population of animal cells. The animal cell culture became a routine laboratory technique in 1950s after **George Gay** established the first human cell line (HeLa) from cervix cancer that led to several important discoveries in medical sciences. The need for cell culture, especially at large scales, became apparent with the need for viral vaccines.

Table 1. Advantages and limitations of animal cell culture.

Advantages	Disadvantages
• Homogenous genetic population. • Controlled physico-chemical environment. • Easy to add genes (Transfection) or regulate protein levels (RNAi). • Available in adequate numbers to do chemical study. • Easy production of biopharmaceuticals. • No ethical clearance required. • Cost effective screening assays.	• Small size (high sensitive techniques to detect changes). • Scale-up is challenging. • May not represent *in vivo* phenotype/genotype.

Advent of recombinant DNA technologies in the 1970s, enabled the expression of mammalian genes in bacteria. However, soon it was realized that large proteins or glycoproteins of therapeutic value could not be produced in bacteria. This prompted the usage of animal cell lines for large-scale production of therapeutic proteins. The development of hybridoma technology allowed us to produce large quantities of monoclonal antibodies of immense diagnostic and therapeutic values. Further, recent advances in the area of stem cells, tissue engineering and gene therapy are likely to open new treatment regimens.

Before you learn about the details of animal cell culture technology, it is instructive to know some essential features of animal cell growth in culture.

6.3.2. Animal Cell Culture Techniques

Features of animal cell growth in culture

Animal cells can be grown in glass or plastic vessels, in the presence of nutritive media that need to be is periodically replenished. However, depending on the tissue they have been isolated from, they can be grown only for limited generations even in the best nutritive media. There is a mortality associated with all normal animal cells. Another important feature of animal cells is that they divide and fill the surface of the culture vessel and then stop growing. Relate this to what happens in the animal body. The infant animal grows only to adulthood and not any further. Cells comprising tissues and organs such as the liver grow only to a certain size after which they cease to grow. This phenomenon which occurs in the normal body is observed also in cell culture and is termed **"contact inhibition"**. This means that when cells grow and reach the walls of the container (i.e., reach confluency) they stop growing further. Another important feature of cell growth in culture is that their environment is different from that *in vivo*. These differences affect the adherence of cells to culture vessels, their shape and rate of proliferation. It is of interest to know that in culture, cancer cells appear very different from normal cells. Cancer cells loose contact inhibition and pile on each other due to uncontrolled growth and among other features, appear more rounded in shape. Such differences in growth patterns in normal versus cancer cells are utilized by Oncologists (cancer biologists) to determine whether tumors are cancerous or not using **`Colony formation assay'**. Let us now learn about various types of cell cultures and the technology associated with it.

Primary Cell Cultures

Cells are dissociated from the parental tissue (such as kidney, liver) by mechanical or enzymatic methods and maintained in suitable culture medium and vessels. The most frequently used enzymes for separating cells from a given tissue (dispersion) are crude preparations of trypsin and collagenase that cleave the proteinaceous cementing material between cells in a tissue. The maintenance of growth of such cells under laboratory conditions is known as primary cell culture.

The characteristics of cells in culture usually depend on their original source within the animal.

Cells can be grown as **adherent** (anchorage-dependent) or **suspension cultures** (anchorage-independent). Adherent cells are usually derived from tissues of organs such as kidney where they are not mobile and are embedded in connective tissue. They grow adhering to the cell culture vessel. On the other hand, suspension cells do not attach to the surface of the culture vessel. Virtually all suspension cultures are derived from cells of the blood system. This is because, these cells (e.g., lymphocytes) are also suspended in plasma *in vivo*. The drawbacks of primary culture are that they are time consuming and require the use of live animals or fresh tissue. There can be considerable variation from one preparation to another particularly if prepared by different people. These difficulties can be overcome by the use of **secondary cell cultures** or **cell lines**.

Secondary Cell Cultures and Cell Lines

Once the primary culture is subcultured, it is known as secondary culture or cell line. Subculturing or "splitting cells," is required to periodically provide fresh nutrients and growing space for continuously growing cell lines. The frequency of subculture or density of cells to be plated, depends on the characteristics of each cell type. If cells are split too frequently or at too low a density, the line may be lost. If cells are not split frequently enough, the cells may exhaust the medium and die. Sub-culturing involves: removing the growth media, washing the plate, disassociating the adhered cells, usually enzymatically (e.g. with trypsin), although some cells may be removed by repeated pipeting or gentle scraping), and diluting the cell suspension into fresh media. **Rous** and **Jones** were first to introduce proteolytic enzyme trypsin for the subculture of adherent cells. Such cultures may be called **secondary cultures**. Sometimes, certain cells of these secondary cell cultures can spontaneously become altered (transformed) and give rise to continuous cell lines which show immortality, as they can grow indefinitely without dying in culture. These cultures can contain mixed cell types or can consist predominantly of a single cell type.

Types of Cell Lines

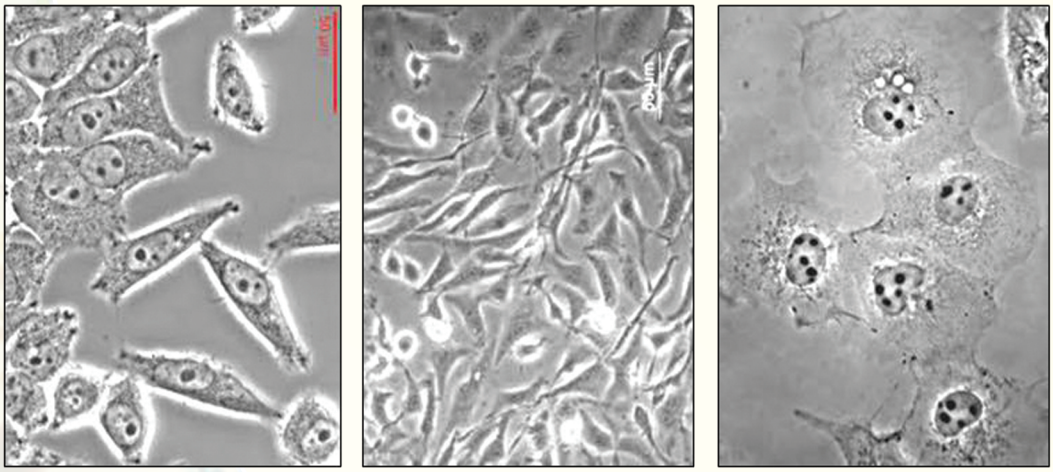
The various types of cell lines are categorized into two types, i.e., finite cell line and continuous cell line.

