GENETIC Engineering

SMITA RASTOGI
Department of Biotechnology
Delhi Technological University
Delhi

NEELAM PATHAK
Department of Biotechnology
Integral University
Lucknow

OXFORD
UNIVERSITY PRESS

Contents

PART II : VECTORS USED IN GENE CLONING

PART III : GENERATION AND SCREENING OF RECOMBINANTS

PART IV: APPLIED GENETIC ENGINEERING

1 An Introduction to Gene Cloning

Key Concepts

In this chapter we will learn the following:

- Cloning vectors
- Steps involved in gene cloning
- Subcloning
- Advantages of gene or cDNA cloning

1.1 INTRODUCTION

Gene cloning, developed in the early 1970s, was a significant breakthrough in the field of molecular biology. The technique was pioneered by Paul Berg, Herbert Boyer, and Stanley Cohen and was the fruit of several decades of basic research on nucleic acids (Exhibits 1.1 and 1.2). Further contributing factors for the development of gene cloning technique were critical understanding of *E. coli* genetics, information about nucleic acid enzymology (i.e., having enzymes that cut, join, and replicate DNA or reverse transcribe RNA), availability of techniques for monitoring the cutting and joining reactions, and information about bacterial plasmids. The impetus for gene manipulation *in vitro* came with the discovery of natural gene transfer capability of viruses and *Agrobacterium tumefaciens* into hosts and development of techniques for genetic transformation of *E. coli*. Another area of research that laid the foundation of gene cloning was renaturation analysis that led to the discovery of base pairing characteristics of two complementary sequences (i.e., hybridization), indicating that complementary DNA or RNA probes can be used to detect specific nucleotide sequences.

The technique of gene cloning (also called molecular cloning) essentially involves the insertion of target DNA (also called foreign DNA/passenger DNA/exogenous DNA/insert DNA/DNA fragment/gene of interest) into a cell through a

Exhibit 1.1 Deoxyribonucleic acid (DNA)

DNA is a high molecular weight biopolymer comprising of mononucleotides as their repeating units, which are joined by 3′ → 5′ phosphodiester bonds. Each mononucleotide unit of DNA consists of purine and pyrimidine nitrogenous bases, phosphorus, and a pentose sugar 2′-deoxy-*D*-ribose (Figure A). Examples of purine bases are adenine (A) and guanine (G), and examples of pyrimidine bases are cytosine (C) and thymine (T) (Figure A).

The nitrogenous bases are planar due to their π-electron clouds, hydrophobic, and relatively insoluble in water at the near neutral pH of the cell. Purines exist in both *syn* or *anti* forms, while pyrimidines due to steric interference between the sugar and carbonyl oxygen at C2 position (carbon atom at position 2) of pyrimidine exist in *anti* form. Besides, the major nitrogenous bases, some minor bases called modified nitrogenous bases also occur in polynucleotide structures, for example, 5,6-dihydrouracil, pseudouracil, 4-thiouracil, 5-methylcytosine (5-MeC), and 5-hydroxymethylcytosine, etc. Phosphorus is present in the sugar–phosphate backbone of DNA as a constituent of phosphodiester bond linking the two sugar moieties. Sugars are always in closed ring β-furanose (ketose) form and hence are called furanose sugars. The base is linked to the sugar by β-*N*-glycosidic linkage. Five atoms in the sugar ring are denoted as 1′ to 5′, while six atoms in

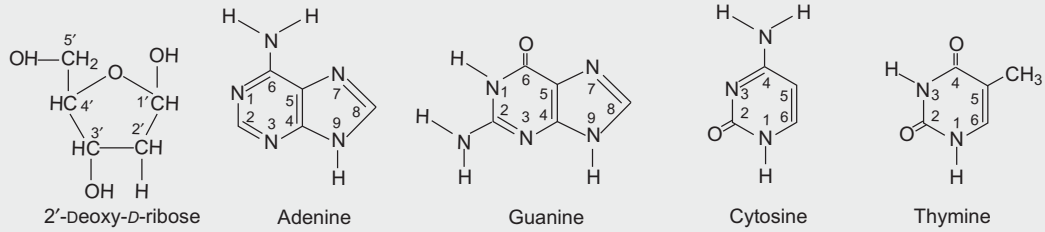


Figure A Structures of 2-deoxy-*D*-ribose sugar and four nitrogenous bases (A, G, C, and T)

pyrimidine ring (or nine in purine ring) are denoted as 1 to 6 (or 1 to 9) as shown in Figure A. The compounds in which the nitrogenous bases are conjugated to the pentose sugars by β-*N*-glycosidic linkages are called deoxyribonucleosides or deoxyribosides. These linkages are formed between the N1 atom of pyrimidine (or the N9 atom of purine) with the C1′ carbon atom of sugar. Thus, the purine deoxyribonucleosides are N-9 glycosides and the pyrimidine deoxyribonucleosides are N-1 glycosides. These are stable in alkali. The purine deoxyribonucleosides are readily hydrolyzed by acid, whereas pyrimidine deoxyribonucleosides are hydrolyzed only after prolonged treatment with concentrated acid. These deoxyribonucleosides are generally named for the particular purine and pyrimidine present. The nomenclature of deoxyribonucleosides differs from that of the bases. The trivial names of purine deoxyribonucleosides end with the suffix –sine and those of pyrimidine deoxyribonucleosides end with suffix –dine, for example, deoxyribonucleosides containing A, G, C, and T are called deoxyadenosine, deoxyguanosine, deoxycytidine, and deoxythymidine, respectively. The phosphate esters of deoxyribonucleosides are called deoxyribonucleotides or deoxyribonucleoside triphosphates. The phosphate is always esterified to the sugar moiety. Esterification can occur at any free hydroxyl group, but is most common at the 5′ and 3′ positions in sugars. The phosphate residue at α position is joined to the sugar ring at C5 by a phosphomonoester bond, while β- and γ-phosphate groups are joined in series by phosphoanhydride bonds. These deoxyribonucleotides occur either in the free form or as subunits in DNA linked by phosphodiester bond as shown in Figure B.

Chargaff (1950) formulated important generalizations about the structure of DNA, which form the Chargaff's equivalence rule. According to the rule, (i) The base composition of DNA varies from one species to another; (ii) DNA specimens isolated from different tissues of the same species have the same base composition; (iii) The base composition of DNA in a given species does not change with age, nutritional state, or changes in environment; (iv) Purines and pyrimidines are always equal such that amount of A is equal to T, and the amount of G is always equal to C (molar equivalence of bases); (v) Base ratio A+T/G+C may vary from one species to other, but is constant for a given species, and this ratio is used to identify the source of DNA; and (vi) The deoxyribose sugar and phosphate components occur in equal proportions in the sugar–phosphate backbone. Based on the X-ray data of Franklin and Wilkins and the base equivalence observed by Chargaff, Watson and Crick (1953) proposed a model for the 3-D structure of DNA. This model accounted for many of the observations on the chemical and physical properties of DNA and also suggested a mechanism for the accurate replication of genetic information.

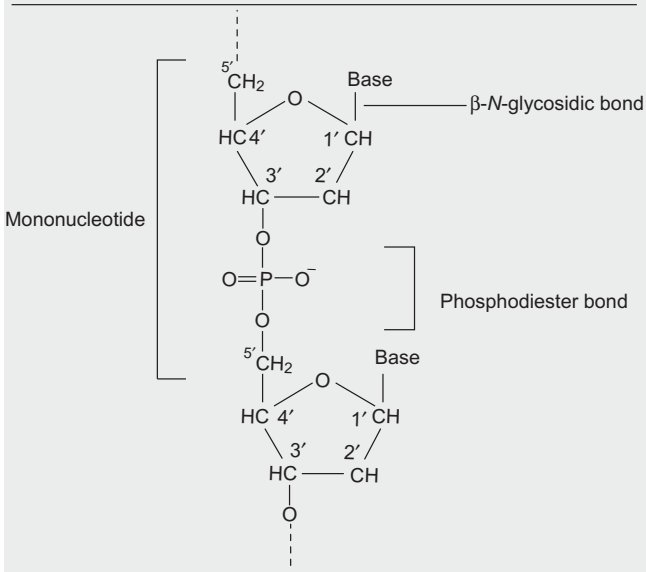


Figure B Structure of deoxyribonucleotide

DNA contains two polynucleotide chains that are coiled in helical fashion around the same axis in right-handed or counterclockwise direction, thus forming a double helix. The two chains or strands are antiparallel, i.e., the 3′,5′-internucleotide phosphodiester linkages run in opposite directions on two strands. The antiparallel orientation is a stereochemical consequence of the way that A pairs with T and G pairs with C. All the phosphodiester linkages have the same orientation along the chain, giving each linear DNA strand a specific polarity, and distinct 5′ and 3′ ends. The backbone of helix consists of sugar and phosphate groups, while purine and pyrimidine bases are stacked inside the helix with their planes parallel to each other and perpendicular to the helix axis. Hydroxyl groups of sugar forms H-bonds with water. DNA is negatively charged due to its phosphate groups and negative charges are generally neutralized by ionic interactions with positive charges of protein, metals, and polyamines. Bases being hydrophobic, project inwards to the center, and hence shielded from water. Backbone is found on the periphery of the helix and is hydrophilic. It means that single stranded structure, in which the bases are exposed to aqueous environment, is unstable. Thus DNA is double helix, in which two strands are held together by H-bonding interactions between complementary base pairs and hydrophobic [(π–π) stacking interactions between adjacent bases] interactions. A perfect Watson–Crick base pair consists of a perfect match between hydrogen donor and acceptor sites on the two bases. Thus, a purine on one strand base pairs with a specific pyrimidine on the other strand; the resulting base pair exhibits proper spatial arrangement. Thus the base A is base paired to T by double H-bonds and G is bonded to C by triple H-bonds (Figure C). This forms the concept of specific base pairing.

The individual H-bond is weak in nature, but a large number of such bonds involved in the DNA molecule confer stability to it. However, the stability of DNA is primarily a consequence of van der Waals forces and hydrophobic (base stacking) interactions between the planes of stacked bases. On one hand, H-bonding is specific, which is responsible for complementarity of two strands, on the other hand, hydrophobic interactions are nonspecific and are responsible for the stability of the macromolecule. Note that the two DNA strands tend to stick together even in the absence of specific H-bonding interactions, although these specific interactions make the association stronger. The two helices are wound in such a way so as to produce two interchain spacings or grooves, a major or wide groove (width 12 Å, depth 8.5 Å), and a minor or narrow groove (width 6.0 Å, depth 7.5 Å). The major groove is slightly deeper than the minor one. The two grooves arise because the glycosidic bonds of a base pair are not diametrically opposite each other. The minor grove contains the O2 position of pyrimidine and the N3 positions of the purine of

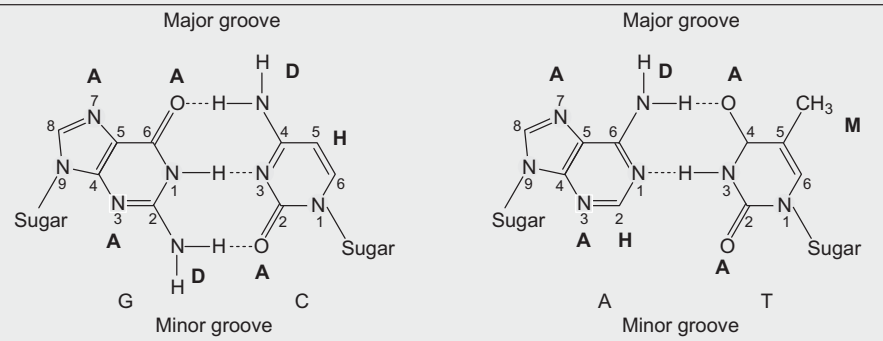


Figure C Watson–Crick base pairs: G:C and A:T

the base pair, while the major groove is on the opposite side of the pair. At each groove, potential H-bond donor and acceptor atoms are exposed to external environment, which serve as the interaction sites for the DNA-binding proteins. Thus the sequence of H-bonding donors and H-bonding acceptors exposed toward the major groove in a G:C base pair is AADH, while in an A:T base pair is ADAM (A is H-bond acceptor, D is H-bond donor, H is nonpolar hydrogen, and M is methyl group). Similarly, the sequence is MADA in T:A base pair, and HDAA in C:G base pair toward the major groove. The sequence of H-bonding donors and acceptors exposed toward the minor grove in G:C and C:G base pairs is ADA, and AHA in A:T and T:A base pairs (Figure C). Thus the major groove displays more information as compared to the minor groove. The double helices of DNA are plectonemic coils, i.e., coils that are interlocked about the same axis. These strands cannot be pulled apart, but can be separated by the unwinding process. The Watson and Crick structure of DNA is referred to as B-DNA (normal form). It is the biologically important one and exists under physiological conditions described below. The structural details of ds DNA as suggested by Watson and Crick (i.e., B-DNA) are shown in Figure D and Table A.

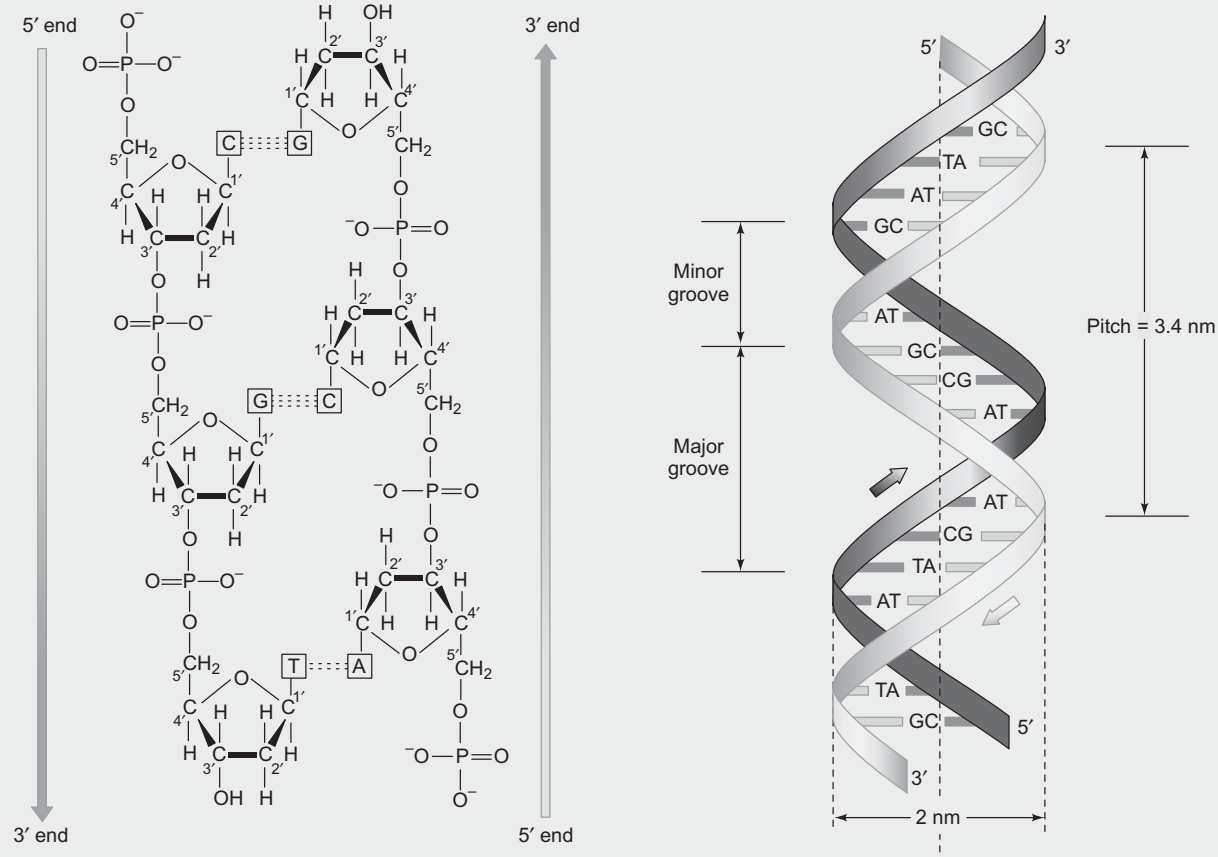


Figure D Watson–Crick double helical structure of DNA

Table A General Characteristics of Three Helical Conformers of DNA

Characteristics	Helical conformations of DNA		
	A	B	Z
Relative humidity	75%	92%	–
Ions required/salt concentration	Na^+, K^+, Cs^+ ions	Low ion strength	Very high salt concentration
Shape	Broadest	Intermediate	Narrowest
Helical state	Right	Right	Left
Pitch (base pairs per turn)	11	10.5	12 (=6 dimers)
Major groove	Deep, narrow	Wide	Flat
Minor groove	Broad, shallow	Narrow	Narrow and very deep
Helix diameter	~26 Å	~20 Å	~18 Å
Sugar pucker conformation	$C^{2'}$ endo	$C^{3'}$ endo	Alternating
Glycosidic bond angle	*anti*	*anti*	Alternating *anti/syn*
Displacement	–4.4	0.6	3.2
Twist	33	36	–49/–10
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Helix pitch	25.30 Å	35.36 Å	45.60 Å
Base tilt normal to helix axis	20°	6°	7°
Inclination	22	–2	–7
Rotation per base pair	+32.72°	+34.61°	–60° (per dimer)

DNA is strongly acidic molecule owing to the presence of phosphate groups. Double helical DNA is maximally stable between pH 4.0 and 11.0 (physiological range). Outside these physiological limits, DNA becomes unstable and unwinds. DNA is very flexible in nature, and due to thermal fluctuation, bending, stretching, and unpairing (melting) of strands can occur. However, when melted DNA is incubated at a temperature ~25°C below the denaturation temperature, the two separated strands reassociate (or reanneal or renature) to form a ds DNA molecule due to complementary sequences present in the two strands. DNA exhibits a strong positive rotation. Upon denaturation, optical rotation is highly decreased and becomes more negative. The solution of ds DNA possesses an absorption maximum at 260 nm. This characteristic absorption maximum is the property of its individual bases, and their corresponding deoxyribonucleotides. Upon denaturation of DNA, an increase in absorption of light (up to 40%) occurs even though the amount of DNA remains the same. This phenomenon is called hyperchromic effect. The temperature at the midpoint of melting curve is called melting temperature (T_m). Because of the rigidity of the double helix and the immense length of DNA in relation to its small diameter, the solution of DNA is highly viscous at pH 7.0 and room temperature (25°C). Mammalian DNA forms a band at position corresponding to 1.7 g/cm^3 buoyant density in a CsCl density gradient upon centrifugation.