Finite Cell Lines

Finite cell lines are those cell lines which have a limited life span and they grow through a limited number of cell generations. Finite cell lines show the property of contact inhibition, density limitation and anchorage dependence. The mode of growth is in the monolayer form. The growth rate is slow and doubling time is around 24 to 96 hours.

Continuous Cell Lines

Cell lines transformed under *in vitro* culture conditions give rise to continuous cell lines **(Fig. 1)**. The various properties associated with continuous cell lines are: the **ploidy** (change in basic number of chromosomes), no contact inhibition and no anchorage dependence. The mode of growth is either monolayer or suspension form. The growth rate is rapid and doubling time is between 12 to 24 hours. The density limitation is reduced or lost.



HeLa cell line from human cervical carcinoma

CHO continuous cell line from Chinese hamster ovary

Cos-1 cell line from monkey kidney

Fig. 1. Different mammalian cell lines.

Physical environment for culturing Animal Cells

The culturing of animal cells under *in vitro* condition involves creation of appropriate physical, nutritional and hormonal environments in which the cells can grow. The physical environment includes controlling the temperature, pH, osmolality and gaseous environment by providing a supporting surface and protecting the cells from chemical, physical and mechanical stresses.

Temperature

The mammalian cells are grown in incubators maintained at 37°C. This temperature is chosen because it is the core body temperature of *Homo sapiens*. Further, it has been observed that most cells derived from the warm blooded animals will grow at this temperature.

pH

The regulation of extra-cellular and intra-cellular pH is essential for the survival of mammalian cells. The correct pH is not only important for maintaining the appropriate ion balance but also for maintaining optimal function of cellular enzymes and for optimal binding of hormones and

growth factors to cell surface receptors. Even transient changes in pH can alter cell metabolism which can lead to cell death. Most media strive to achieve and maintain the pH between 7 and 7.4. The regulation of pH is done using a variety of buffering systems. Most media use the bicarbonate - CO_2 buffering system. The interaction of CO_2 derived from cells or atmosphere with water leads to a drop in pH described by the equation below:

$$H_2O + CO_2 \leftrightarrows H_2CO_3 \leftrightarrows H^+ + HCO_3^-$$

The bicarbonate content of the medium neutralizes the effect of increased CO_2 according to the following equation: $NAHCO_3 \rightarrow Na^+ + HCO_3^-$. The increased HCO_3^- ion drives the equation above to the left until the equilibrium is reached at pH 7.4. This kind of system is called an open system.

Osmolality

The osmolality of the culture medium also has a significant bearing on cell growth and function. It preserves the membrane integrity of cells. If the outside osmotic pressure becomes higher or lower than that which must be maintained inside the cell, it will shrink or swell accordingly. The osmolality of the medium used is determined by the media formulation. Salt and glucose are the major contributors to the osmolality of the medium, although amino acids may also contribute significantly. Almost all commercial media are formulated to have a final osmolality of around 300 mOsm. Osmolality can be checked directly with an osmometer.

Medium

The most commonly varied factor in culture systems is the growth medium. Medium is a mixture of inorganic salts and other nutrients capable of sustaining cell survival *in vitro*. Having the correct nutrient mixture can often be the determining factor in failure or success in cell culture. The medium provides essential nutrients that are incorporated into dividing cells, such as, amino acids, fatty acids, sugars, ions, trace elements, vitamins, cofactors, and ions necessary to maintain the proper chemical environment for the cell. Some components may perform both roles; for example, the sodium bicarbonate may be used as a carbonate source but also may play an important role in maintaining the appropriate pH and osmolality. All media contain an energy source, usually glucose. Many of the media contain phenol red as a pH indicator. This is very helpful in monitoring the pH of the culture medium in an incubator. Highly acidic conditions turn the phenol red into yellow while highly alkaline conditions turns the phenol red into pink color.

Serum and antibiotics

Serum is one of the most important components of animal cell culture, as it supports cell proliferation and their attachment to culture vessels. The peptide hormones or hormone-like growth factors that promote healthy growth are often derived from animal blood, such as **foetal**

bovine serum (FBS). Serum is also a source of various amino acids, hormones, lipids, vitamins, polyamines and salts containing ions such as calcium, chloride, ferrous, ferric, potassium etc. Current practice is to minimize using blood-based supplements and switch to **serum-free medium** due to some complications of FBS usage. Although not required for cell growth, **antibiotics** such as penicillin and streptomycin are often used in culture medium to control the growth of bacterial and fungal contaminants.

Vessels and Equipments required for Animal Cell Culture

Cultures should be examined daily for their morphology, colour of the medium and density of cells. The animal cells are usually grown and maintained in Petri dishes, Culture flasks or Multi-well plates of various shapes and sizes **(Fig. 2)** at an appropriate temperature and gas mixture (typically, 37°C, 5% CO_2 for mammalian cells) in an incubator. Culture conditions vary widely for each cell type, and variation of conditions for a particular cell type can result in different phenotypes being expressed.

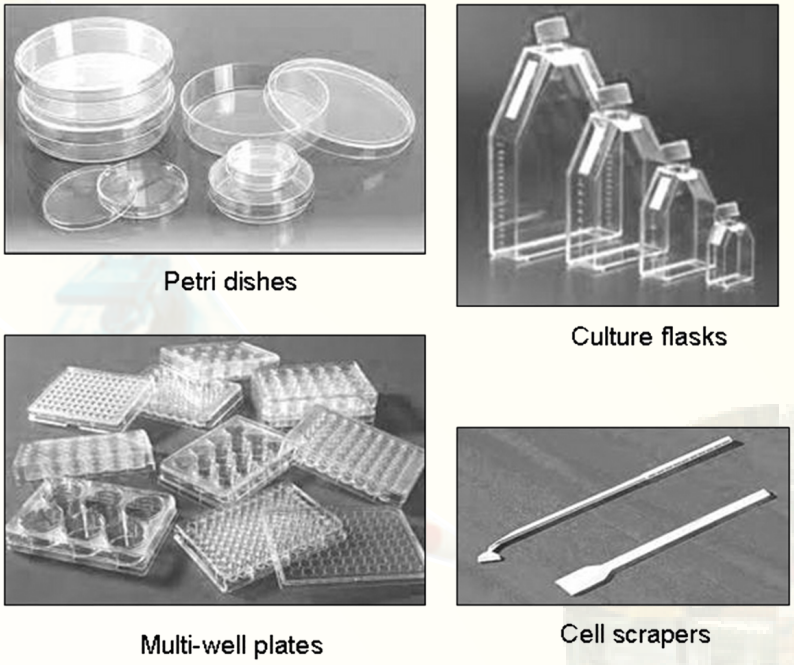


Fig. 2. Vessels and accessories for animal cell culture.

A cell culture laboratory should have equipments like tissue culture hood, CO_2 incubator, inverted microscope, Centrifuge etc. for doing animal cell culture work.

Tissue Culture Hood

All tissue/cell culture manipulations must be performed asceptically, i.e., without any bacterial or fungal contamination. Otherwise, animal cell culture media can easily get contaminated with

bacteria or fungi which will outgrow animal cells. The Laminar Air Flow (LAF) hoods allow the work area to be free of such contamination. A LAF hood essentially performs two functions:

1. Protects the tissue culture from the operator (by providing a sterile environment).
2. Protects the operator from the tissue culture (from possible infection risk).

Depending on the nature of cells/tissue being handled (especially infective agents), the biology safety cabinets are designated as Class I to class III. The LAF hoods have continuous displacement of air that passes through a high efficiency particle air **(HEPA)** filter that removes particulates from the air. The hoods are equipped with a short-wave UV light source that can be turned on for a few minutes to sterilize the surfaces of the hood just before use.

CO_2 Incubator

The CO_2 incubator is designed to reproduce as closely as possible the environmental conditions of the living cells. The essential functions of the incubator are to maintain, the sterility of the chamber, a constant temperature, an atmosphere with a fixed level of CO_2 and high relative humidity. A pan of water is kept at all times in the incubator chamber to maintain high relative humidity and prevent desiccation of the culture medium and maintain the correct osmolarity **(Fig. 3A)**. The animal cells are grown in an atmosphere of 5-10% CO_2 because the medium used is buffered with sodium bicarbonate/carbonic acid and the pH must be strictly maintained.

A. CO_2 Incubator B. Centrifuge C. Inverted Microscope

Fig. 3. Equipments for animal cell culture.

Centrifuge

For most cell culture only low-speed centrifuges are required **(Fig. 3B)**. A gentle braking action helps prevent disruption of the separated bands of cells. In most cases cells should be centrifuged at 20°C; nevertheless low operation temperature is useful to avoid exposing cells to uncontrolled higher temperatures.

Inverted Microscope

In tissue culture vessels, for example a petri dish, the cells are present at the bottom with the culture medium above. The inverted microscope allows the cells at the bottom to be visualized because the optical system is at the bottom with the light source on top **(Fig. 3C)**. Observation of cultures in this way will give an immediate idea of the health and growth of cells. Microscopes should be kept covered to protect from dust and the lights turned down when not in use.

6.3.3. Characterization of Cell Lines

In order to analyze the growth characteristics of a particular cell type or cell line, a growth curve can be established from which one can obtain a population doubling time, a lag time, and a saturation density. A growth curve generally will show the cell population's lag phase, that is, the time it takes for the cells to recover from subculture, attach, and spread; the log phase, in which the cell number begins to increase exponentially and a plateau phase, in which the growth rate slows or stops due to depletion of growth factors and nutrients. An increase in cell number is also a frequently used method of assessing the effect of hormones, nutrients, and so forth on a specific cell type. The culture doubling time, allows prediction of the likely cell concentration at any time in the future **(Fig. 4)**.

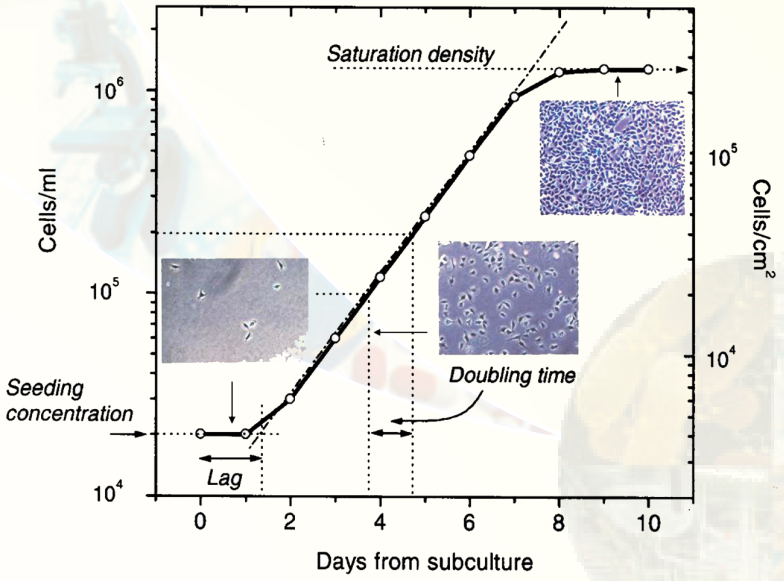


Fig. 4. Normal growth curve of animal cells.

There are laboratories and institutions which maintain various cell lines for scientists to use (for example American Type Culture Collection or ATCC, Virginia, USA). The advantage of using established cell lines is that their growth characteristics, media requirements and responses to selected reagents are established and therefore, convenient for the scientist to use. However, it

is important to check the identity of any newly acquired cell line for species of origin, tissue of origin and the maintenance of specific properties.

Once a cell line has been established, its karyotype has to be determined. This will confirm the species of origin, at least for those cells karyotyped, and determine the extent of gross chromosomal changes in the line. Karyotypes may vary from being near normal (i.e., the vast majority of cells in the culture have normal karyotypes) to being aneuploid. While a normal karyotype is desirable, the presence of an abnormal karyotype does not preclude using the cells for *in vitro* studies, especially if it has been demonstrated that the cells retain the normal function expected of them. If a normal karyotype is required (e.g., cell lines to be used to create transgenic animals), then special care must be taken in handling the cells to minimize chromosomal changes. The stability of the karyotype depends on the species from which the cell lines was derived, the growth conditions used, the way in which the cells are sub-cultured and whether or not the cells are frozen.

Storage and revival of cells

Liquid nitrogen is used for storing cells at very low temperature (-180°C to -196°C). Cells are stored at liquid nitrogen temperatures because the growth of ice crystals is retarded below -130°C. Freezing can be lethal to cells due to the effects of damage by ice crystals, alterations in electrolyte concentration, dehydration and changes in pH. To minimize the effects of freezing, a **cryoprotective agent** such as glycerol or DMSO, is added. A typical freezing medium is 90% serum, 10% DMSO. Further, it is desirable to use healthy cells that are growing in log phase and to replace the medium 24 hours before freezing. Also, the cells should be slowly cooled from room temperature to -80°C to allow the water to move out of the cells before it freezes.