Some structural variants of DNA may also arise due to difference in possible conformation of deoxyribose, rotation about phosphodiester bonds in the sugar–phosphate backbone, and free rotation about C1′-β-*N*-glycosidic bonds (*syn* or *anti*). Major helical conformers that are formed at different humidities are A and Z forms. These differ in the gross morphological features, bond angles, base inclination, displacement of the base pairs from the helical axis resulting from dehydration, helix parameters as base pairs per helical turn, helical twist, size, and shape of grooves. The variations in minor and major grooves potentially influence the nature of protein–DNA interactions and consequently affect the regulatory property of DNA. However, the key properties of DNA in different forms are not changed. A-DNA is observed when the relative humidity is reduced below 75% and is also favored in many solutions relatively devoid of water. Moreover, the double stranded regions of RNA (as in hairpins) and RNA:DNA hybrids also adopt a conformation similar to A-DNA. The structure of Z-DNA is characterized by alternating helical parameters and torsion angles with a two-base pair periodicity, causing the backbone of the helix to zig-zag and hence the name. The general characteristics of A-, B-, and Z-DNAs are tabulated in Table A.

Sources: Watson JD, Crick FHC (1953) *Nature* 171: 737–738; Dickerson RE (1983) *Sci. Am.* 249: 94–111; Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R (2004) *Molecular Biology of the Gene*, Pearson Education (Singapore), pp. 97–128.

suitable vector in such a way that inserted DNA replicates independently and transferred to progenies during cell division, thereby generating genetically identical organisms. Note that the availability of different kinds of restriction endonucleases and DNA ligases made it feasible to treat sequences of DNA as modules, which can be moved at will from one DNA molecule to another. The artificially created DNA molecule formed by the insertion of foreign DNA into a vector is called recombinant DNA molecule or DNA chimera because of its analogy with the Chimera, a creature in mythology with the head of a lion, body of a goat, and tail of a serpent. The term cloning signifies that the technique leads to generation of a line of genetically identical organisms, all of which contain the recombinant DNA molecule that can be propagated

Exhibit 1.2 Ribonucleic acid (RNA)

RNA is a long, unbranched macromolecule consisting of nucleotides joined by 3′ → 5′ phosphodiester bonds. Chemically, RNA is very similar to DNA. It contains nitrogenous bases such as A, G, C, and uracil (U), phosphorus, and a *D*-ribose sugar (Figure A). The structural differences between the components of RNA and DNA include presence of U instead of T (its nucleoside is called uridine), and *D*-ribose sugar instead of 2′-deoxy-*D*-ribose sugar.

The linkages between sugar and phosphate in the sugar–phosphate backbone, and sugar and base (perpendicular to backbone) are the same as described in the structure of DNA in Exhibit 1.1. The structural difference in the sugars of RNA and DNA confers very different chemical and physical properties on RNA. RNA is much stiffer due to steric hindrance and more susceptible to hydrolysis in alkaline conditions (high pH). Unlike DNA, under physiological conditions, RNA is single stranded. However, the single strand can fold back on itself having potentially much greater structural diversity than DNA. These secondary structures arise due to intramolecular base pairing. If the two stretches of complementary sequences are near to each other, RNA may adopt stem loop structures in which the intervening RNA is looped out from the end of the double helical segment (e.g., secondary structures include hairpin, bulge, or simple loop). Similar to DNA, weak interactions, especially base stacking interactions, play a major role in stabilizing RNA structures. Where complementary sequences are present, the predominant double stranded structure is an A-form right handed double helix. The presence of 2′-OH on the sugar residue in the RNA backbone prevents RNA from adopting a B-form helix. A feature of RNA that adds to its propensity to form secondary structures is an additional non Watson–Crick G:U base pair. Note that G:U base pair contains two H-bonds, one between N3 position of U and carbonyl on C6 of G, and the other between the carbonyl on C2 of U and N1 of G, and due to these additional G:U base pairs, RNA chains have an enhanced capacity of self-complementarity.

On basis of size, function, and stability, there are three types of RNAs, *viz.*, ribosomal RNA (rRNA), messenger RNA (mRNA), and transfer RNA (tRNA). The rRNAs are the components of ribosomes (16S, 23S, and 5S rRNAs in prokaryotic 70S ribosomes; 18S, 28S, 5.8S, and 5S rRNAs in eukaryotic nuclear 80S ribosomes), which are the sites for protein synthesis. The mRNA carries genetic information from one or a few genes to a ribosome. The protein coding region(s) of each mRNA is composed of a contiguous, nonoverlapping string of codons called open reading frame (ORF), which is a sequence of DNA consisting of triplets that is translated into amino acids starting from initiation or start codon at 5′-end

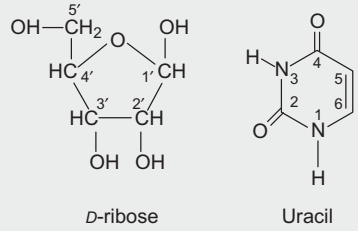


Figure A Structure of *D*-ribose and uracil

(e.g., AUG in most bacteria and eukaryotes or GUG or UUG in some bacteria), and ending with a termination or stop codon [e.g., UAG (amber) or UGA (opal) or UAA (ochre)] at 3′-end. Each ORF specifies a single polypeptide, and starts and ends at internal sites within the mRNA, i.e., the ends of an ORF are distinct from the ends of mRNA. Prokaryotic mRNA is polycistronic, i.e., a single mRNA molecule codes for two or more polypeptide chains or simply it contains multiple ORFs. The prokaryotic mRNA contains a ribosome-binding site (RBS; also referred to as Shine Dalgarno sequence). It is complementary to a sequence located near the 3′-end of 16S rRNA. RBS base pairs with 16S rRNA, thereby aligning the ribosome with the beginning of mRNA during the process of translation. Note that some mRNAs lack RBS and have translational coupling of termination with initiation due to the presence of sequence 5′-AUGA-3′ (in this overlapping sequence, UGA marks termination for previous transcript and AUG marks initiation for the next). In contrast, eukaryotic mRNA is monocistronic, i.e., a single mRNA codes for single polypeptide chain or simply it contains single ORF. There is no RBS, rather the eukaryotic mRNA is recognized by translation machinery by 5′ cap and the start codon (AUG) is reached by scanning. Note that most of the eukaryotic mRNAs contains a 5′ cap and a 3′ poly (A) tail. The tRNA serves as adapter molecule in translating the language of nucleic acids in mRNA into the language of proteins by serving as carriers of specific amino acids to specific sites on the ribosome. Thus anticodons of charged tRNAs (i.e., tRNAs covalently linked to an amino acid at 3′-end in a reaction catalyzed by amino acyl tRNA synthetase) pair with the codons of mRNA in such a way that amino acids are joined to form a polypeptide chain in a correct sequence.

Sources: Uhlenbeck OC, Pardi A, Feigon J (1997) *Cell* 90: 833–840; Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R (2004) *Molecular Biology of the Gene*, Pearson Education (Singapore), pp. 97–128.

and grown in bulk hence amplifying the recombinant DNA molecule and any gene product whose synthesis it directs. The creation of this artificial recombinant DNA molecule by a variety of sophisticated techniques, and in many cases its subsequent introduction into living cells, is referred to as genetic engineering or gene manipulation because of the potential for creating novel genetic combination by biochemical means. By this technique, even the DNA sequences not usually found together in nature can be brought together. In the developed world, there is a precise legal definition of gene manipulation as a result of government legislation to control it. In the UK, for example, gene manipulation is defined as 'the formation of new combinations by the insertion of nucleic acid molecules, produced by whatever means outside the cell,

into any virus, bacterial plasmid, or other vector system so as to allow their incorporation into a host organism in which these do not naturally occur but in which these are capable of continued propagation'. The resulting transformants are selected and screened for containing the DNA sequences of interest and later used either for research purpose or for commercial production of useful compounds related to the areas of medicine, agriculture, animal husbandry, environment, public health, forensics, etc. In this chapter, an overview of gene cloning procedure is presented; however, the details of each step are discussed in separate chapters in the book, a reference to which is given in each step. Here an emphasis is laid on the properties of ideal cloning vector and the procedures for nucleic acid isolation and purification.

1.2 CLONING VECTORS

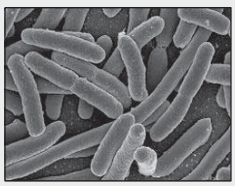
A prime requisite for a gene cloning experiment is the selection of a suitable cloning vector (or cloning vehicle), i.e., a DNA molecule that acts as a vehicle for carrying a foreign DNA fragment when inserted into it and transports it into a host cell, which is usually a bacterium although other types of living cells can also be used.

A wide variety of natural replicons exhibit the properties that allow them to act as cloning vectors, however, vectors may also be designed to possess certain minimum qualifications to function as an efficient agent for transfer, maintenance, and amplification of target DNA (for details see Section 1.2.2). As will be evident from the list (Section 1.2.1), most of the naturally occurring or artificially constructed cloning vectors in use today are for *E. coli* as the host organism. This is not surprising in view of the central role that this bacterium has played in basic research over the last 50 years (Exhibit 1.3).

1.2.1 Examples of Cloning Vectors

The examples of naturally occurring or artificially constructed cloning vectors include vectors based on *E. coli* plasmids, bacteriophages (e.g., λ, M13, P1), viruses [animal viruses

Exhibit 1.3 *Escherichia coli* (*E. coli*)



Theodor Escherich, a German pediatrician and bacteriologist, cultured '*Bacterium coli*' in 1885 from the feces of healthy individuals and concluded that this bacterium can be found almost universally in the large intestine or colon, and named it '*coli*'. It was renamed *Escherichia coli* (*E. coli*) in 1919 in a revision of bacteriological nomenclature to lend more specificity to this particular form of bacterium.

E. coli (Domain: Bacteria; Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Enterobacteriales; Family: Enterobacteriaceae; Genus: *Escherichia*; Species: *coli*) is a gram negative, unicellular, facultatively anaerobic chemoorganotroph capable of both respiratory and fermentative metabolism, mesophilic, nonphotosynthetic, and nonsporulating eubacteria. It is small (length ~2 μm, diameter 0.5 μm, and cell volume of 0.6–0.7 μm^3) and straight rod-shaped bacterium. The motile bacterium contains peritrichous flagella. It is one of the characteristic members of the normal intestinal (lower intestine) flora of warm-blooded animals, i.e., it is a coliform or enteric bacterium. However, it can also survive when released into the natural environment, allowing widespread dissemination to new hosts. The strains that are part of the normal flora of the gut are harmless and can benefit their hosts by producing Vitamin K2 or by preventing the establishment of pathogenic bacteria within the intestine. Most *E. coli* strains are harmless, but some pathogenic *E. coli* strains are responsible for infection of the enteric, urinary, pulmonary, and nervous systems, for example, serotype O157:H7 can cause serious food poisoning in humans.

The genome of *E. coli* is well studied. Sequence analysis of *E. coli* K-12 has revealed that the bacterium has a single circular chromosome of 46,39,221 bp and molecular weight of 2.7×10^9 Da. Its 4,288 protein-coding genes have been annotated, out of which 38% have no attributed function. The genome of *E. coli* K-12, like other *E. coli* genomes, has a 50.8% G+C content. Genes that code for proteins account for 87.8% of the genome, stable RNA-encoding genes make up 0.8%, noncoding repeats contribute to 0.7%, and about 11% is for regulatory and other functions. Comparison with five other sequenced microbes reveals ubiquitous as well as narrowly distributed gene families. Many families of similar genes within *E. coli* are also evident. The largest family of paralogous proteins contains 80 ABC transporters. The genome as a whole is strikingly organized with respect to the local direction of replication. The genome also contains insertion sequence (IS) elements, phage remnants, and many other patches of unusual composition indicating genome plasticity through horizontal transfer. *E. coli* and related bacteria possess the ability to transfer DNA via bacterial conjugation, transduction, or transformation, which allow genetic material to spread horizontally through an existing population.

E. coli is easy to cultivate and it can live on a wide variety of substrates. It can be easily cultivated on synthetic medium, with a minimal medium comprising of ingredients (g/l): 5 glucose, 6 Na_2HPO_4, 3 KH_2PO_4, 1 NH_4Cl, 0.5 NaCl, 0.12 $MgSO_4$, and 0.01 $CaCl_2$. It can also thrive well in complex medium, for example, Luria Bertani (LB) medium. The pH of the media is 7.0 and sterilized by autoclaving [15 psi (1.05 kg/cm^2), 121°C, 20 min]. On rich media, bacteria grow with a doubling time of 20 min, hence readily visible colonies can be seen overnight when plated on agar (Figure A). Specialized medium, like MacConkey's

agar, was developed for the selective isolation and identification of *E. coli*, as this was used as a global indicator for the pollution of water supplies. Optimal growth of *E. coli* occurs at 37°C, but some laboratory strains can multiply at temperatures of up to 49°C.

From the beginning, although pathogenic strains were also found, *E. coli* was used as a representative, harmless bacterium that could be safely and easily cultivated, and it became one of the most studied bacterium in various fields of science. *E. coli* was an integral part of the first experiments to understand bacterial and phage genetics, process of replication, conjugation, gene regulation, the concept of operon, etc. Thus, it has served as prokaryotic model organism in various studies including molecular genetics, microbiology, and genetic engineering. First cloning experiments were also undertaken in *E. coli* and this organism became the primary cloning host. However, under some circumstances, it may be desirable to use a different host for a gene cloning experiment. This is especially true in biotechnology, where the aim may not be to study a gene, but to use cloning to control or improve synthesis of an important metabolic product (e.g., a hormone such as insulin) or to change the properties of the organism (e.g., to introduce herbicide resistance into a crop plant). Hence subsequently, cloning techniques were extended to a range of other microorganisms, such as *B. subtilis*, *Pseudomonas* spp., *A. tumefaciens*, yeasts, filamentous fungi, and higher eukaryotes. Despite these advances, *E. coli* remains the most widely used cloning host even today because gene manipulation in this bacterium is technically easier than in any other organism, and the widest variety of cloning vectors are available for this organism. As a result, it is unusual for researchers to clone DNA directly in other organisms. Rather DNA from the organism of choice is first manipulated in *E. coli* and subsequently transferred to the original host or another organism. Without the ability to clone and manipulate DNA in *E. coli*, the application of recombinant DNA technology to other organisms is greatly hindered. The dominant role that *E. coli* plays in recombinant DNA technology is due to construction of many well-characterized mutants, good understanding of gene regulation, establishment of growth medium, isolation of many plasmids, and establishment of genetic transformation procedure. *E. coli* is considered a very versatile host for the production of heterologous proteins and hence by introducing genes into these microbes, mass production of recombinant proteins in industrial fermentation processes is possible. One of the first useful applications of recombinant DNA technology is the manipulation of *E. coli* to produce human insulin. Other applications of modified *E. coli* include vaccine development, bioremediation, production of immobilized enzymes, etc.

Several *E. coli* strains find application in genetic engineering experiments. Few such examples include cultivated strain K-12 (isolated at Stanford University in 1922 from human feces), DH5α, C600, and XL-1 Red. The K-12 strain has universally been adopted for fundamental work in biochemistry, genetics, and physiology. *E. coli* K-12 serves as the precursor for almost all strains used by molecular biologists for propagating cloned DNA. This is because this strain is well-adapted to the laboratory environment, and unlike wild-type strains, it is nonpathogenic and has lost the ability to thrive in the intestine. Note that the K-12 strain harbors

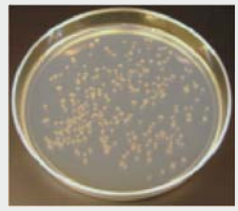


Figure A *E. coli* colonies on solid medium

a lysogenic λ bacteriophage and a number of plasmids. Many laboratory strains lose their ability to form biofilms. These features protect wild-type strains from antibodies and other chemical attacks, but require a large expenditure of energy and material resources. Improvements in *E. coli* K-12 for recombinant DNA experiments include: (i) Removal of *Eco* K restriction system (*hsd* R^-); (ii) Removal of *mcr* A/*mcr* B genes (*mcr* A^-/*mcr* B^-) that are responsible for degrading methylated foreign DNA, but not at the sites that *E. coli* K-12 recognize as its own (e.g., human and mouse DNA is CpG methylated); (iii) *rec* A^- mutation that suppresses homologous recombination, which makes it more sensitive to UV light; and (iv) *end* A^- mutation in the endonuclease A gene that greatly improves the quality of DNA isolated with biochemical techniques. Some derivatives of K-12 are XL1-Blue strain [*rec* A1 *end* A1 *gyr* A96 *thi*-1 *hsd* R17 *sup* E44 *rel* A1 *lac* {F′ *pro* AB *lac* I^q ZΔM15 *Tn* 10 (*tet*^{r})}], and XL1-Blue MR strain {Δ(*mcr* A) 183 Δ(*mcr* CB – *hsd* SMR – *mrr*) 173 *end* A1 *sup* E44 *thi*-1 *rec* A1 *gyr* A96 *rel* A1 *lac*}. Like many cloning strains, *E. coli* DH5α strain with a genotype [*fhu* A2 Δ(*arg* F - *lac* Z) *U* 169 *pho* A *gln* V44 ϕ80 (*lac* Z) ΔM15 *gyr* A96 *rec* A1 *rel* A1 *end* A1 *thi* -1 *hsd* R17] has several important features, which make it useful for recombinant DNA methods. These are as follows: (i) The strain transforms with high efficiency; (ii) The *end* A1 mutation inactivates an intracellular endonuclease that degrades plasmid DNA in many miniprep methods; (iii) The *hsd* R17 mutation eliminates the restriction endonuclease of the *Eco* KI restriction–modification system, so DNA lacking the *Eco* KI methylation will not be degraded. DNA prepared from *hsd* R strains that are wild-type for *hsd* M will be methylated and can be used to transform wild-type *E. coli* K-12 strains; (iv) *lac* ZΔM15 is the α-acceptor allele needed for blue-white screening with many *lac* Z-based vectors; (v) *rec* A eliminates homologous recombination. This reduces deletion formation and plasmid multimerization; and (vi) *sup* E44, with a systematic name *gln* V44, is an amber suppressor mutation. The chromosomal genotype of strain C600 is *sup* E44 *hsd* R *thi*-1 *leu* B6 *lac* Y1 *ton* A21 *hfl* 150 *chr :: Tn* 10 (*tet*^{r}). The genotype of XL-1 Red is *end* A1 *gyr* A96 *thi*-1 *hsd* R17 *sup* E44 *rel* A1 *lac mut* D5 *mut* S *mut* T. This strain is used for introducing random mutations into a gene of interest. Though this strain has the *Tn* 10 insertion with tetracycline, it should not be used for selection, as it is frequently lost.

Sources: Blattner FR, Plunkett IIIG, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF (1997) *Science* 277: 1453–1474; http://ecoliwiki.net/colipedia/index; Stanier RY, Lingraham JL, Wheelis ML, Painter PR (1986) *The Microbial World*, Prentice-Hall India, pp. 145–182.

(e.g., retrovirus, adenovirus, adeno-associated virus, Herpes simplex virus, Sindbis virus, Semliki forest virus, Lentivirus, *Vaccinia* virus, SV40, etc.), insect viruses (e.g., baculovirus), and plant viruses (e.g., cauliflower mosaic virus, gemini virus, tobacco mosaic virus, potato virus X, etc.)], *A. tumefaciens* tumor-inducing plasmid-based vectors, *A. rhizogenes* root-inducing plasmid-based vectors, chimeric plasmids (e.g., cosmid, phagemid, phasmid, fosmid), artificial chromosomes (e.g., yeast artificial chromosome, bacterial artificial chromosome, P1-derived artificial chromosome, mammalian artificial chromosome, and human artificial chromosome), and non-*E. coli* vectors (e.g., *Bacillus* and *Pseudomonas* vectors, etc.) (for details see Chapters 8 to 12).

1.2.2 Properties and Construction of a Vector DNA Molecule

A vector can be used for cloning to get DNA copies of the fragment inserted into it or as an expression vector to obtain expression of the cloned gene, i.e., to get either the RNA copies or protein corresponding to the cloned gene. To be ideal as a cloning vector, these should exhibit certain minimal properties; however, a few vectors should also possess certain specialist functions. These minimal or specialist properties are acquired either naturally or artificially (by retaining genes/loci for useful quality attributes and deleting nonessential ones from the potential vector DNA, or by gaining or combining useful genes/loci from different DNA molecules). The properties of ideal vectors to be used simply for multiplication of DNA or for expression of the cloned gene are as follows:

Capability of Autonomous Replication It is preferable to design a vector molecule that is capable of replication within the host cell so that numerous copies of recombinant DNA molecules are produced and passed to the daughter cells. Moreover, isolation of vector DNA should be possible independently of the host's genome. Thus, replicons containing an origin of replication, i.e., DNA capable of self-replication, are preferred as cloning vectors. Note that bacterial and viral genomes contain only one origin of replication, while eukaryotes contain multiple origins (i.e., autonomously replicating sequences (*ARS*).

On the contrary, some vectors are designed to lack the capability of autonomous replication (i.e., lack origin of replication). The purpose is to allow long-term expression of cloned genes to get several copies of DNA before expression or to regulate the expression of the genes toxic to the host cell. As these vectors lack origin of replication, the only way for their multiplication is by integration into the host genome. Such integrating vectors have low copy number and do not suffer from the problem of gene-dosage effect. Note that for integrating vectors, mapping of foreign DNA and the surrounding host sequences is required so as to facilitate integration by homologous recombination.

Small Size A cloning vehicle needs to be reasonably small in size (lower molecular weight) and manageable. Small molecules are easy to handle, isolate, and manipulate, as these are less susceptible to damage by shearing stress during purification. On the other hand, large molecules tend to break down or degrade due to shearing and are more difficult to handle and manipulate. Moreover, in small molecules, the chance of occurrence of unique sites for restriction enzyme(s) increases. In addition, the efficiency of gene transfer is also high with small vector molecules. Furthermore, small plasmids are usually high copy number plasmids and hence give higher yield of product (RNA or proteins) expressed from the cloned gene. The only disadvantage of high copy number plasmids is gene-dosage effect. Small wild-type plasmids are usually nonconjugative and hence can be biologically contained. Thus for cloning, a small wild-type vector may be used directly for cloning or a large vector DNA molecule may be subjected to size (or gene) diminution, i.e., reduction in size by deleting sites that do not affect its replicative ability, growth, and viability.

Presence of Selectable Marker Gene(s) Vector molecules should possess a selectable marker gene(s) (e.g., antibiotic resistance genes, *lac* Z, or resistance to toxin, etc.) that confer readily scorable phenotypic traits on host cells harboring recombinant DNA molecule, thereby allowing easy detection of the recombinants (for details see Chapter 16). If not present naturally, these selectable marker genes are artificially introduced into the vector molecule by different strategies including restriction digestion of source DNA followed by ligation of fragment containing selectable marker gene(s) into vector molecule, recombineering, or transposition. Note that some antibiotic resistance genes that are located on transposons are moved from one bacterial cell to another by transposition [e.g., *Tn* 1, *Tn* 5, *Tn* 9, and *Tn* 10 contain resistance genes for ampicillin (amp^r or ap^r), kanamycin (kn^r or kan^r), chloramphenicol (cm^r), and tetracycline (tet^r), respectively]. Some expression vectors also possess reporter gene(s) to determine the expression of the cloned gene and its level of expression (for details see Chapter 16).

Presence of Unique Restriction Enzyme Site(s) or Multiple Cloning Sites for Inserting the Target DNA For cloning the gene of interest, unique restriction enzyme site(s) should be present on the vector DNA molecule either individually or as cluster in the form of multiple cloning site (MCS; polylinker). The position of these restriction sites should be such that the insertion of a segment of DNA in any of these sites brings about a phenotypic change in the characteristics of vector DNA molecule, for example, loss of resistance to antibiotic or loss of expression of a gene whose

product is an enzyme that normally carries out a reaction, leading to an easily recognizable trait (e.g., change in color). Note that MCS is chemically synthesized cluster of restriction enzyme recognition and cleavage sites. The clustering of restriction enzyme sites in the form of MCS allows a DNA fragment with two different ends to be cloned in particular orientation (i.e., directional cloning is possible provided the restriction enzyme sites in MCS are asymmetric) without resorting to additional manipulations such as linker/adaptor attachment. This also increases the flexibility in the use of restriction enzymes.

Ease of Purification Generating and purifying large amounts of vector DNA from the host cell should be easy and straightforward. This is possible when the size of the vector DNA molecule is small so that it is not susceptible to damage by shearing and the purification procedure is simple.

No Effect on the Replicative Ability of Vector due to Insertion of Target DNA The introduction of target DNA into vector DNA molecule to form recombinant DNA molecule followed by its introduction into host cell should not change the replicative ability of vector DNA molecule.

Ease of Reintroduction into Host Cell with High Efficiency The transformation/transfection protocols should be well-established and easy to perform. Moreover, the efficiency of these procedures should be high.

Biological Containment The self-catalyzed transfer of recombinant DNA from one host to another is not preferred. In simple words, vectors should be biologically contained with no possibility of gene escape. To achieve this, one way is to use nonconjugative and nonmobilizable plasmid vectors. The conjugative or mobilizable plasmids may be made nonconjugative or nonmobilizable by mutating or deleting responsible loci, for example, *tra*, *mob*, *nic*, or *bom*. Their deletion also leads to extensive reduction in plasmid size (for details see Chapter 8). For biological containment of λ bacteriophage, the strategy for biological containment is based on amber mutations in the lysis genes, for example, *S*, *R*, head, and tail genes, and their suppression using sup^+ host at a later stage when required (for details see Chapter 9).

Presence of Promoters and Ribosome Binding Site Through a vector, it is possible to give a cloned DNA fragment additional characteristics that may expand its use considerably. For example, expression vector containing promoter *in frame* with the cloned gene (transcriptional fusion vector) allows *in vitro* transcription of the cloned DNA fragment and that containing *in frame* promoter and ribosome binding site (RBS) (translational fusion vector) allows *in vitro* translation of the cloned DNA fragment (for details see Chapter 12). However, the transcription and/or translation of the cloned gene can also be achieved in a vector lacking promoter and RBS by including them *in frame* with the cloned gene itself. Note that if we are interested in simply getting the DNA copies of the cloned gene, then expression vectors need not be used.

Presence of Two Different Origins of Replication or Broad Host-range Origin of Replication Most of the routine manipulations involved in gene cloning experiments use *E. coli* as the host organism; consequently, most cloning vectors have origin of replication for multiplication in *E. coli*. *E. coli* is particularly used when the aim of the cloning experiment is to study gene structure and function. However, when the aim is not to study a gene, but is controlling or improving synthesis of an important metabolic product (e.g., a hormone such as insulin) or changing the properties of the organism (e.g., to introduce herbicide resistance into a crop plant), it is desirable to use a different host for a gene cloning experiment after gene manipulation studies in *E. coli*. Thus for such purposes, initial isolation and analysis of DNA fragments is almost always carried out using *E. coli* as the host organism and further manipulations are done in the second host. To easily carry out these steps, there should be some strategy that allows multiplication both in *E. coli* and other host. A solution to this is a shuttle vector, i.e., the vector containing two different origins of replication or a broad host-range origin of replication that allows multiplication in two different hosts, for example, *E. coli* and *A. tumefaciens* (for details see Chapter 12).

Specialist Vectors Expression vectors may be designed for some special purposes (for details see Chapter 12). For example, the expression vector design may be needed for the following: (i) Vector DNA molecules targeting the expressed protein corresponding to the gene of interest to extracellular medium for easy purification and reduced contamination with cellular extracts. This involves insertion of the target DNA *in frame* with signal sequence of secretory proteins; (ii) Vector DNA molecules allowing surface display of the expressed heterologous protein so that recombinants are easily recognized through immunological procedures. For this, the target DNA is cloned within the gene encoding outer membrane protein (in bacteria) or in the gene encoding a capsid protein (in phage); and (iii) Tagged vector DNA molecules that allow easy purification of the expressed heterologous protein through affinity purification or antigen–antibody reaction. This involves appending of a His tag or a marker peptide to the gene of interest.

1.3 STEPS INVOLVED IN GENE CLONING

As evident from the previous section, the prime requisite of a gene cloning experiment is the choice of a suitable vector.

Once this selection is done, a gene cloning experiment can be performed through a series of steps. The first step is the isolation of pure DNA from a particular source. This is followed by isolation of DNA fragments using different strategies, which are discussed in Section 1.3.3. Note that the gene of interest may become a part of a single or a few DNA fragments. DNA fragments containing the gene to be cloned (i.e., fragmented genomic DNA) are inserted into cloning vectors to produce composite molecules called the recombinant DNA molecules. Note that each DNA fragment becomes inserted into a different vector molecule to produce a family of recombinant DNA molecules, each carrying a different gene or part of a gene. The resulting mixture of recombinant DNA molecules thus represents the entire genetic complement of an organism, one (or more) of which carries the gene of interest. In the next step, recombinant DNA molecules are introduced into host cells, usually bacteria, or any other living cells. These recombinants multiply when plated on a suitable selection medium, producing colonies or plaques depending upon the vector used. The individuals in the colonies or plaques are genetically identical and are referred to as 'clones' and the collection of all the clones is called a 'genomic library'. Note that usually only one recombinant DNA molecule is transported into any single host cell, so that the final set of clones may contain multiple copies of just one molecule. The library is then screened to isolate the clone(s) of interest and the desired clone(s) is stored.

For cDNA library construction, total RNA is isolated from the source, followed by mRNA purification from total RNA and cDNA synthesis through reverse transcription (for details see Chapter 5). The end result is several cDNA molecules; one from each mRNA. Each cDNA is then cloned into vector DNA molecules and the resulting recombinant DNA molecules are introduced into host cells to get a cDNA library, which represents the entire complement of the expressed sequences within a given cell or tissue type at a particular period.

On the contrary, if the gene is small and its sequence is known from same or different source, it can be directly synthesized chemically taking account of codon bias (for details see Chapter 4) or amplified using gene-specific primers through polymerase chain reaction (PCR) (for details see Chapter 7). In either case, a single DNA fragment containing the gene of interest is obtained, which can be directly cloned into a vector DNA molecule resulting in the formation of a single recombinant DNA molecule, i.e., a library is not obtained.

A diagrammatic representation of the steps involved in a gene cloning experiment is presented in Figure 1.1, and in the following sections, these steps are discussed in detail.

1.3.1 Isolation and Purification of Total Cellular DNA

The first step in a gene cloning experiment is isolation of DNA from a particular source. There are several procedures for the isolation and purification of DNA. However, the fundamentals of all these procedures are same, and are discussed in this section. For easy understanding, the simplest type of DNA purification procedure from bacteria is discussed in detail, while only specific requirements are mentioned for the isolation and purification of DNA from other sources. The procedures for isolation and purification of vector DNA, for example, plasmids and phage (λ and M13) DNA is also based on the same principle as described below; however, the exact procedures are discussed in Chapters 8 to 10.

Isolation and Purification of Total DNA from Bacterial Cells

For all the procedures in molecular cloning experiments, the first step is the isolation of pure and good quality DNA. The procedure for total DNA preparation from a bacterial culture is divided into four stages: (i) Growing and harvesting a bacterial culture; (ii) Lysis of bacterial cells to release their contents and removal of cell debris (i.e., to get cell extract); (iii) Treatment of cell extract to remove all contaminating components (i.e., purification of DNA); and (iv) Precipitation and concentration of DNA (i.e., recovery of purified DNA). These steps are discussed in detail below and also summarized in Figure 1.2.

Growing and Harvesting a Bacterial Culture The first step in the isolation of DNA from bacterial cells is growing and harvesting a bacterial culture.