Revival

Frozen cells should be thawed as rapidly as possible by placing the vial containing them into a 37°C water bath with moderate shaking. This is to minimize ice crystal formation, which may damage cells. After thawing the cells can be transferred directly into a tissue culture vessel containing suitable media for further growth.

6.3.4. Methods of Gene Delivery into Cells

A number of methods have been developed over the years for efficient transfer of genes in cell culture. Some common methods of plasmid **DNA transfection** are outlined below:

Calcium phosphate

Here HEPES-buffered saline solution is mixed with a calcium chloride solution containing DNA for transfection to form a fine precipitate of calcium phosphate with DNA. The suspension of the precipitate is then added to the monolayer of cells. The cells take up the calcium-phosphate- DNA complexes by endocytosis and express genes.

Lipofection

In this case, gene is transferred with the help of tiny vesicles of bipolar phospholipids that fuse with the cell membrane, releasing the DNA into the cytoplasm.

Microinjection

It is the most efficient method of gene delivery into cells. Here, DNA is directly injected into the nucleus using a fine glass capillary under a microscope. However this method acquire a great effort as each and every cell has to be injected individually.

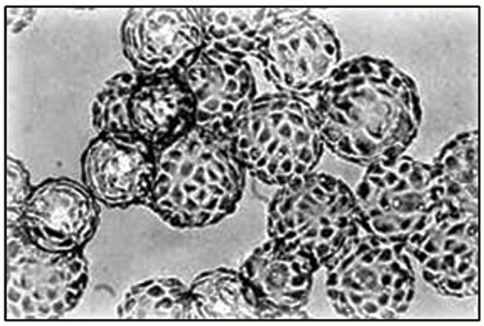
Electroporation

In this method, cells are mixed with the DNA and placed in a small chamber with electrodes connected to a specialized power supply. A brief electric pulse is applied, which is thought to 'punch holes' in the cell membrane, enabling the cell to take up DNA.

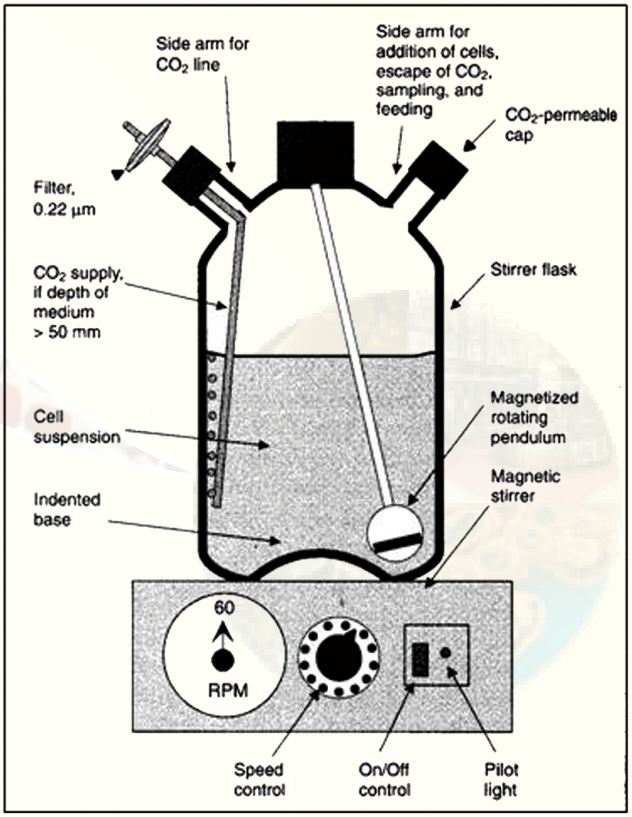
6.3.5. Scale-up of Animal Culture Process

The various scale-up methods include roller bottles with micro carrier beads for adherent cell cultures and spinner flasks for suspension cultures **(Fig. 5).**

A. Roller bottle



B. Microcarrier beads



C. Spinner cell culture

Fig. 5. Tools for scale-up of animal culture process.

Roller bottles

In roller bottles, the cells adhere to the total curved surface area of the micro carrier beads, thereby markedly increasing the available space for growth. These tissue culture bottles can be used in specialized CO_2 incubators with attachments that rotate the bottles along the long axis. After each complete rotation of the bottle, the entire cell monolayer has transiently been exposed to the medium. The volume of medium need only be sufficient to provide a shallow covering over the monolayer **(Fig. 5A)**.

Micro carrier beads

These beads are used to increase the number of adherent cells per flask and are either dextran or glass-based and come in a range of densities and sizes. The beads are buoyant and therefore, can be used with spinner culture flasks. The surface area available for cell growth on these beads is huge **(Fig. 5B)**. Microcarrier beads when re-suspended at the recommended concentration provide an area of 0.24 m^2 for every 100 ml of culture flasks. Under these conditions, adherent cells can be grown to very high densities before crowding becomes a problem. Cells growing at such high densities will rapidly exhaust the medium, which may need replacing the medium during culture.

Spinner cultures

Spinner cultures are used for scaling up the production of suspension cells. They consist of a flat surface glass flask with a suspended central teflon paddle that turns and agitates the medium when placed on a magnetic stirrer. Commercial versions incorporate one or more side arms for sampling and/or decantation. The cells are not allowed to settle to the bottom of the flask and thus cell crowding occurs only at very high densities. Stirring the medium improves gas exchange **(Fig. 5C)**.

6.3.6. Applications of Animal Cell culture

Several medically important protein pharmacueticals have been produced using animal cell culture and recombinant DNA technology. The important ones are listed in **Table 2.**

Table:2

Proteins	Animal Cell Line used	Therapeutic use
Erythropoietin (EPO)	CHO cells	Anemia
Factor VIII	CHO cells	Hemophilia A
Factor IX	CHO cells	Hemophilia B
Follicle Stimulating Hormone (FSH)	CHO cells	Infertility
Human Growth Hormone (hGH)	CHO cells	GH deficiency
Interleukin 2 (IL2)	CHO cells	Cancer therapy
Tissue Plasminogen Activator (t-PA)	CHO cells	Stroke
Monoclonal antibodies (mAbs)	Hybridoma cells	Cancer therapy & Autoimmune diseases

Erythropoietin

Erythropoietin (EPO) is a glycoprotein hormone that is involved in RBC production (erythropoiesis) and wound healing. EPO stimulates the bone marrow to produce more red cells and thereby increase the oxygen-carrying capacity of the blood. It is produced in response to hypoxia (shortage of oxygen) or anoxia (lack of oxygen) caused by anaemia. EPO is useful in the treatment of certain types of anemia such as anemia due to cancer, chronic renal failure and treatment of AIDS. Recombinant human EPO (r-HuEPO) has been produced using Chinese Hamster Ovary (CHO) cell lines. The use of r-HuEPO is advantageous over blood transfusion as it does not require donors or transfusion facilities, and there is no risk of transfusion-associated disease.

Factor VIII

Haemophilia A is a common heritable genetic disorder where the body lacks the ability to produce Factor VIII required for blood clotting. Like EPO, factor VIII is also a glycoprotein and has been produced in CHO cells due to its large structure.

Factor IX

Hemophilia B or Christmas disease is the second most common type of bleeding disorder due to deficiency of factor IX. Recombinant Factor IX produced in CHO cells is used to treat haemophilia B.

Tissue Plasminogen Activator (tPA)

tPA is a serine protease that catalyzes the conversion of plasminogen to plasmin which is responsible for dissolving blood clots. It is approved for use in certain patients having a heart attack or stroke. tPA is the first drug to be produced through mammalian cell culture **(Fig. 6)**.

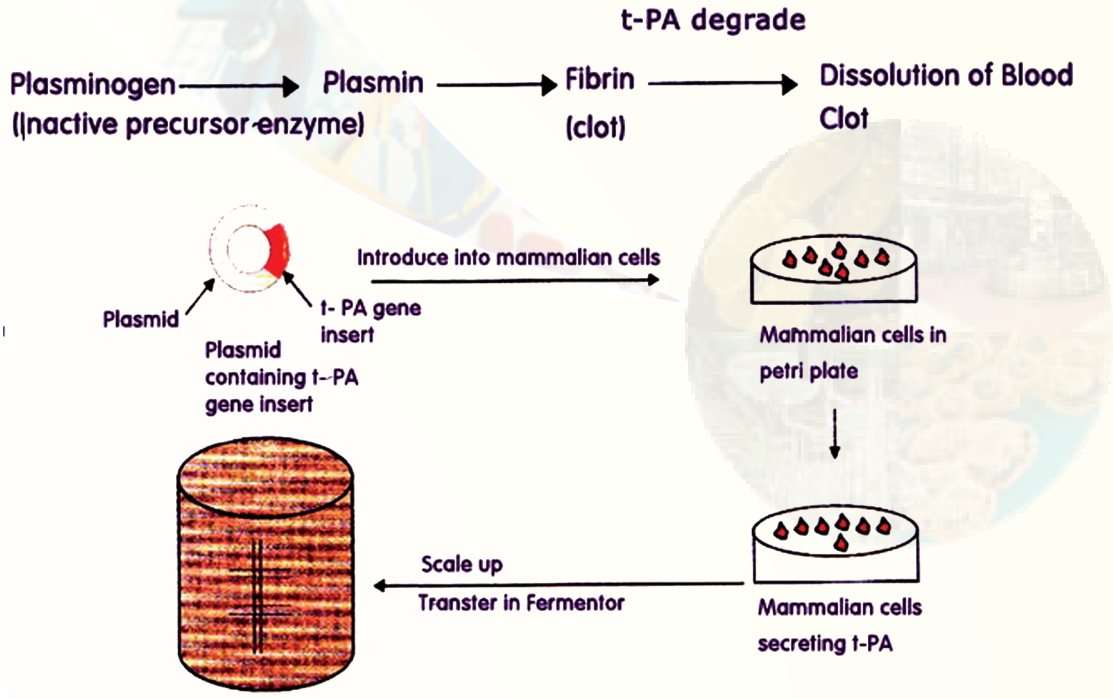


Fig. 6. Production and mode of action of tPA.

Hybridoma Technology for Monoclonal Antibody Production

It has been observed that antibodies bind to specific domains of macromolecules **(antigens)** known as **epitopes**. Antibodies present in serum are a heterologous population released by different populations of B-lymphocytes and therefore are known as polyclonal antibodies. Monoclonal antibodies (mAbs), on the other hand bind specifically to an epitope on an antigen and therefore are useful in detecting specific antigens **(diagnostics)** or blocking their binding by other molecules. mAbs are produced by antigen-activated B lymphocytes that have been immortalised by hybridising (fusing) them with a myeloma cell (cancerous lymphocyte). **Ceasar Milstein** and **George Kohler** (Nobel Prize winners) developed **hybridoma technology** by fusing antibody producing B cells with myeloma cells using polyethylene glycol. The hybrid cells retain the ability of B cells to secrete antibody and the ability of myeloma cells to grow indefinitely. The hybrid clones when grown in culture produces epitope-specific mAb **(Fig. 7)**. This technology has revolutionised the area of diagnostics and antibody-based therapies. The availability of monoclonal antibodies has helped in early detection of many infectious diseases like hepatitis and AIDS.

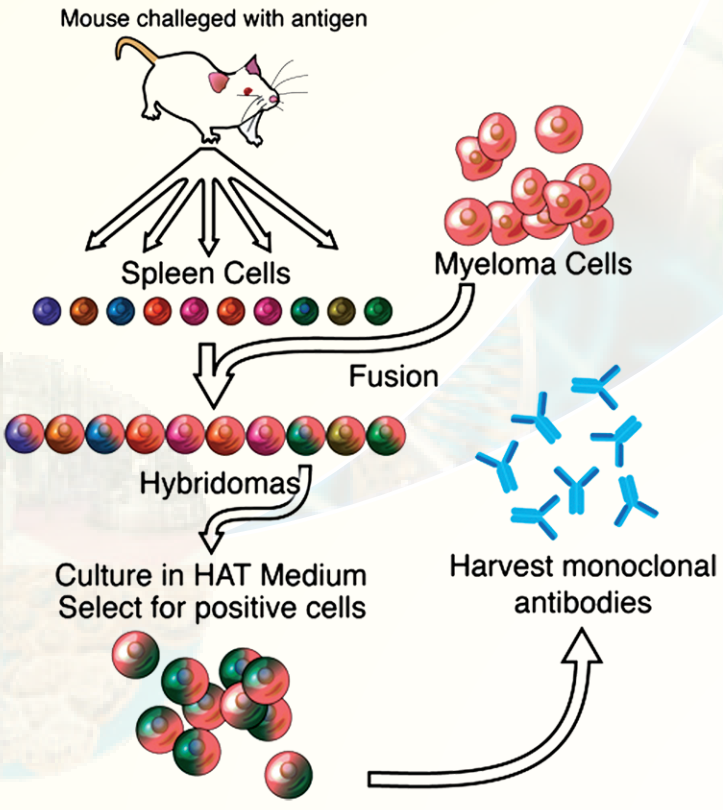


Fig. 7. Production of monoclonal antibodies.