Growth medium Most bacteria are grown easily in a liquid medium (broth culture), which provides a balanced mixture of all the essential nutrients at concentrations that allow efficient growth and division of bacteria. Two typical bacterial growth media are LB and M9; however, there are a variety of other specific media. LB medium comprises of tryptone (10 g/l), yeast extract (5 g/l), NaCl (10 g/l), and pH adjusted to 7.0. Note that it is a complex or undefined medium, since the precise identity and quantity of its components are not known. This is because two of the ingredients, tryptone and yeast extract, are complicated mixtures of unknown chemical compounds. Tryptone supplies amino acids and small peptides, while yeast extract (a dried preparation of partially digested yeast cells) provides nitrogen along with sugars and inorganic and organic nutrients. This medium needs no further supplementation and supports the growth of a wide range of bacterial species. On the other hand, M9 is a defined medium in which all the components and their quantities are known. This medium contains a mixture of inorganic

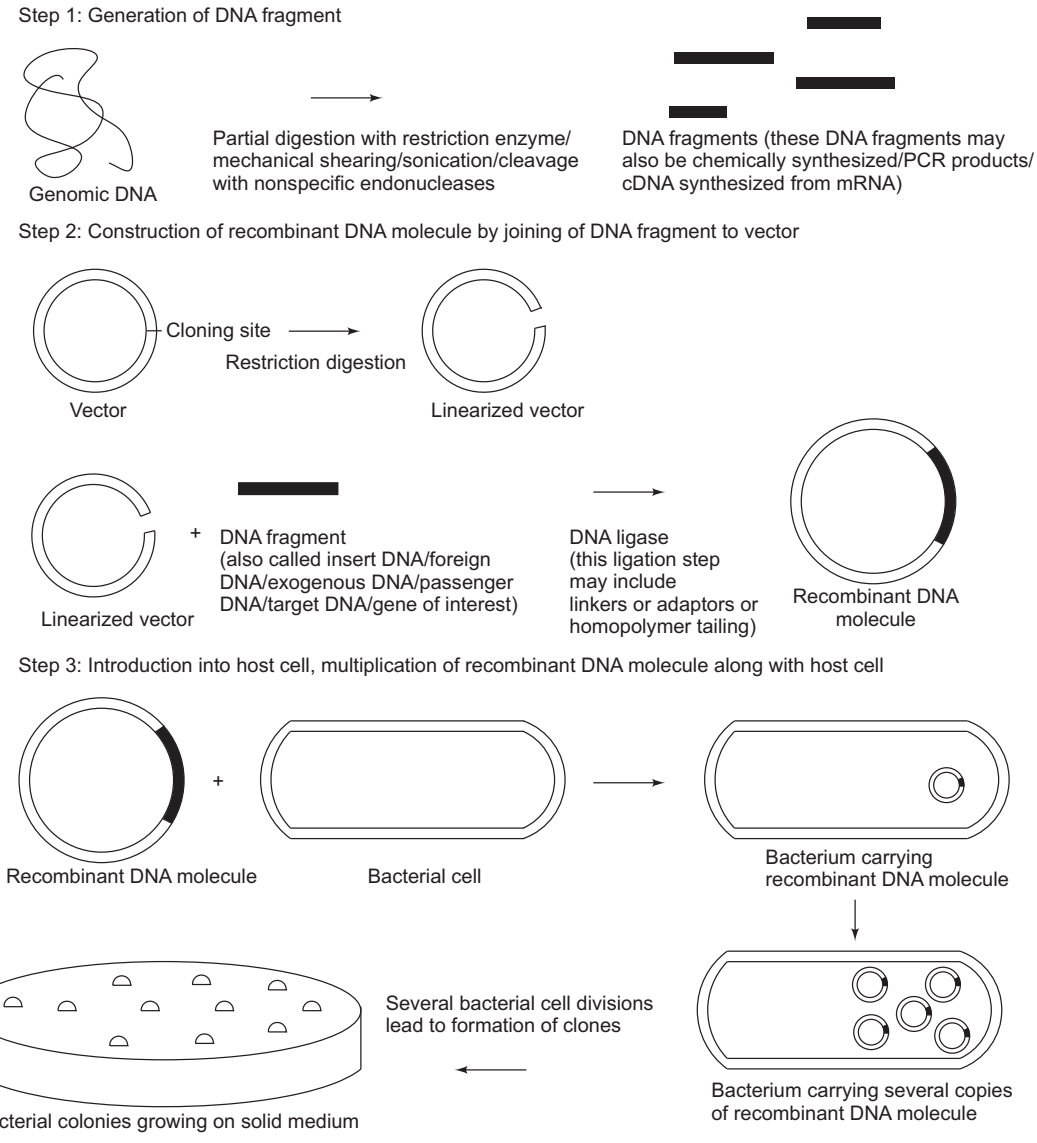


Figure 1.1 Basic steps in gene cloning [In this figure, the genomic DNA and vector are cut with blunt end cutter, however, a sticky end cutter can also be used depending upon the requirement.]

nutrients to provide essential elements such as nitrogen, magnesium, and calcium, as well as glucose to supply carbon and energy. The ingredients (per liter) of M9 minimal medium are 5 X M9 salts (200 ml; comprising 64.0 g/l $Na_2HPO_4 \cdot 7H_2O$, 15.0 g/l KH_2PO_4, 2.5 g/l NaCl, 5.0 g/l NH_4Cl), 1 M $MgSO_4$ (2 ml), 20% solution of appropriate carbon source (e.g., 20% glucose; 20 ml), 1 M $CaCl_2$ (0.1 ml), and rest water; pH adjusted to 7.0. Note that separately sterilized $MgSO_4$ and $CaCl_2$ solutions (by autoclaving) and glucose (by membrane sterilization through 0.22 μm filter) are added to autoclaved and diluted solution of 5 X M9 salts. In practice, additional growth factors such as amino acids, vitamins and trace elements are added to the M9 minimal medium to support bacterial growth, the selection of which depends on the species concerned. Defined media is used when the bacterial culture has to be grown under precisely controlled conditions, while complex medium is appropriate when the culture is being grown simply as a source of DNA.

Growth conditions As *E. coli* is mesophilic and facultative anaerobe, it is grown in LB medium at 37°C and aerated by shaking at 150–250 rpm on a rotary platform. Under these conditions, *E. coli* cells divide once in every 20 min approximately.

Harvesting by centrifugation after acquisition of appropriate cell density The density of a bacterial culture is measured by monitoring the optical density (OD) at 600

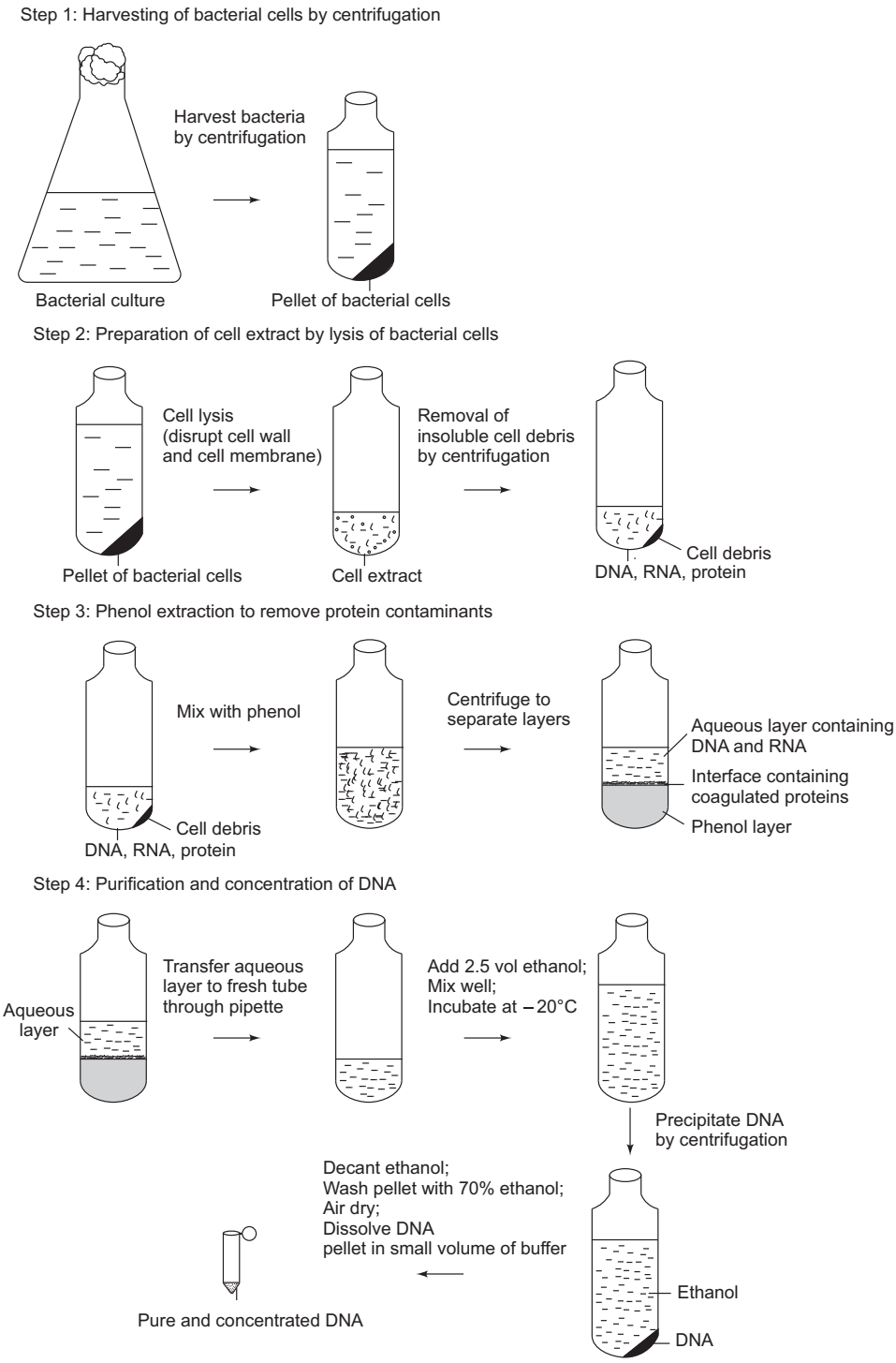


Figure 1.2 The basic steps in preparation of total cell DNA from a culture of bacteria

nm. At this wavelength, one OD corresponds to about 0.8×10^9 cells/ml. Once appropriate growth is obtained, the bacterial cells are harvested by spinning the culture in a centrifuge at fairly low centrifugation speeds for small time intervals (3,000–8,000 *g* for ≤5 min) to prevent the formation of compact pellets. After centrifugation, the bacterial cells pellet at the bottom of the centrifuge tube, allowing the culture medium to be decanted off. The harvested bacterial cells are

then resuspended in as small a volume as possible. Note that the bacteria from 1,000 ml culture at maximum cell density are resuspended into a volume of 10 ml or less.

Bacterial Cell Lysis and Removal of Cell Debris (Preparation of Cell Extract) Having harvested bacterial cells, the next step is to prepare a cell extract. As *E. coli* cells are enclosed by various barriers (cytoplasmic membrane, rigid cell wall, and an outer membrane enveloping the cell wall) preventing the release of cell components, all these barriers have to be disrupted. The techniques that are employed for breaking open bacterial cells fall into two categories: physical and chemical methods.

Physical methods These methods involve bacterial cell lysis by mechanical forces, for example, vortexing and sonication. The physical methods are used either alone or in combination with chemical methods.

Chemical methods Chemical methods involve disruption of cells by exposure to chemical agents that affect the integrity of the cell barriers. These chemical disruption methods are the most commonly used methods for DNA preparation; however, a combination of both chemical and physical methods may also be employed. Generally, two chemical agents are employed, one that attacks the cell wall and the other that disrupts the cell membrane, and their choice depends on the species of bacterium. With *E. coli* and related organisms, lysozyme, EDTA (disodium salt), or a combination of both is used to weaken the cell wall. Note that lysozyme (an enzyme present in egg-white or secretions such as tears, saliva) catalyzes the hydrolytic cleavage of the *N*-acetyl muramyl–*N*-acetyl glucosamine (NAM–NAG) linkage in the glycan strands (i.e., glycosidic linkage between NAM and NAG), which destroys the integrity of the peptidoglycan, consequently disrupting the rigidity of cell wall. The resultant weakening of the peptidoglycan in the growing cell leads to osmotic lysis. On the other hand, Na_2EDTA is a chelating agent that sequesters divalent metal cations such as Mg^{2+} and Ca^{2+} and hence disrupts the overall structure of cell envelope. Removal of Mg^{2+} leads to inactivation of cellular DNases and hence prevents degradation of DNA. Note that Mg^{2+} ions are essential for preserving the overall structure of cell envelope, as well as for the action of cellular nucleases. Disodium salt of EDTA (Na_2EDTA) is used instead of tetrasodium salt (Na_4EDTA), because many molecular biology enzymes are highly sensitive to salt concentration. Buffers prepared with Na_4EDTA are initially more alkaline than those made with Na_2EDTA and the subsequent addition of HCl to establish pH 7.0–8.0 greatly increases the NaCl concentration in the final buffer, and likely have a negative influence on the outcome of an experiment. Under some conditions, weakening of the cell wall with lysozyme or Na_2EDTA is sufficient to cause bacterial cells to burst, but usually a detergent such as sodium dodecyl sulfate (SDS) is also added. These detergents dissolve lipids and hence stimulate disruption of cell membranes, consequently releasing the cellular components.

After lysis, the final step in the preparation of a cell extract is removal of insoluble cell debris. Partially digested cell wall fractions are pelleted by high-speed centrifugation, leaving the cell extract as a reasonably clear supernatant.

Removal of Contaminating Biomolecules from the Cell Extract A bacterial cell extract contains significant quantities of proteins and RNA in addition to DNA. These contaminating biomolecules need to be removed, leaving DNA in the pure form for use in a gene cloning experiment. A variety of simple procedures in use today for this purpose are as follows:

Deproteination The removal of proteins can be achieved by two different ways. These are phenol:chloroform extraction and treatment with proteases.

Phenol:chloroform extraction The removal of proteins is often carried out by extracting the cell extract with organic solvents (e.g., phenol, chloroform, isoamyl alcohol), which denature and precipitate proteins very efficiently, leaving the nucleic acids (DNA and RNA) in aqueous solution. The cell extract is first extracted with an equal volume of phenol:chloroform (1:1). For this purpose, molecular biology grade phenol equilibrated at pH 8.0 (by 1 M Tris-HCl, pH 8.0) should be used. This is essential because nucleic acids tend to partition into the organic phase if phenol has not been adequately equilibrated to a pH of 7.8–8.0. Moreover, phenol should be distilled before use to remove impurities. This is because phenol is very unstable and oxidizes rapidly into quinones (phenol oxidation products), which form free radicals that break phosphodiester linkages, cross-link nucleic acids, and impart a pinkish tint to phenol. Optionally, 0.1% w/v 8-hydroxyquinoline may be added to phenol during equilibration. As an antioxidant, 8-hydroxyquinoline stabilizes phenol by retarding the oxidation of phenol and consequent formation of quinones. It also imparts a bright yellow color to the phenol to which it is added, and hence to the organic phase; this facilitates to keep track of the organic and aqueous phases or aids in easy visualization of separated layers. Chloroform facilitates partitioning of the aqueous and organic materials. The extraction with phenol:chloroform involves proper mixing of the contents of the tube until the formation of an emulsion followed by centrifugation at high speed (12,000 *g* for 10 min). After centrifugation, aqueous phase forms the upper phase, while organic phase (yellow if 8-hydroxyquinoline is added) forms the lower phase. In case the aqueous phase contains >0.5 M salt or >10% sucrose, it becomes denser than organic phase and forms the lower phase. The precipitated protein molecules form a white coagulated mass at the interface between the aqueous and organic layers.

As nucleic acids (DNA and RNA) contain highly charged phosphate backbone, these are polar and hydrophilic (water soluble) in nature and hence get partitioned in the aqueous layer along with salts. This is explained by polar nature and high dielectric constant of water. [Note that water is also polar molecule with a partial negative charge near the oxygen atom due the unshared pairs of electrons and partial positive charges near the hydrogen atoms. It has a high dielectric constant (80.1 at 20°C), indicating that electrical force between any two charges in aqueous solutions is highly diminished compared with the force in vacuum or air. At an atomic level, this diminishing of force acting on charges results from water molecules forming hydration shells (solvation shells) around them. It makes water a very good solvent for charged compounds (nucleic acids or salts)]. The aqueous phase (containing nucleic acids and salts) is pipetted out and transferred to a fresh tube, while organic phase and interface (containing denatured proteins) are discarded. Some cell extracts contain such a high protein content that a single phenol extraction is not sufficient to completely purify nucleic acids. This problem can, however, be circumvented by performing several phenol:chloroform extractions one after the other until no protein is visible at the interface. Moreover, deproteination is more efficient when two or more different organic solvents are used instead of one. Hence, it is preferable to perform extraction with an equal volume of phenol:chloroform: isoamyl alcohol (25:24:1) after phenol:chloroform extraction. Isoamyl alcohol in the mixture reduces foaming generated by the mechanics of the extraction procedure. After organic extraction(s), the aqueous phase is extracted with an equal volume of chloroform. As phenol in extensively soluble in chloroform, any lingering traces of phenol are removed by extracting with an equal volume of chloroform. The removal of trace amounts of phenol from the nucleic acid preparation is essential so that quinones do not compromise the integrity of the nucleic acid sample. Note that phenol, chloroform, and their combinations are caustic and carcinogenic reagents that must be handled with extreme care. Phenol is highly corrosive. It anesthetizes the skin and can result in severe burns that scar on heating. It is therefore essential to work in a fume hood and wear a lab coat, gloves, and eye protection glasses. If any phenol comes into contact with skin, the affected area should be washed immediately with copious amounts of water. Note that ethanol (EtOH) should not be used for washing since this enhances phenol absorption and increases the severity of burn.