Therapeutic mAb - OKT3

OKT-3 is monab-CD3, an immunosuppressant drug given intravenously to reverse acute rejection of transplanted organs such as heart, kidney and liver. OKT3 is the first monoclonal antibody to be used for the treatment of patients. OKT3 acts by blocking the function of T cells which play a major role in acute graft rejection **(Fig. 8)**. OKT3 binds and blocks the function of a cell surface molecule called CD3 in T cells. The binding of OKT3 to T cells results is followed by blocking of their functions. After OKT3 therapy is over, T cell function usually returns to normal within a week.

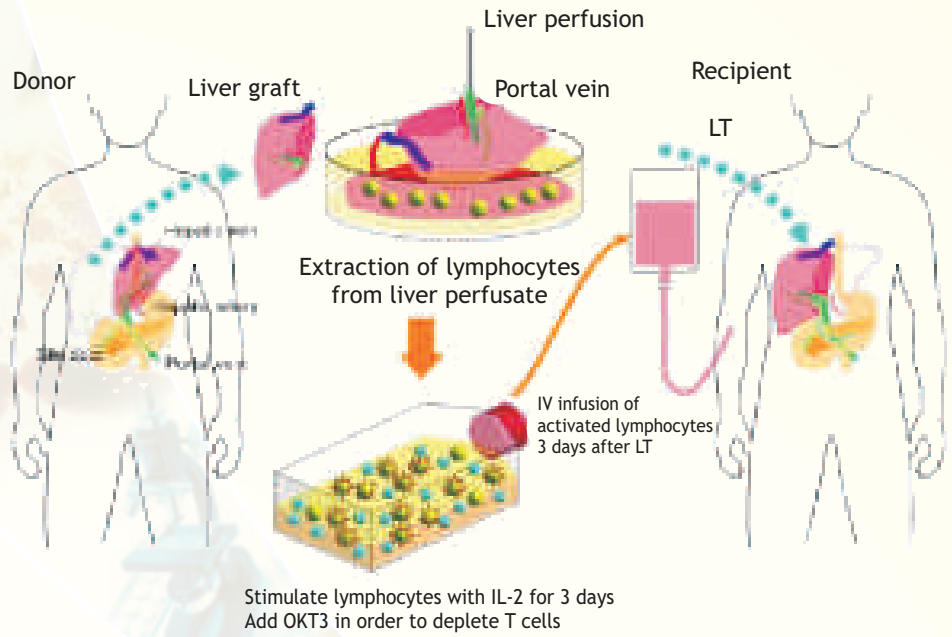


Fig. 8. Adoptive immunotherapy with OKT3 during liver transplantation (LT).

Therapeutic mAb - Herceptin

Herceptin (trastuzumab) is a monoclonal antibody approved for therapy of early-stage breast cancer that is Human Epidermal growth factor Receptor 2-positive (HER2+). These cell surface receptors receive signals that help cells to grow and multiply. Herceptin works by attaching itself to HER2 receptors by blocking them from receiving growth signals. The result is impaired growth of breast cancer.

6.3.7. Stem Cell Technology

Stem cells are characterized by their ability to renew themselves through mitotic cell division and differentiate into a diverse range of specialized cell types. Stem cells are found in all multi cellular organisms. Stem cells are like good shares in the stock market which can either be multiplied **(self renewal)** by getting bonus shares or sold to buy goods **(differentiate)**. Tissues like skin, blood and intestinal epithelium are subject to continuous renewal throughout life and

must maintain an adequate number of cells (stem cells) that retain the potential to proliferate to make good such losses. The most well studied process has been the formation of blood cells (haematopoiesis). It was known in case of mouse that haematopoiesis occurs in the spleen and bone marrow. In human being about 100,000 haematopoietic stem cells produce one billion RBC, one billion platelets, one million T cells, one million B cells per Kg body weight per day. The field of stem cell research was established in 1960s by **Ernest McCulloch** and **James Till** at the University of Toronto.

The two broad types of mammalian stem cells are: **embryonic stem (ES) cells** that are isolated from the inner cell mass of blastocysts, and **adult stem cells** that are found in adult tissues. The ES cells are **pleuripotent** and can differentiate into all types of specialized tissues **(Fig. 9)**. The adult stem cells are **multipotent** (lineage restricted) and act as a repair system for the body by maintaining the normal turnover of regenerative organs, such as, blood, skin, or intestinal tissues **(Fig. 10)**. Stem cells are now routinely grown and transformed into specialized cells such as, muscles or nerves through cell culture and used in medical therapies. The stem cells are useful in many medical conditions where cells are either dead or injured or abnormal, such as:

- Leukemia (cancerous blood cells).
- Heart disease, heart attack (cardiac tissue damage).
- Paralysis (spinal cord injury).
- Alzheimer's, Parkinson's, Huntington's (dead brain cells).
- Burns (damaged skin cells).

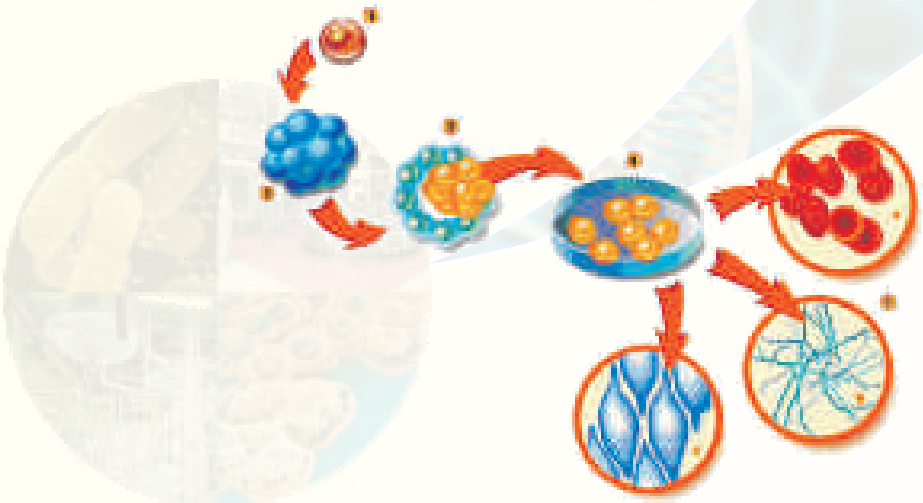


Fig. 9. Cultivation of Embryonic Stem cells. 1, *In vitro* fertilized eggs; 2, Morula; 3, Blastula with inner stem cell mass; 4, Cultured undifferentiated stem cells; 5, Differentiated cells - (a) blood, (b) neural, and (c) muscle cells

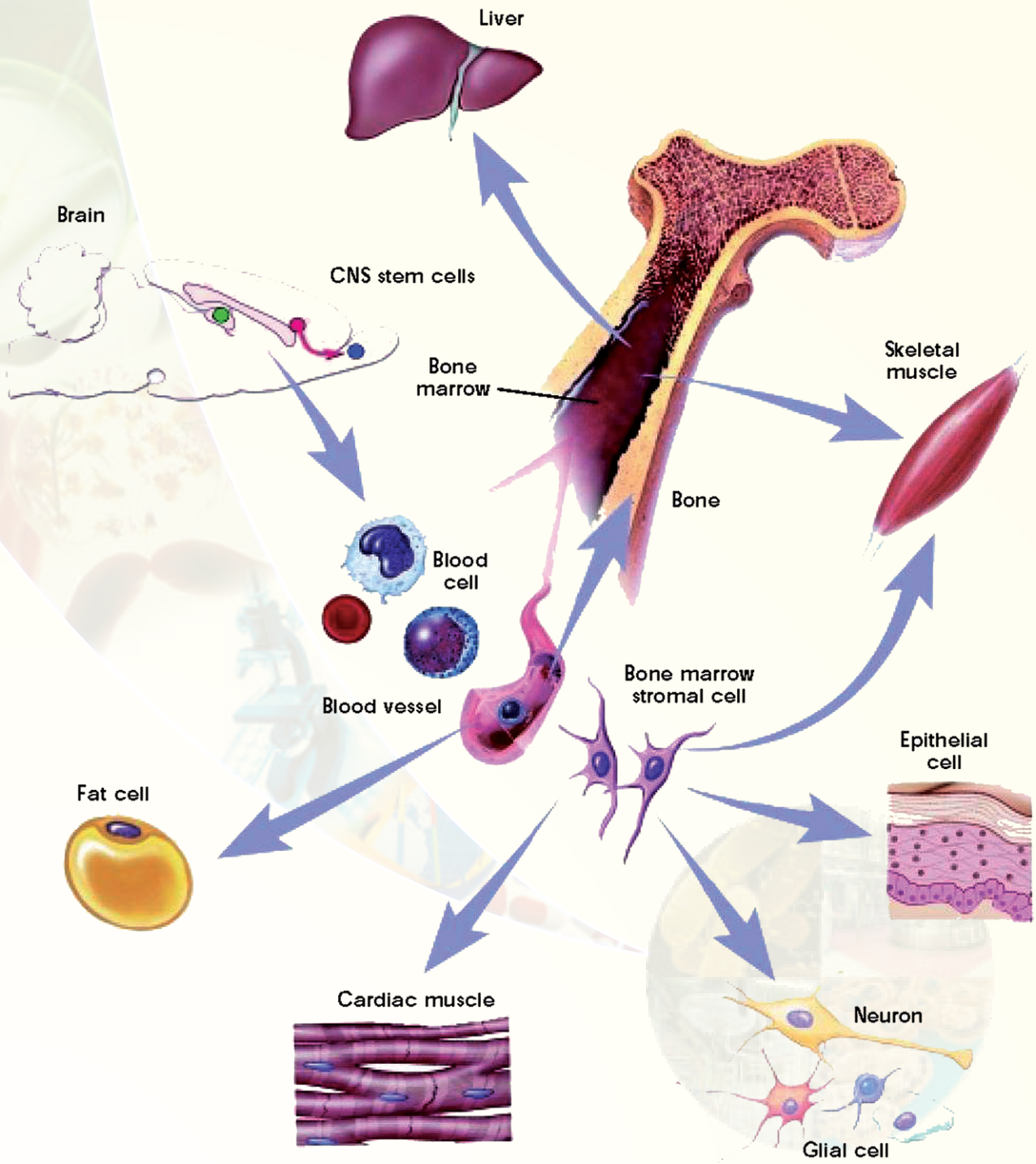


Fig. 10. Cultivation of adult stem cells from bone marrow and their differentiation into specialized cells.

ES Cell culture and its applications

The ES cells are cell lines derived directly from the inner cell mass of growing embryos without use of immortalising or transforming agents. The **inner cell mass** (ICM) of embryos can be maintained in cell culture in the presence of irradiated fibroblast cells. The stem cells:

a. retain the characteristics of founder cells, even after prolonged culture and extensive manipulation.

b. reintegrate fully into embryogenesis if transferred.

c. could be used to create chimeric mice by taking ES cells from a black mouse and implant it into the embryo of an albino mouse (white). The progeny so developed had skin color of black and white (a chimera **Fig. 11**).

d. could maintain a stable euploid karyotype.

e. could self renew without differentiating in culture.

Now it is possible to selectively remove a gene **(gene knock outs)** and make other precise genetic modifications in the mouse ES cells and create mouse models of human diseases. Such mouse models have been extremely useful not only in understanding the genetic basis of a disease but also in search for new diagnostic and therapeutic modalities.

In 1998, **James Thomson** developed a technique to isolate and grow human ES cells in culture. The human ES cells can be derived from the inner cell mass of blastocyst or from human germ cells before they initiate meiosis and cultured in a petri dish. Specialised cells can be grown in the presence of specific growth factors such as fibroblast growth factor and platelet-derived growth factor. The human ES cells have opened new possibilities for stem cell therapy in clinics.