Treatment with proteolytic enzymes The removal of protein contaminants by several phenol:chloroform extractions is undesirable as each mixing and centrifugation step results in a certain amount of breakage of the DNA molecules, resulting in decrease in DNA yield. As an alternative, it is usual to remove most of the proteins before phenol extraction by digestion with proteolytic enzymes (proteases) such as pronase or proteinase K. These enzymes are active against a broad spectrum of native proteins and break down the polypeptides into smaller units that are easily removed by phenol. Proteinase K is proteolytic enzyme (a serine protease) that is purified from the mold *Tritirachium album*. In solution, it is stable over pH range 4–12.5 with an optimum of pH 8.0 and a temperature range 25–65°C. Although the enzyme has two binding sites for Ca^{2+}, in the absence of this divalent cation, some catalytic activity is retained for the degradation of proteins commonly found in nucleic acid preparations. Occasionally, proteinase K digestion is carried out in the presence of Na_2EDTA at 50°C to inhibit labile, Mg^{2+}-dependent nucleases. Proteinase K is commonly prepared as a 20 mg/ml stock solution in sterile water (stable for 1 year at –20°C) or in a solution of 50 mM Tris-HCl (pH 8.0) and 1 mM $CaCl_2$ (stable for months at 4°C). It is generally used at a working concentration of up to 50 μg/ml in any of a number of buffer formulations, one example of which is 10 mM Tris-HCl (pH 8.0), 1 mM Na_2EDTA, and 0.5% SDS. Maximum proteinase K activity is observed with the inclusion of 1 mM Ca^{2+} in the reaction buffer. Instead of proteinase K, pronase isolated from *Streptomyces griseus* can be used. In some cases, it may be necessary to perform a pronase self-digestion to eliminate contaminating DNase activity. If necessary, this is easily accomplished by incubation of the pronase stock [20 mg/ml in 10 mM Tris-HCl (pH 7.5) and 10 mM NaCl] at 37°C for 1 hour. This extra task is usually avoided by purchasing predigested pronase. In either case, suitable aliquots of pronase are stored at –20°C. Reaction conditions for pronase are identical to those for proteinase K except that the recommended working concentration for pronase is about 1 mg/ml.

Removal of RNA contamination The contaminating RNA molecules must also be removed from DNA before subjecting it to cloning experiments. Some RNA molecules, especially mRNAs, are removed by phenol extraction as described earlier. However, this is not a useful strategy for most RNA molecules, which remain with DNA in the aqueous layer even after phenol extraction. Traditionally used isopycnic ultracentrifugation (for details see Chapter 6) for partitioning grossly contaminated samples is also not the method of choice for removal of RNA. This is because the method is time-consuming, requires expensive and highly specialized equipment, and not preferable when the combined mass of DNA and RNA in the sample totals only a few micrograms. The strategies routinely employed for the removal of contaminating RNA molecules include the following:

Ribonuclease treatment After deproteination, the cell extract comprises of nucleic acids. Contaminating RNA molecules need to be removed from this nucleic acid preparation

without compromising the integrity of DNA. The most effective way to remove RNA is to treat aqueous layer with the enzyme, ribonuclease (RNase). Two most commonly used RNases in DNA isolation procedure are RNase A and RNase T1 [stock solution of 10 mg/ml in water or TE buffer (10 mM Tris-HCl and 1 mM Na_2EDTA, pH 8.0)]. These enzymes rapidly degrade RNA into ribonucleotide subunits, which are then easily removed by phenol extraction. As crude preparations of RNases, if not specially treated, can harbor significant levels of DNase activity as well, RNase preparations should be purged of all intrinsic DNase activity (i.e., made DNase-free) in the laboratories. This is routinely done by heating RNase stock solutions to near boiling (90°C) for 10 min, followed by cooling on ice. This is because at 90°C, DNase activity is quickly eliminated without compromising the RNase activity of the reagent. The DNase-free RNase thus obtained is stored frozen in aliquots. DNase-free RNase that harbors no intrinsic DNase activity is also commercially available. Note that RNase is costly and should be used in minute amounts. Hence, RNase treatment of nucleic acid preparation to remove RNA contamination is done after concentration step.

Application of monophasic reagent The application of monophasic reagent containing acidified phenol, guanidinium or ammonium thiocyanate, and a phenol solubilizer as lysis buffer allows single-step simultaneous isolation of DNA, RNA, and proteins. This method involves lysis of cells with a monophasic solution of guanidine or ammonium isothiocyanate and phenol. Note that guanidinium salts are chaotropic and denaturing agents that function by destroying the 3-D structure of proteins and convert most proteins to a randomly coiled state through an unclear mechanism. Addition of chloroform generates a second (organic) phase into which DNA and proteins are extracted, leaving RNA in the aqueous supernatant. DNA is purified from the organic phase by precipitation with EtOH. DNA recovered from the organic phase is ~20 kbp in size and is a suitable template for PCR. The proteins, however, remain denatured as a consequence of their exposure to guanidine and are used chiefly for immunoblotting. As RNA is partitioned in aqueous phase, its contamination in DNA preparation is easily eliminated. Note that the same procedure is used for the isolation of pure RNA, where RNA is precipitated from aqueous phase with isopropanol and the contaminating DNA and proteins are eliminated in the organic phase (for details see Section 1.3.2). Contaminating biomolecules may also be removed by silica-based purification in combination with guanidinium thiocyanate. These methods also form the basis of some commercially available nucleic acid isolation kits.

Precipitation and Concentration of DNA The next step in a DNA isolation and purification procedure is recovery of DNA through easy and rapid processes, which are discussed below.

Precipitation with EtOH Coulomb's law dictates the electrostatic attractions between the positively charged ions in solution and the negatively charged phosphate ions, and this interaction is also affected by the dielectric constant of the solution. As water has a high dielectric constant, it becomes fairly difficult for the sodium and phosphate ions to come together. Consequently, even in the presence of positively charged ions in aqueous solution, the negatively charged phosphate groups on nucleic acid backbone forms relatively weak electric force, which prevents them from precipitating out of solution. This is in principle similar to weakening of electric force, which normally holds salt crystals together by way of ionic bonds in the presence of water, thereby separating ions from the crystal, and spread through solution. The addition of EtOH increases the stability of ionic interactions between positively charged ions and negative charged phosphate residues on nucleic acid backbone, thereby allowing nucleic acid precipitation. The principle behind EtOH precipitation is that EtOH being less polar than water with a much lower dielectric constant (24.3 at 25°C) disrupts the screening of charges by water (i.e., exposes negatively charged phosphate groups of nucleic acids by depleting the hydration shell from them), makes it much easier for positively charged ions to interact with phosphates, and makes the nucleic acid less hydrophilic, causing it to drop out of solution. In other words, the repulsive forces between the polynucleotide chains are reduced to an extent that a precipitate can form. Thus, if sufficient EtOH (~64% of the solution) is added, the electrical attraction between phosphate groups and any positively charged ions present in solution becomes strong enough to form stable ionic interactions.

Based on the principle described, it is best to recover nucleic acids from aqueous solutions by standard precipitation with 2–3 volumes of absolute EtOH (calculated after addition of salt) in the presence of monovalent cations. The most commonly used cations for this purpose are 0.3 M sodium acetate (pH 5.2), 0.2–0.5 M sodium chloride, and 2–2.5 M ammonium acetate (Table 1.1). These positively charged ions neutralize the negative charge on nucleic acid, making it far less hydrophilic and therefore much less soluble in water. Note that these ions should be removed at a later stage as these may interfere with the downstream applications of isolated DNA.

After EtOH precipitation, the next step is to recover the precipitated nucleic acids. This is done either by spooling or centrifugation at high speeds. Often a successful preparation of nucleic acids results in a very thick solution that does not need to be concentrated any further. With a thick solution of nucleic acid, EtOH is layered on top of the sample, causing

Table 1.1 Salt Solutions Used in Ethanol Precipitation of Nucleic Acids

Salt	Stock solution (M)	Final concentration (M)	Important remarks
Sodium acetate	3.0 (pH 5.2)	0.3	Stock solution (3.0 M) is added one-tenth the volume of aqueous phase to get a final concentration of 0.3 M; 0.3 M sodium acetate (pH 5.2) is used for most routine precipitations of DNA and RNA.
Ammonium acetate (NH_4OAc)	5.0	2.0–2.5	Stock solution (5.0 M) is added in a volume equal to that of aqueous phase to get a final concentration of 2.5 M; It is frequently used to reduce the coprecipitation of unwanted contaminants (e.g., dNTPs or oligosaccharides) with nucleic acids. For example, two sequential precipitations of DNA in the presence of 2.0 M ammonium acetate result in the removal of >99% of the dNTPs from preparations of DNA; It is also the best choice when nucleic acids are precipitated after digestion of agarose gels with agarase (for details see Chapter 6). This is because the use of ammonium ions reduces the possibility of coprecipitation of oligosaccharide digestion products; It is used frequently for the removal of unincorporated nucleotides following a DNA labeling reaction; It is not used when the precipitated nucleic acid is to be phosphorylated, as bacteriophage T4 polynucleotide kinase is inhibited by ammonium ions.
Sodium chloride (NaCl)	5.0	0.2	0.2 M should be used if the DNA sample contains SDS. The detergent remains soluble in 70% ethanol; It is not as soluble as NH_4OAc, NaOAc, or LiCl in EtOH-water, or isopropanol water.
Lithium chloride (LiCl)	8.0	0.8	Stock solution (8.0 M) is added one-tenth the volume of aqueous phase to get a final concentration of 0.8 M; It is frequently used when high concentrations of EtOH are required for precipitation (e.g., when precipitating RNA). LiCl is very soluble in ethanolic solutions and is not coprecipitated with the nucleic acids; Small RNAs (tRNAs and 5S rRNAs) are soluble in solutions of high ionic strength (without ethanol), whereas large RNAs are not. Because of this difference in solubility, precipitation in high concentrations of LiCl (0.8 M) can be used to purify large RNAs.

molecules to precipitate at the interface. A spectacular trick is to push a glass rod through the EtOH into the nucleic acid solution. When the rod is removed, nucleic acid molecules adhere to it and are pulled out (spooled out) of the solution in the form of a long fiber. In case the aqueous phase is in large volumes or the nucleic acid preparation is dilute, it becomes important to consider methods for increasing the concentration of nucleic acids. Thus, for precipitation of nucleic acids from dilute solutions, the organic (EtOH) and aqueous phases are mixed by vortexing when isolating small DNA molecules (<10 kbp) or by gentle shaking when isolating DNA molecules of moderate size (10–30 kbp). When isolating large DNA molecules (>30 kbp), precautions must be taken to avoid shearing, for example, using large bored pipette tips for transfer from one tube to another. If SDS is present, it can be readily removed in the supernatant by adding NaCl to a final concentration of only 0.2 M or by using LiCl (0.5 M) before adding the alcohol.

After centrifugation, the supernatant is discarded and the pellet (nucleic acid) is washed with 70% EtOH (its volume should be sufficient to at least cover the pellet and wet the sides of the tube) and air-dried (by leaving the tube open on the laboratory bench for a few minutes or in speed vac). Note as the pellet formed after EtOH precipitation contains both nucleic acids and salts and as nucleic acid molecules and salt form aggregates, demonstrating dramatically reduced solubility in alcohol, washing with 70% EtOH is important. Washing with 70% EtOH removes much of the salt present in the leftover supernatant or bound to DNA pellet and often loosens the precipitates from the wall of the tube, however, it is

not sufficiently aqueous to redissolve DNA. DNA pellets should not be allowed to dry out completely, as it leads to denaturation of DNA and make it harder to resuspend. Moreover, it has been observed that slightly damp DNA pellets with EtOH dissolve more readily than completely dried ones. When dealing with small amounts (ng or pg) of DNA, it is prudent to save the ethanolic supernatant from each step (i.e., precipitation with absolute EtOH and washing with 70% EtOH) until the entire DNA has been recovered. The air-dried pellet (nucleic acids) is redissolved or rehydrated in an appropriate volume of 1 X TE buffer (pH 8.0) or sterile water. The aqueous solution of nucleic acids thus obtained is treated with RNase, extracted with phenol, and the DNA is precipitated with EtOH in the presence of salt. Besides the recovery of nucleic acids, EtOH precipitation is used for desalting and has the added advantage of leaving short-chain and monomeric nucleic acid components in solution. Hence ribonucleotides produced by RNase treatment are lost at this stage. The recovered pellet is washed with 70% EtOH, air-dried, and redissolved in TE buffer (pH 8.0). The resulting solution contains only DNA, which is stored at 4°C. Note that for long-term storage, DNA is stored as suspension in EtOH at –20°C.

The concentration of EtOH and salt added, time and temperature of incubation with EtOH, and the length and speed of centrifugation play a significant role in the process. The concentration of salt added should be appropriate, as too much of salt results in large amount of salt coprecipitating with nucleic acid and too little results in incomplete recovery of nucleic acid. EtOH precipitations are commonly performed at –70°C, –20°C, 0°C, or room temperature for periods of 5 minutes to overnight. The optimal incubation time depends on the length and concentration of DNA. For low DNA concentrations, higher final concentrations of EtOH, longer precipitations (for 1 hour to overnight), lower temperatures (–20°C to –70°C), and longer centrifugation times (up to 30 min) may be required for efficient recovery. This is because at lower temperatures, the viscosity of alcohol is greatly increased and centrifugation for longer times may be required to effectively pellet the precipitated DNA. The efficiency of precipitation for small concentrations of DNA may be increased by incubation at –70°C, but these reactions should be brought to 0°C before centrifugation. Due to increase in the viscosity of solution, some researchers are of the view that at lower temperatures precipitation efficiency is lowered and the highest precipitation efficiency is achieved at room temperature. However, when possible degradation at room temperature is considered, it is probably best to incubate nucleic acid preparation on wet ice or lower temperatures. Thus some researchers recommend that nucleic acids at concentrations as low as 20 ng/ml are precipitated at 0–4°C, so incubation for 15–30 min on ice is sufficient. Some researchers have also shown that this precipitation period is unnecessary and recommend direct centrifugation immediately after the addition of alcohol. During centrifugation, as the precipitated nucleic acids have to move through EtOH solution to the bottom of the tube, the time and speed of centrifugation have a significant effect on the recovery rates of nucleic acids. Smaller fragments and higher dilutions (lower nucleic acid concentrations) require longer and faster centrifugation as compared with larger fragments and lesser dilutions. Thus, for very small lengths and low concentrations overnight incubation is recommended. In such cases, the recovery of small quantities of nucleic acids is improved with the addition of carriers or coprecipitants, for example, yeast tRNA (10–20 µg/ml), glycogen (50 µg/ml), or linear polyacrylamide (10–20 µg/ml). Note that the carriers are inert substances that are insoluble in ethanolic solutions and form precipitates that trap the target nucleic acids. During centrifugation, carriers generate a visible pellet that facilitates handling of the target nucleic acids. Yeast tRNA is inexpensive, but it cannot be used for precipitating nucleic acids that are to be used as substrates in reactions catalyzed by polynucleotide kinase or terminal transferase. This is because the termini of yeast tRNA are excellent substrates for these enzymes and compete with the termini contributed by the target nucleic acid. Glycogen is used as a carrier when nucleic acids are precipitated with 0.5 M ammonium acetate and isopropanol. Glycogen does not compete with the target nucleic acids in subsequent enzymatic reactions. However, it interferes with the interactions between DNA and proteins. Linear polyacrylamide is an efficient neutral carrier for precipitating picogram amounts of nucleic acids with EtOH. Note that in the absence of a carrier, DNA concentrations as low as 20 ng/ml form a precipitate that can be quantitatively recovered by centrifugation in a microfuge, however, when lower concentrations of DNA or very small fragments (<100 nucleotides in length) are processed, more extensive centrifugation may be necessary to cause the pellet of nucleic acid to adhere tightly to the centrifuge tube. Centrifugation at 1,00,000 *g* for 20–30 min allows the recovery of picogram quantities of nucleic acid in the absence of carrier. In general, the length of time of centrifugation is more important for precipitating DNA than chilling the solution in –20°C or –70°C freezer.

Precipitation with isopropanol EtOH can also be replaced with isopropanol. This is advantageous because its precipitation efficiency is higher making 0.6–1.0 v/v (just half the volume of EtOH) enough for precipitation, although higher concentrations may be helpful when the DNA is at low concentration. Moreover, precipitation is done at room temperature, allows coprecipitation of fewer salts, and the

pellet adheres less tightly to the tube. Isopropanol precipitation is more effective than EtOH in separating primers from PCR products. However, many salts are less soluble in isopropanol than in EtOH, hence a second 70% alcohol rinse of the pellet is recommended to more efficiently desalt the DNA pellet or else salts that are more soluble in isopropanol are preferred. Moreover, as isopropanol is less volatile than EtOH, more time is required for air-drying in the final step. Similar to EtOH precipitation procedure, subsequent desalting of the DNA pellet involves rinsing in 70% alcohol, recentrifugation, air-drying, and resuspension in appropriate buffer or water.