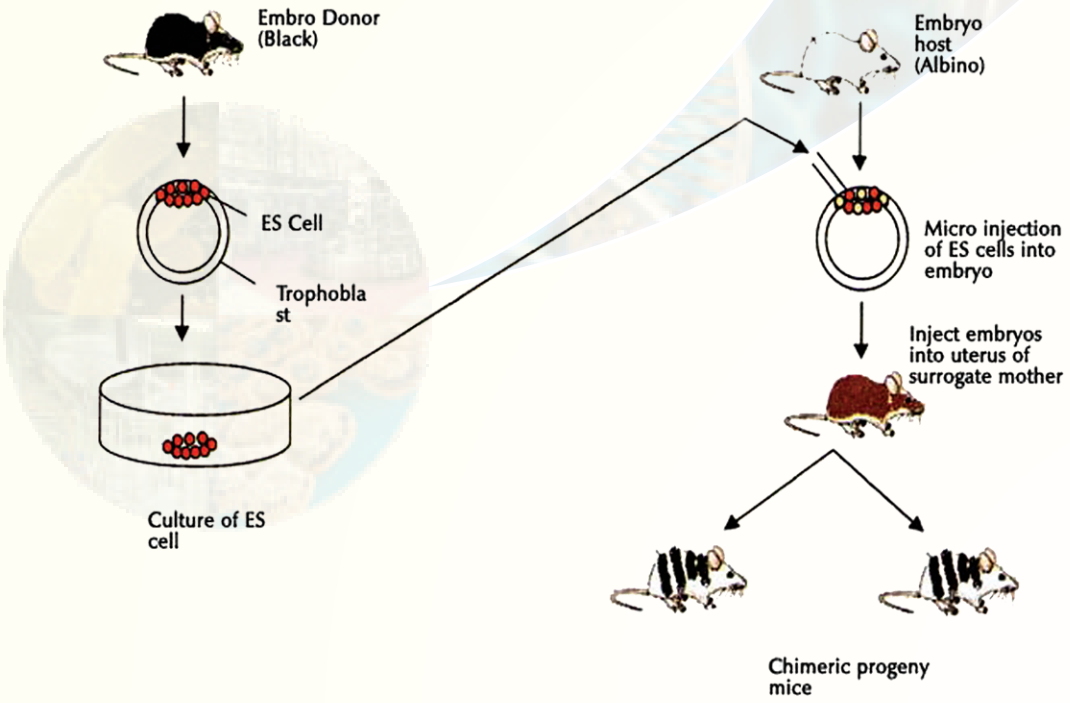


Fig. 11. Creation of chimeric mice using ES cells

6.3.8. Tissue engineering

Recent advances in the fields of cell biology, biomedical engineering and materials science have given rise to the inter-disciplinary field of tissue engineering. The aim of tissue engineering is to supply body parts for repair of damaged tissue and organs, without causing an immune response or infection or mutilating other parts of the body. Tissue engineering potentially offers dramatic improvements in low-cost medical care for hundreds of thousands of patients annually. Large-scale culturing of human or animal cells-including skin, muscle, cartilage, bone, marrow, endothelial and stem cells-may provide substitutes to replace damaged components in humans. Naturally derived or synthetic materials may be engineered into "scaffolds" that when implanted in the body could provide a template that allows the body's own cells to grow and form new tissues **(Fig. 12)**. Such implants could function like neo-organs in patients without triggering immune responses. **Genetically-modified animals** may also provide a source of cells, tissues, and organs for xenografts.

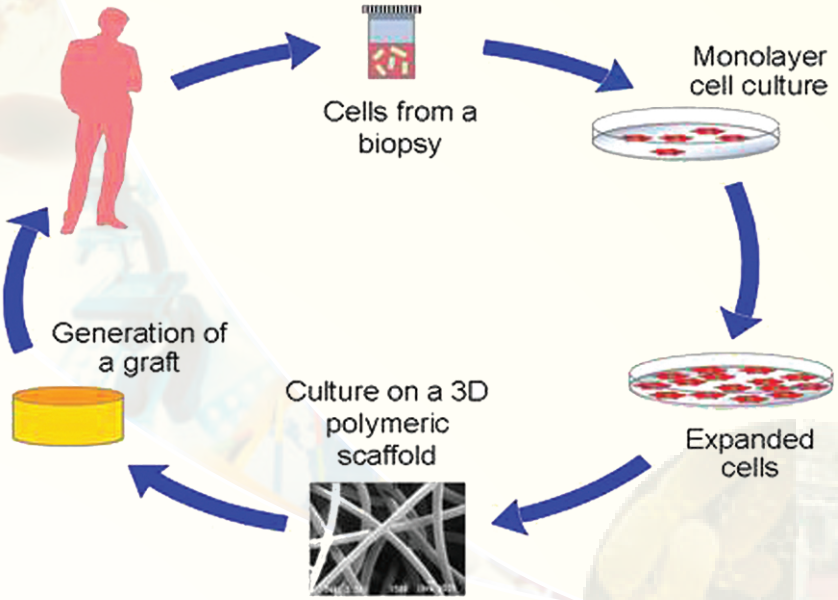


Fig. 12. Basic principle of Tissue engineering.

Review Questions

1. Name two important products from Animal Cell Culture Technology. What are their functions?
2. Write two important features of cultured animal cells. Differentiate between primary and secondary cell cultures.
3. What are cell lines? How the growth characteristics of cell lines determined?

4. What is the importance of pH while culturing animal cells? How is the pH maintained in culture media?
5. What are some of the characteristic features of normal and transformed cells?
6. What is the role of serum for culturing animal cells?
7. Why are $C0_2$ incubators required for animal cell culture?
8. Differentiate between roller bottle and spinner cultures?
9. How are animal cells cryopreserved? Give two examples of cryopreservatives.
10. How is erythropoietin produced by animal cell culture? Write down the procedure involved.
11. What is the mode of action of tPA? How is it produced by animal cell culture technology?
12. How are monoclonal antibodies different from polyclonal antibodies? Write one therapeutic application of monoclonal antibody.
13. What are stem cells? Describe the application of embryonic stem cell technology.
14. What is gene knock out? How is this useful in generating genetic models of human disease?
15. What is meant by tissue engineering? Discuss some important medical applications of tissue engineering.

References

1. Das HK. 2010. Text Book of Biotechnology - 4th ed., Wiley India Pvt Ltd., New Delhi.
2. Ian Freshney R. 2010. Culture of Animal Cells: A Manual of Basic Techniques and Specialized Applications - 6th ed., Wiley-Blackwell, New Jersey.
3. Masters JRW. 2000. Animal Cell Culture. A Practical Approach - 3rd ed., Oxford University Press, Oxford.
4. Morgan SJ & Darling. 1993. Animal Cell Culture, BIOS Scientific Publishers, Oxford.
5. Ratledge C & Kristiansen B. 2006. Basic Biotechnology - 3rd ed., Cambridge University Press, Cambridge.
6. Sinha BK & Kumar R. 2008. Principles of Animal Cell Culture, International Book Distributing Co., Lucknow.
7. Smith JE. 2009. Biotechnology - 5th Ed, Cambridge University Press, Cambridge.