Extraction with butanol followed by EtOH precipitation DNA samples may also be concentrated by extraction with solvents such as secondary butyl alcohol (isobutanol) or *n*-butyl alcohol (*n*-butanol). This results in a reduction in the volume of DNA preparation to the point where the DNA can be recovered easily by precipitation with EtOH. This is because some of the water molecules are partitioned into the organic phase and by carrying out several cycles of extraction, the volume of a DNA solution can be reduced significantly. Thus slightly more than one volume of isobutanol (or *n*-butanol) is added and the solution is vortexed vigorosly and centrifuged to separate the two phases. The upper butanol phase is discarded and extracted once with water-saturated diethyl ether to remove residual isobutanol (or *n*-butanol). The nucleic acids are then precipitated with EtOH as described earlier.

Ultrafiltration Another alternative to alcohol precipitation for the concentration and desalting of DNA solutions is ultrafiltration. In this method, nucleic acid solution is forced under pressure or by centrifugal force through a semipermeable membranous disk. Aqueous medium and small solute molecules pass through the semipermeable membrane, while large molecules are retained, thus a more concentrated nucleic acid solution is left behind. The ultrafiltration membranes are available with different pore sizes and hence the technique can be used to separate different sized nucleic acids. By selecting ultrafiltration device (e.g., Microcon cartridge from Millipore) with a nucleic acid cut-off (the size of the smallest particle that cannot penetrate the membrane) equal to or smaller than the molecular size of the nucleic acid of interest, efficient desalting and concentration of nucleic acid samples can be achieved. This method does not require any phase change and is particularly useful for dealing with very low concentrations of nucleic acids. The Microcon cartridge is first inserted into a vial and to concentrate (without affecting salt concentration), 500 µl sample of DNA (or RNA) is pipetted into reservoir and centrifugation is performed for the recommended time, not exceeding the *g* force. To exchange salt, proper amount of appropriate diluent is added to bring the concentrated sample to 500 µl. The reservoir is removed from the vial, inverted into a new vial, centrifuged at 500–1,000 *g* for 2 min, and nucleic acids are recovered. Note that ultrafiltration does not change the buffer composition. The salt concentration in a sample concentrated by spinning in a Microcon is the same as that in the original sample. For desalting, the concentrated sample is diluted with water or buffer to its original volume and spun again. This is called discontinuous diafiltration and is used to remove salt by the concentration factor of the ultrafiltration. For example, if a 500 µl sample containing 100 mM salt is concentrated to 25 µl (20% concentration factor), 95% of the total salt in the sample is removed. Rediluting the sample to 500 µl in water brings the salt concentration to 5 mM. Concentrating to 25 µl once again removes 99% of the original total salt. The concentrated sample now contains only 0.25 mM salt. For more complete removal of salt, an additional redilution and spinning cycle removes 99.9% of the initial salt content.

Dialysis Another alternative to EtOH precipitation is dialysis, a process that separates molecules according to size through the use of semipermeable membrane (dialysis tubing or bag) containing pores of less than macromolecular dimensions. Small molecules (lower than cut-off value), such as salts, small biochemicals, and water diffuse across the membrane (through pores) driven by the concentration differential between the solutions on either side of the membrane but the passage of larger molecules is blocked. Dialysis allows desalting and concentration of DNA preparation. The technique is also used for the removal of CsCl after isopycnic ultracentrifugation and purification of DNA fragments from agarose gel (for details see Chapter 6). The most commonly used dialysis material is cellophane (cellulose acetate), although several other substances such as cellulose and colloidon are similarly employed. These are available in a wide variety of molecular weight cut-off values that range from 0.5 to 500 kDa. Besides the size of the molecule, the exact permeability of a solute is also dependent on the shape of the molecule, its degree of hydration, and its charge. These parameters are in turn influenced by the nature of the solvent, its pH, and its ionic strength. Before use, the dialysis tubing is pretreated in the given order: cutting into appropriately sized pieces (10–20 cm), washing for 10 min in a large volume of 2% w/v sodium bicarbonate and 1 mM Na_2EDTA (pH 8.0), thorough rinsing with distilled water, boiling for 10 min in 1 mM Na_2EDTA (pH 8.0) [or else autoclaving at 20 psi (i.e., 1.40 kg/cm^2) for 10 min on liquid cycle in a loosely capped jar filled with water], cooling, and storage at 4°C in submerged condition. After this treatment, the tubing is always handled with gloves, and before use, it is washed inside and outside with sterile distilled water. For concentrating DNA in the preparation, the DNA solution is transferred into a standard cellulose acetate dialysis bag with the help of a wide-bored pipette or cut tip. The dialysis bag is then placed on a bed of solid sucrose and additional sucrose is

packed on top of the bag (this packing is best done at 4°C on a piece of aluminium foil spread on the bench in a cold room). Dialysis is allowed to proceed until the volume of the fluid in the dialysis bag is reduced by a factor of 5–10. The outside of the bag is rinsed with TE buffer (pH 8.0) to remove all of the adherent sucrose. The solution of DNA is gently massaged to one end of the bag and then the tubing is clamped just above the level of the fluid with a dialysis clip. The sample is dialyzed against large volume (4 l) of TE buffer (pH 8.0) for 16–24 hours with at least two changes of buffer. This method works more efficiently as compared to concentration in ultrafiltration devices, or in colloidon bags, and results in smaller losses of DNA.

Another variant called drop dialysis is used to remove low molecular weight contaminants from DNA in solution when high molecular weight DNA is to be used as a template in DNA sequencing reactions or if restriction enzymes fail to digest the DNA to completion. The method involves spotting of a drop (~50 µl) in the center of a Millipore Series V membrane (0.025 µm) floating shiny side up on 10 ml of sterile water in a petridish with a 10 cm diameter. DNA is then dialyzed for 10 minutes. The drop is then removed to a clean microfuge tube and aliquots of dialyzed DNA are used for restriction enzyme digestion and/or DNA sequencing experiments.

Isolation and Purification of Genomic DNA from Higher Eukaryotes

Total cell DNA from plants and animals is also required if the aim of the genetic engineering experiment is to clone genes from them. The basic steps in DNA purification are the same with all the organisms, however, some modifications need to be introduced to take account the special features of the cells being used. In this section, these special requirements are discussed.

Cell Lysis The major modifications from the basic approach discussed above are very likely to be needed at the cell breakage stage, as the chemicals used for disrupting bacterial cells (e.g., lysozyme) do not usually work with plant and animal cells. Like bacteria, a combination of physical and chemical disruption techniques is employed with plants and animals. Often physical techniques, such as grinding frozen source material with a mortar and pestle or homogenizer, are more efficient. For chemical disruption, specific plant wall degradative enzymes available for most cell wall types are used. On the other hand, most animal cells have no cell wall at all and can be lysed simply by treating with a detergent.

Removal of Contaminating Biomolecules and Recovery of DNA Another important consideration is the biochemical content of the cells from which DNA is being extracted. With most bacteria, the main biochemicals present in a cell extract are proteins, DNA, and RNA, so phenol extraction and/or protease treatment followed by removal of RNA with RNase or the application of monophasic reagents leaves a pure DNA sample. These treatments are, however, not sufficient to get pure DNA if the cells also contain significant quantities of other biochemicals, for example, polysaccharides or polyphenolics. Plant tissue is especially notorious for being a difficult source from which to isolate high-quality DNA with good yield. Problems encountered in the isolation and purification of high molecular weight DNA from plant species include degradation of DNA due to endonucleases, coisolation of highly viscous polysaccharides, and inhibitor compounds like polyphenols and other secondary metabolites released during cell disruption, which directly or indirectly interfere with the enzymatic reactions in the downstream reactions. Polysaccharides are not removed by phenol extraction, but these form complexes with nucleic acids during tissue extraction and coprecipitate during subsequent alcohol precipitation steps. Depending on the nature and the quantity of these contaminants, the resulting alcohol precipitates can be gelatinous and difficult to dissolve. Thus, DNA purification procedure from plant tissues should include a step for the removal of carbohydrates. The method used for removal of contaminating carbohydrates makes use of a detergent cetyl trimethyl ammonium bromide (CTAB), which forms an insoluble complex with nucleic acids. When CTAB is added to a plant cell extract, the nucleic acid–CTAB complex precipitates, leaving carbohydrate, protein, and other contaminants in the supernatant. The precipitate is then collected by centrifugation and resuspended in 1 M NaCl, which causes the complex to break down. The nucleic acids are then concentrated by EtOH precipitation and the RNA removed by RNase treatment. The CTAB method for purification of DNA from plant source is presented in Figure 1.3.

To eliminate browning effect of polyphenols and secondary metabolites leached out from plant tissues, polyvinyl pyrrolidone (PVP) may be included in the extraction buffer. PVP inhibits polyphenol oxidase and binds to the phenolic compounds, which are then eliminated by EtOH precipitation.

Other DNA Recovery Methods

The traditional method employed the isolation of genomic DNA bands from the cesium chloride (CsCl) or cesium trifluoroacetate (CsTFA) isopycnic gradient after ultracentrifugation (for details see Chapter 6). In the CsCl density gradient, protein and RNA contaminants are separated on the basis of differences in buoyant densities (for details see Chapter 6). For high yields of DNA, cells are lysed using a detergent and the lysate is alcohol precipitated. Resuspended DNA is mixed with CsCl and EtBr and centrifuged for several hours. The DNA band is collected from the centrifuge tube, extracted with isopropanol to remove EtBr, and then precipitated with EtOH to recover the DNA. This method allows the isolation

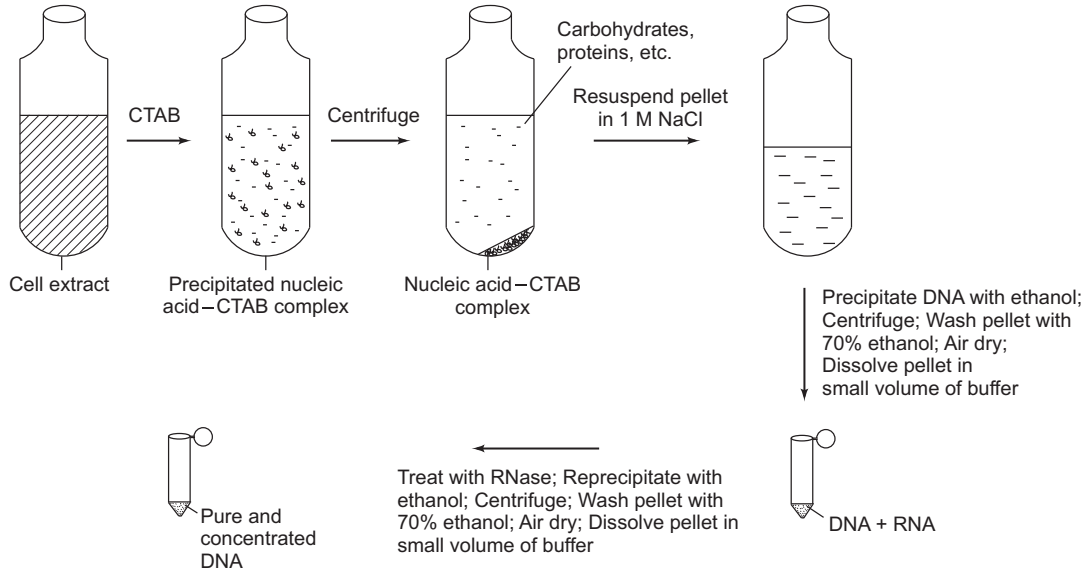


Figure 1.3 CTAB method for purification of plant DNA

of high-quality DNA, but is time-consuming, labor-intensive, and expensive (an ultracentrifuge is required), making it inappropriate for routine use. Moreover, this method uses toxic chemicals (e.g., CsCl) and is also impossible to automate. This traditional method has been replaced by conventional precipitation method involving EtOH or isopropanol as described above. The EtOH precipitation method is rapid, virtually foolproof, efficient, and even minute amounts of DNA can be quantitatively precipitated. For even faster and simpler recovery of DNA, certain variants or alternatives are presently available. These are summarized below.

Silica-based Spin Column Chromatography The spin column method involves selective adsorption of nucleic acids to silica in the presence of high concentrations of chaotropic salts. Earlier, nucleic acids were purified using glass powder or silica beads under alkaline conditions in the presence of chaotropic agents such as sodium iodide or sodium perchlorate. Note that silica structures are much more effective as packing material because these are etched into the microchannels during its fabrication by soft lithography. Moreover, these silica structures are easier to use in highly parallelized designs. This technique was later improved by using guanidinium salts (e.g., guanidinium thiocyanate or guanidinium hydrochloride) as chaotropic agent. The use of beads was later changed to minicolumns. Thus silica may be directly added to the cell extract or more conveniently packed in a chromatography column. In silica-based spin column chromatography, DNA binds tightly to silica particles in the presence of guanidinium thiocyanate, from where it is eluted at a later stage (Figure 1.4a). The principle behind the technique is that a chaotrope denatures biomolecules by disrupting the shell of hydration around them, thereby allowing positively charged ions to form salt bridges between the negatively charged silica and the negatively charged DNA backbone in high salt concentration (Figure 1.4b). Use of optimized buffers in the lysis procedure ensures that only DNA is adsorbed to silica while contaminating biochemicals such as cellular proteins and metabolites remain in solution, and pass through the column upon washing with high salt buffer and EtOH. DNA is ultimately eluted from the silica using a low-salt buffer or water, which destabilizes the interactions between DNA molecules and silica. Thus the method provides an easy way of recovering DNA from the denatured mix of biochemicals. In this technique, alcohol precipitation and resuspension of DNA (which is often difficult if the DNA has been over-dried) are not required. The technology is simpler and more effective than other methods where precipitation or extraction is required. The method is reliable, fast, inexpensive, and provides high-throughput isolation of high-quality DNA. Genomic DNA isolated using silica-based spin column technology is up to 50 kbp in size, with an average length of 20–30 kbp. Moreover, the isolated DNA is suitable for downstream applications such as Southern blotting, PCR, real-time PCR, random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), and amplified fragment length polymorphism (AFLP) analyses (for details see Chapter 18). Currently, the main manufacturer of silica-based columns for purification of DNA is Qiagen and hence these columns are often referred to as Qiagen columns. For example, DNeasy tissue kits (Qiagen) are designed on the basis of silica-based column purification technique, which allows rapid isolation of pure total DNA

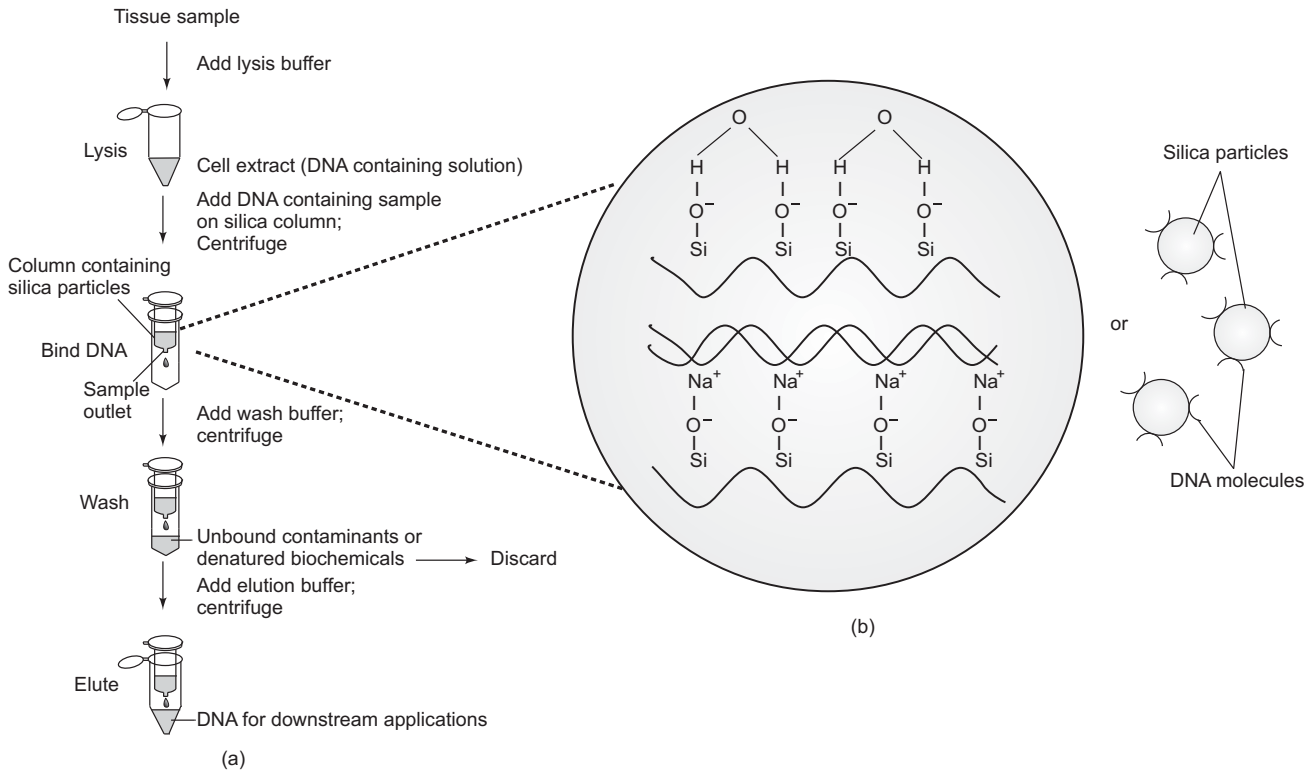


Figure 1.4 Silica-based spin column chromatography for purification of DNA; (a) Procedure for purification on silica-based spin column, (b) Magnified view showing binding of DNA to silica particles in the presence of guanidinium thiocyanate

(genomic/viral/mitochondrial) from a wide variety of sample sources, including fresh and frozen animal cells and tissues, yeasts, and blood. This method has, however, a few associated drawbacks. The beads and resins are highly variable and hence each loading of a microchannel can result in a different amount of packing, consequently changing the amount of DNA adsorbed to the channel. The technology is also not suitable for isolating genomic DNA >50 kbp in size.

Magnetic Bead Capture Technology This technology eliminates the conventional spin steps and employs magnetic bead capture for the isolation of nucleic acids. Using this technique, DNA is isolated with both high yield and purity. Moreover, the high magnetite content of the beads and the ease of handling make this technology highly adaptable to automation for isolation of DNA from different starting volumes as well as different sources. A general scheme for the isolation process is shown in Figure 1.5.

Anion Exchange Chromatography Solid phase anion exchange chromatography is based on the interaction between the negatively charged phosphates of the nucleic acid and positively charged surface molecules on the substrate (i.e., solid support). In this method, DNA is allowed to bind to the solid support under low salt conditions and the impurities such as RNA, cellular proteins, and metabolites are washed away using medium salt buffers. High-quality DNA is then eluted using a high salt buffer and DNA is recovered by alcohol precipitation. The advantage of anion exchange chromatography is that it completely avoids the use of toxic substances and can be used for different throughput requirements as well as for different scales of purification. The isolated DNA is sized up to 150 kbp, with an average length of 50–100 kbp. Moreover, the DNA preparation is suitable for all downstream applications.

DNA Isolation Kits Isolation and purification of DNA from different sources and different tissues can also be done using commercially available kits. These kits are based on one or the other methods described earlier and make the process of DNA isolation and purification rapid and simple.

Analysis of DNA Preparations

The DNA isolated by any of the above-mentioned techniques is then evaluated for quality (purity and integrity) and quantity. The techniques used for the purpose include the following:

Agarose Gel Electrophoresis and Visual Inspection Under UV Illumination The purity and integrity of DNA prepara-

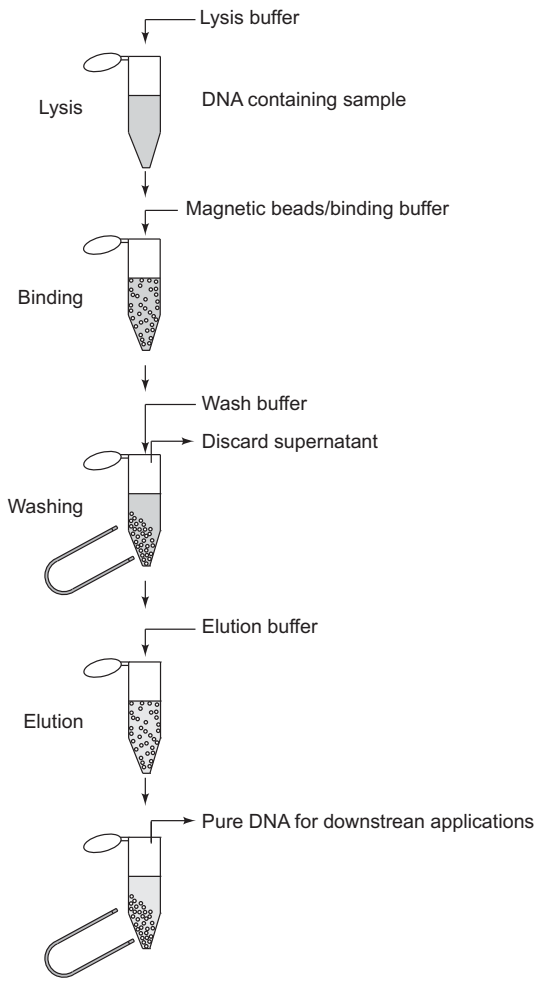


Figure 1.5 Magnetic bead capture technology for the purification of DNA

tion isolated from any source is assessed in the form of DNA–EtBr complex by agarose gel electrophoresis followed by UV illumination (for details see Chapter 6). Conventional agarose gel electrophoresis is used to determine the pattern of DNA fragments ranging in size from 50 bp to ~20 kbp, while pulse field gel electrophoresis (PFGE) is used for DNA molecules >20 kbp in size. Polyacrylamide gels are effective for visualizing and separating fragments of DNA ranging in size between 5 and 500 bp. Figure 1.6 shows the results of a conventional agarose gel electrophoresis of purified total DNA preparation. With this technique, the relative concentration of DNA solution with that of DNA size standard can also be estimated.

Agarose Gel Electrophoresis and Fluorescence Image Analysis For a gene cloning experiment, it is crucial to know the exact concentration of DNA solution. At present, fluorescence image analysis of fluorescent dye-stained DNA on gel is possible using image analysis software. This software affords an opportunity for quantifying nucleic acid concentration by digitizing a fluorescence image of a sample (after labeling with SYBR Green, Gel Star, or SYBR Gold) that has been run out on a gel and comparing it to a known mass standard (e.g., ϕX174 or λ DNA). Image analysis software counts the pixels that make up the image of all bands and smears in a lane and then compares the value for each unknown sample against a mass standard. This approach is also used to determine the quantity of synthesized cDNA. The method is rapid and usually quite accurate.

Visual Inspection of DNA Spots on Agarose Plate Under UV Illumination Another nonspectrophotometric method for determination of DNA concentration involves making an estimate by simple visual inspection. The method involves

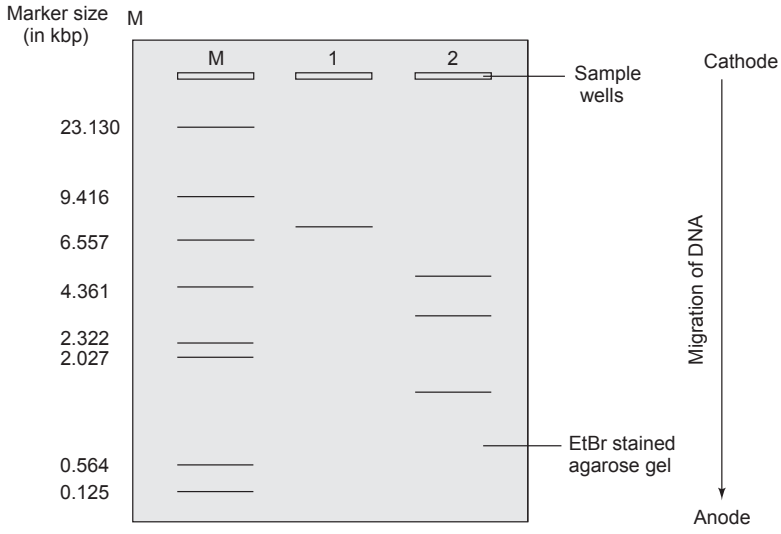


Figure 1.6 A pattern of EtBr-stained DNA bands under UV light [Lane M : DNA size marker; Lane 1 : Purified DNA; Lane 2 : DNA fragments.]

spotting dilutions of known amounts of standard DNA and samples of unknown concentrations on an agarose plate containing a fluorescent dye. On UV irradiation, the dye fluoresces in proportion to the amount of DNA present and comparing fluorescence intensity between the known and unknown samples gives a rough estimate of the amount of DNA in the unknown. Performing this somewhat imprecise assay requires the preparation of agarose plates containing 0.5 µg/ml EtBr or SYBR dyes. Note that the samples should be adsorbed into the agarose before assessing the fluorescence intensity of the dye. For ease of quantification, dilutions of the standards and the test samples are best applied in 5 µl (or less) aliquots, which avoids sample diffusion. These plates can be observed directly on the surface of a UV transilluminator or irradiated from above with a handheld UV monitor. Although this method is outdated, it may be a useful teaching tool when funds are limited.

UV Absorption Spectrophotometry For an accurate measurement of the concentration of DNA in pure solutions (i.e., without significant amounts of contaminants such as proteins, phenol, agarose, or RNA), the method of choice is UV absorbance spectrophotometry. The principle behind the technique is that the purines and pyrimidines in nucleic acids absorb UV light and the extinction coefficients of nucleic acids are the sum of the extinction coefficients of each of their constituent nucleotides. For large molecules, where it is both impractical and unnecessary to sum up the coefficients of all the nucleotides, an average extinction coefficient is used. For small molecules such as oligonucleotides, it is best to calculate an accurate extinction coefficient from the base composition. Because the concentrations of oligonucleotides are commonly reported as mmole/l, a millimolar extinction coefficient (ε) is conventionally used in the Beer–Lambert equation [ε = A (15.3) + G (11.9) + C (7.4) + T (9.3)], where A, G, C, and T are the number of times each nucleotide is represented in the sequence of the oligonucleotide. The numbers in parentheses are the molar extinction coefficients for the given deoxynucleotide at pH 7.0. Note that the extinction coefficient of nucleic acids is affected by ionic strength and pH of the solution. Accurate measurements of concentration can be made only when the pH is carefully controlled and the ionic strength of the solution is low. The method is rapid, simple, accurate, and nondestructive. Because the absorption spectra of nucleic acids are maximal at 260 nm, absorbance data for DNA (or RNA) are almost always expressed in A_{260} units. The OD at 260 nm allows calculation of the concentration of DNA (or RNA) in the sample. A_{260} of 1 corresponds to ~50 µg/ml for ds DNA, 40 µg/ml for ss DNA or RNA, and ~33 µg/ml for ss oligonucleotides. The amount of UV radiation absorbed by a solution of DNA (or RNA) is directly proportional to the amount of DNA (or RNA) in the sample. As the Beer–Lambert law is valid at least to an $A_{260} = 2$, the concentration of a solution of DNA (or RNA) is therefore easily calculated by simple interpolation using following equations:

Concentration of ds DNA (µg/ml)
$= A_{260} \times$ dilution factor $\times$ 50

Concentration of ss DNA (or RNA) (µg/ml)
$= A_{260} \times$ dilution factor $\times$ 40

Concentration of ss oligonucleotides (µg/ml)
$= A_{260} \times$ dilution factor $\times$ 33

The ratio between the readings at 260 and 280 nm (A_{260}/A_{280}) provides an estimate of the purity of the DNA (or RNA) preparation. Pure preparations of DNA have A_{260}/A_{280} value of 1.8. If there is significant contamination with protein or phenol, the A_{260}/A_{280} will be less than 1.8 and accurate quantitation of the amount of DNA will not be possible. Note that aromatic amino acids of proteins absorb strongly at 280 nm.

UV-induced Fluorescence Emitted by DNA–EtBr Complex Although it is possible to estimate the concentration of solutions of DNA and oligonucleotides by measuring their absorption at a single wavelength (260 nm), this is not a good practice. This is because UV absorption spectrophotometry suffers from certain drawbacks, for example, it is comparatively insensitive, most laboratory spectrophotometers require at least 1 µg/ml DNA concentrations to obtain reliable estimates of A_{260}, it cannot readily distinguish between DNA and RNA, and it cannot be used with crude preparations of DNA. Thus, if there is not sufficient DNA (<250 ng/ml) or the DNA may be heavily contaminated with other substances that absorb UV irradiation, accurate spectrophotometric analysis will be impeded. Moreover, A_{260}/A_{280} as a measure of purity of isolated DNA is also not a good choice. This is because DNA absorb so strongly at 260 nm that only a significant level of protein contamination will cause a significant change in the ratio of absorbance at the two wavelengths. To overcome these problems, an alternative method for quantitating ds DNA is spectrophotometric analysis of DNA–EtBr complex. The method involves estimation of the amount of DNA through analysis of UV-induced fluorescence emitted by EtBr molecules intercalated into the DNA. The formation of complexes between DNA and EtBr can be observed with the naked eye because of the large metachromatic shift in the absorption spectrum of the dye that accompanies binding. The original maximum at 480 nm (yellow-orange) is shifted progressively to 520 nm (pink) with a characteristic isosbestic point at 510 nm. Because the amount of fluorescence is proportional to the total mass of DNA, the quantity of DNA in the sample is estimated by comparing the light emitted by the sample at 590 nm (fluorescent yield) with that of series of standards. This provides a simple, faster,

and more sensitive way to estimate the concentration of a sample of DNA, and as little as 1–5 ng of DNA is readily detected by this method.

It is suggested that the absorbance of the sample should be measured at several wavelengths since the ratio of absorbance at 260 nm to the absorbance at other wavelengths is a good indicator of the purity of the preparation. Note that significant absorption at 260 nm indicates contamination by phenolate ion, thiocyanates, and other organic compounds, whereas absorption at higher wavelengths (330 nm and higher) is usually caused by light scattering and indicates the presence of particulate matter.

Fluorometric Quantitation of DNA using Hoechst 33258 Hoechst 33258 is a class of *bis*-benzimidazole fluorescent dye that binds nonintercalatively and with high specificity into the minor groove of ds DNA. Unlike EtBr, Hoechst dyes are cell-permeant. Like many other nonintercalative dyes, Hoechst 33258 binds preferentially to A-T rich regions of the DNA helix. Hoechst 33258 in free solution has an excitation maximum at ~356 nm and an emission maximum at 492 nm. However, when bound to DNA, Hoechst 33258 absorbs maximally at 356 nm and emits maximally at 458 nm. After binding to ds DNA, the fluorescent yield increases from 0.01 to 0.6, with the $\log_{10}$ of the intensity of fluorescence increasing in proportion to the A+T content of DNA. The fluorescent yield of Hoechst 33258 is approximately three-fold lower with ss DNA. Owing to this property, Hoechst 33258 can be used for fluorometric detection and quantification of ds DNA in the solution. Hoechst 33258 is preferred to EtBr for this purpose because of its greater ability to differentiate ds DNA from RNA and ss DNA.

Measuring the concentration of DNA using fluorometry is simple and more sensitive than spectrophotometry and allows the detection of nanogram quantities of DNA. The concentration of DNA in the unknown sample is estimated from a standard curve constructed using a set of reference DNAs (10–250 ng/ml) whose base composition is the same as the unknown sample. In this assay, DNA preparations of known and unknown concentrations are incubated with Hoechst 33258 fluorochrome. Absorption values for the unknown sample are compared with those observed for the known series and the concentration of the unknown sample is estimated by interpolation. The assay can only be used to measure the concentration of DNAs whose sizes exceed ~1 kbp, as Hoechst 33258 binds poorly to smaller DNA fragments. Measurements should be carried out rapidly to minimize photobleaching and shifts in fluorescence emission due to changes in temperature. Either a fixed wavelength fluorometer (e.g., Hoefer model TKO 100) or a scanning fluorescence spectrometer (e.g., Hitachi Perkin–Elmer model MPF-2A) is used. Binding of Hoechst 33258 is adversely influenced by pH extremes, presence of detergents near or above their critical micelle concentrations, and salt concentrations above 3.0 M. Note that fluorometry assays with Hoechst 33258 do not work at extremes of pH and are affected by both detergents and salts. Assays are therefore usually carried out under standard conditions [0.2 M NaCl, 10 mM Na_2EDTA, pH 7.4]. However, two different salt concentrations are required to distinguish ds DNA from ss DNA and RNA. The concentration of Hoechst 33258 (M_r = 533.9) in the reaction should be kept low (5×10^{-7} M to 2.5×10^{-6} M), since quenching of fluorescence occurs when the ratio of dye to DNA is high. However, two concentrations of dye are sometimes used to extend the dynamic range of the assay. All DNAs and solutions should be free of EtBr, which quenches the fluorescence of Hoechst 33258. However, as Hoechst 33258 has little affinity for proteins or RNA, measurements can be carried using cell lysates or purified preparations of DNA.

Membrane Immobilization and Chromogenic Assay Another method for determining DNA concentration involves immobilization of a small aliquot of a DNA sample onto a membrane followed by a chromogenic (colorimetric) assay. Briefly, dilutions of DNA samples are spotted onto the membrane, followed by a series of washes and incubation in chromogenic substrate. DNA concentration is determined by comparing the resulting color intensity to standards. This technique permits efficient quantitation of very small quantities of ss DNA, ds DNA, RNA, and oligonucleotides, down to as little as 1 ng in aqueous buffer. Based on this technique, DNA Dipsticks kit (Invitrogen) does not register individual nucleotides and are thus quite handy for quantitating dilute template samples for PCR amplification and for monitoring the progress of the reaction.

Colorimetric Analysis using Diphenylamine (DPA) DPA indicator confirms the presence of DNA. The principle behind the procedure is that when a DNA solution is heated (at ≥95°C or simply in boiling water bath) with DPA under acid conditions, a blue compound is formed with a sharp absorption maximum at 595 nm. Note that DPA solution is prepared in a mixture of glacial acetic acid and concentrated sulfuric acid. In acid conditions, the straight chain form of 2-deoxypentose sugar of DNA is converted to the highly reactive β-hydroxylevulinyl aldehyde, which reacts with DPA to produce a blue-colored complex. DNA concentration is determined by measuring the intensity of absorbance of the blue-colored solution at 595 nm with a spectrophotometer, and comparing it with a standard curve of known DNA concentrations. In DNA, only the deoxyribose of the purine nucleotide reacts so that the value obtained represents half of the total deoxyribose present.

1.3.2 Isolation and Purification of Total RNA

RNA isolation is required for purification of mRNA, followed by cDNA synthesis to be used for the preparation of cDNA library. Besides cloning, other applications of RNA or mRNA include analysis of gene expression, poly (A) RNA selection, *in vitro* translation, reverse transcriptase-PCR (RT-PCR), northern hybridization, dot/slot blotting, RNase protection assay, S1 nuclease protection assay, RNase mapping, and primer extension experiments. Efficient methodologies have been empirically derived to accommodate the expedient isolation of RNA. In general, these methods yield cytoplasmic RNA, nuclear RNA, or mixtures of both, commonly known as cellular RNA. The conditions required for a successful isolation of RNA also differ significantly both between species and for the same species when grown under different environmental conditions. Similar to DNA isolation strategies, RNA isolation also requires same basic steps, i.e., cell lysis, removal of contaminating biomolecules, and recovery of RNA, however, a seemingly endless list of successful permutations on a few fundamental RNA extraction techniques exists. In the following sections, details of basic steps involved in RNA isolation strategies and the rationale behind each step is discussed. Note that throughout the RNA isolation and purification procedure, five Rs of molecular biology should be kept in mind. These are: rapid, representative, reproducible, RNase-free, and reliable.

Cell Lysis and Preparation of Cell Extract

The first step in RNA recovery process is cell disruption. The reagents included in the cell lysis buffer are chaotropic and denaturing agent [e.g., guanidinium salts such as guanidinium isothiocyanate (GIT) and guanidinium chloride], denaturant [e.g., β-mercaptoethanol (β-ME), dithiothreitol (DTT), and urea], deproteination agent [e.g., proteinase K, phenol, and chloroform], detergent [e.g., SDS, *N*-lauryl sarcosine (sarkosyl), Triton X-100, and sodium triisopropyl naphthalene sulfonate], chelating agent [e.g., sodium 4-amino salicylate and Na_2EDTA], and RNase inhibitor [e.g., heparin, iodoacetate, dextran sulfate, polyvinyl sulfate, macaloid, vanadyl ribonucleoside (VDR), and cationic surfactant].

Roles of Various Ingredients of Cell Lysis Buffer The inclusion of these reagents in the cell lysis buffer targets two aims, *viz*., membrane solubilization and complete inhibition of RNase activity, and the selection of these reagents depends on the RNA population intended to be isolated, i.e., RNA population from the subcellular compartment(s) of interest.

Membrane solubilization The method of cell lysis determines the extent of subcellular disruption of the sample. For example, a lysis buffer that is used successfully with tissue-cultured cells may be entirely inappropriate for whole tissue samples. The method by which membrane solubilization is accomplished dictates whether additional steps are required to remove DNA from the RNA preparation and whether compartmentalized nuclear RNA and cytoplasmic RNA species can be purified independently of one another. Note that it is difficult, if not impossible, to determine the relative contribution of RNA from the nucleus and the cytoplasm once RNA from these two subcellular compartments have been copurified. A particular lysis procedure must likewise demonstrate compatibility with ensuing protocols once it has been recovered from the lysate.

Complete inhibition of RNase activity Setting up of conditions for the control of RNase activity is equally important. This includes speedy extraction and purification, purging RNase from reagents and equipments, and controlling RNase activity in a cell lysate. Some lysis reagents inhibit nuclease activity, while other lysis reagents require additional nuclease inhibitors for the elimination of RNase activity that are compatible with the lysis buffer. The methods used for prevention or inactivation of endogenous and exogenous RNase activities at different steps are summarized in Exhibit 1.4.

Exhibit 1.4 Ribonuclease (RNase) in RNA isolation and purification

Ribonucleases (RNases) are a family of enzymes that are small, very stable, and omnipresent (present in virtually all the living cells). These enzymes are present both in endogenous (intrinsic or internal) and exogenous (extrinsic or external) sources. These enzymes, if present, easily and quickly degrade RNA during extraction, purification, as well as storage. RNA exhibits short half-life due to its intrinsically labile nature and its degradation is further compounded by the ubiquity of these RNases. Further the resilient nature of RNases aggravates the problem, as these renature quickly following treatment with most denaturants even after boiling. This property is attributable to reformation and maintenance of their tertiary configurations by virtue of four disulfide bridges. RNases have minimal cofactor requirements and are active over a wide pH range.

To maintain the stability of the RNA before, during, and after its isolation from the cell (i.e., to isolate full-length RNA and store it for longer time periods), following precautions and preventive measures to overcome the problem of both endogenous and exogenous RNases should be taken into account. Extrinsic sources of potential RNase contamination must be identified and neutralized from the onset of the experiment. These include, but are not limited to, bottles and containers in which chemicals are packaged, RNase containing water, gel-boxes and combs, bacteria and molds present on airborne dust particles causing contamination of buffers, the hair or beards of investigator, and oil from user's fingertips, etc. These extrinsic sources of RNase lead to accidental contamination of an RNA preparation. Beyond the potential for accidental contamination of an RNA preparation with

RNases from the laboratory environment, one must be acutely aware of the fact that intracellular RNases, normally sequestered within the cell, are liberated on cellular lysis. The problems with exogenous RNase can be entirely avoided by vigilant use of prophylactic measures and the prudent application of common sense. Following precautions and preventive measures should be taken to eliminate the risk of contamination with both exogenous and endogenous RNases.

Wear Gloves Finger greases are notoriously rich in RNases and are generally accepted as the single greatest source of RNase contamination. Hence, rule number one for controlling extrinsic RNase activity is to wear rubber gloves throughout the isolation and purification process, i.e., during the preparation of reagents, handling of reagents and apparatus, and especially during the actual RNA extraction procedure. Furthermore, changing of gloves several times during the course of RNA related experiment is recommended because door-knobs, micropipettors, handles of refrigerator's door, and telephone receivers are also potential sources of RNase contamination.

Proper Handling and Maintenance of Cleanliness and Aseptic Conditions Proper microbiological sterile techniques should be observed for handling and preparation of reagents. Tubes should be kept closed when handling RNA samples. As reagents are used repeatedly, contamination must be prevented during opening and closing the reagent tubes. If the barrel or the metal ejector of the automatic pipettor comes in contact with the sides of tubes, it becomes a very efficient vector for the dissemination of RNase and hence one should be very careful in pipetting solutions.

Reserve Separate Glassware, Plasticware, Equipments, and Reagents for RNA Work Materials or stock solutions that have been used for purposes other than RNA isolation and purification in the laboratory should not be used for RNA work. Rather items of glassware, batches of plasticware, electrophoresis devices, and buffers that are to be reused should be reserved exclusively for RNA work and not be in general circulation in the laboratory. A special set of automatic pipettors for use when handling RNA should also be kept aside. Chemicals should be set aside for RNA work and should be handled with disposable spatulas or RNase-free spatula or more safely dispensed by tapping the bottle rather than using a spatula. Solutions/buffers should be stored in small aliquots of suitable volumes, rather than drawing repeatedly from the stock bottles, and each aliquot used once should be discarded. Although such actions may at first seem excessive, these may well preclude the accidental introduction of RNases and facilitate recovery of the highest possible quality RNA.

Use RNase-free and Disposable Plasticware Conical tubes, both polypropylene and polystyrene, are considered sterile if already capped and racked by the manufacturer. Individually wrapped serological pipettes are always preferred for RNA work because no special pretreatment is required. Sterile (marked as tissue culture sterile) disposable tips and microfuge tubes certified by a reputable manufacturer to be free of RNase should be used preferably. These materials are generally RNase-free and thus do not require pretreatment to inactivate RNase. Bulk packed polypropylene products are potential sources of RNase contamination, mainly due to handling and distribution from a single bag. This pertains to microfuge tubes and polypropylene micropipette tips because these can become contaminated and, in turn, contaminate stock solutions. Thus, any plastic product that comes into contact with an RNA sample at any time, either directly or indirectly, and that can withstand autoclaving should be autoclaved [sterilization using moist heat done under high temperature (121°C) and high pressure (15 psi)]. These microfuge tubes are then handled only with gloves and set aside exclusively for RNA work. To reduce the chances of contamination, it is best to use sterile forceps or gloved hands (use fresh gloves) for the distribution of small items of plasticware from original packages to laboratory racks or beakers. After distribution, these can be covered with aluminium foil and autoclaved. For the manipulation of organic extraction buffers, which typically contain mixtures of the organic solvents phenol and chloroform, individually wrapped borosilicate glass pipettes are strongly preferred.

Make Glassware, Plasticware, Equipments, and Reagents RNase-free Clearly, it is incumbent on the investigator to ensure that equipment, glassware, and plasticware are purged of RNases from the outset of the experiment. The temperature and pressure generated during the autoclaving cycle for the sterilization of solutions, laboratory plastics, and other apparatuses do not ensure complete elimination of RNase activity. Hence, some other strategies as mentioned below should be adopted.

- All reagents should also be maintained RNase-free at all times. Pre-made solutions that are certified as being RNase-free are widely available and it is worthwhile to invest in such solutions. Alternatively, stock solutions and buffers prepared in the laboratory can be treated directly or indirectly with diethyl pyrocarbonate (DEPC) (Figure A);

Moreover, all reagents should be prepared in high-purity biochemical quality water and in RNase-free glassware and plasticware. For indirect treatment with DEPC, solutions are made in DEPC–water (DEPC–water is prepared by treating water with 0.1% DEPC for at least 1 hour at 37°C and removed by autoclaving for 15 min at 15 psi on liquid cycle), followed by autoclaving of the prepared solution. Alternatively, DEPC is removed from large volumes of DEPC–water by boiling for 1 hour in a fume hood. On the other hand, for direct DEPC treatment, 0.1% DEPC is added to solutions and treatment is allowed for the required length of time, and later DEPC is removed by autoclaving. Thus wherever possible, solutions are treated with 0.1% DEPC for at least 1 hour at 37°C, or overnight at room temperature, and then autoclaved for 15 min at 15 psi on liquid cycle. For the preparation of solutions and buffers, DEPC is added to a final concentration of 0.1%, shaken for several hours on an orbital platform, or stirred vigorously with a magnetic stirrer for 20–30 min, and following

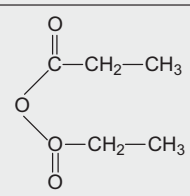


Figure A Structure of diethyl pyrocarbonate (DEPC)

this treatment, DEPC is destroyed completely by autoclaving. Complete removal of DEPC is further promoted by rapidly stirring the hot solutions with a nuclease-free magnetic stir bar. Frequently, if the autoclaving time is not adequate, the distinctive odour of the residual DEPC can be noticed. Alternatively, some solutions are maintained at 60°C overnight. Certain solutions such as those containing SDS, NP-40, or NaOH are not autoclaved and these nonautoclavable components are added to complete the solution formulation after the other components have been autoclaved as needed, or otherwise made RNase-free. DEPC is not added to any buffer containing mercaptans or primary amine groups, with which DEPC is reactive. Perhaps the most common buffers to which DEPC exposure is to be avoided are the Tris buffers [tris (hydroxymethyl) aminomethane]. This is because the hydrolysis of DEPC to CO_2 and ethanol is greatly accelerated by Tris and other amines, which themselves become consumed in the process. Note that in aqueous solution, DEPC is hydrolyzed rapidly to CO_2 and ethanol; half-life in phosphate buffer is ~20 min at pH 6.0 and 10 min at pH 7.0. Thus for RNase decontamination of Tris buffers, DEPC–water is used and the solution is autoclaved again. Those buffers consisting of chemicals that demonstrate or are known to have DEPC incompatibility are filtered twice through a nitrocellulose membrane to remove RNase activity, and other trace proteins. As DEPC is too difficult to remove from reagents and interferes with PCR and other reactions, a suitable alternative to DEPC treatment of water to render it RNase-free is to simply purchase RNase-free water;

- Nondisposable plasticware should also be treated before use to make them RNase-free. Plasticware is therefore filled with DEPC–water and allowed to stand for 1 hour at 37°C, or overnight at room temperature. The items are rinsed several times with DEPC-treated water and then autoclaved for 15 min at 15 psi on liquid cycle. Besides DEPC–water, commercially available products that inactivate RNase upon contact (e.g., RNaseZap from Ambion Inc.) can also be used to remove RNase contamination from pipettes or table-tops;
- Baking the glassware in a dry heat oven is a very effective method for purging glassware of RNase activity. Hence glassware should be cleaned scrupulously, rinsed with RNases-free water, and then baked for 3–4 hours or overnight at 300°C. It is important to note here that not all laboratory implements can withstand the heat generated in a dry heat oven. Such glassware (e.g., COREX® tubes) should be filled with DEPC–water and allowed to stand for 1 hour at 37°C or overnight at room temperature. The items are rinsed several times with DEPC-treated water and then autoclaved for 15 min. at 15 psi on liquid cycle; and
- The electrophoresis unit, including the comb, casting tray, and electrophoresis chambers, are treated to remove RNase contamination by rinsing and soaking in DEPC–water. For RNase decontamination, after cleaning of electrophoresis devices with detergent solution, these are rinsed in water, dried with ethanol, and then filled with a 3% solution of H_2O_2. H_2O_2 is an inexpensive and powerful oxidizing agent that can be used to render a surface nuclease-free by soaking in it for 20–30 min, then rinsing with copious amounts of sterile water. More concentrated form of H_2O_2 (e.g., 30%) commonly available from standard chemical supply companies should not be used, as concentrated solutions of H_2O_2 are extremely dangerous and may cause irreparable damage to acrylic gel-box components. After 10–20 min at room temperature, electrophoresis tank is washed thoroughly with sterile water. Note that the unit should never be exposed to DEPC, because acrylic is not resistant to DEPC.

Discard Old Chemicals By careless use of aseptic technique, buffers may become contaminated with bacteria or other microorganisms. As the growth of these microorganisms is not usually visible to the naked eye, solutions that are even suspected of being contaminated should be discarded. All stocks and working solutions in the laboratory should be labeled with a predetermined expiration date in addition to the date of preparation. A sad consequence of the use of out-of-date solutions in the laboratory is the unintentional introduction of RNase from microorganisms that may have taken up residence in the solution bottles.

Immediately Freeze Harvested Tissue The use of fresh tissue is preferred for RNA isolation. However, as it is easier to collect tissues in advance, the interest lies in preservation of the samples for days, weeks, or even months after tissue collection without sacrificing the integrity of the RNA. One way is to store tissue samples in RNAlater® tissue storage:RNA stabilization solution (Ambion). Dissected tissue or collected cells are simply dropped into the RNAlater® solution at room temperature. The solution permeates the cells, stabilizing the RNA. The samples are then stored at 4°C. Samples can be shipped on wet ice or even at room temperature if shipped overnight. Note that the use of RNAlater® for tissue storage is compatible with most RNA isolation procedures. Alternative to storage in RNAlater®, harvested tissues or cell pellets may be frozen immediately in liquid nitrogen and stored at –70°C to –80°C. When stored in this way, RNA can be purified up to a year later. Tissues stored in RNAlater® or in liquid nitrogen are simply removed and processed by homogenization or other mechanical apparatus in the lysis buffer specified in the RNA isolation procedure. Some investigators homogenize fresh tissue in guanidinium buffer on receipt and then freeze the homogenate at –80°C, continuing the RNA isolation procedure at a later date. Purified RNA is most stable when stored as an ethanol precipitate at –80°C. Under these conditions, the investigator can confidently store RNA for several months or even longer because the half-life of RNA is a direct function of the biological source.

Speedy Extraction Cellular RNases (endogenous RNases) are free to initiate degradation of the RNA that the investigator is attempting to isolate, unless these are inhibited without delay. Hence intrinsic RNases should be inactivated as quickly as possible at the very first stage in the RNA purification process (i.e., extraction). This demands speed as well as precautions while handling. Once the endogenous RNases have been destroyed, the immediate threat to the integrity of the RNA is greatly reduced and purification can proceed at a more graceful pace.

Inclusion of Chaotropic Agent and Reducing Agents in RNA Lysis Buffer RNA lysis buffer may combine the disruptive and protective properties of guanidine thiocyanate (GTC) and β-ME to inactivate RNases present in cell extracts. GTC is a chaotropic agent that disrupts nucleoprotein complexes, allowing RNA to be released into solution and isolated free of protein. β-ME is a reducing agent that breaks disulfide bonds in proteins, thereby inactivating RNases. The combination of GTC and β-ME makes the RNA lysis buffer a potent inhibitor of RNase activity.

8-Hydroxyquinoline, a partial RNase inhibitor, when included in lysis buffer, may optimize RNA purification efficiency. It chelates heavy metals that can cause RNA degradation when present with RNA for extended periods.

Application of RNase Inhibitors Creating an RNase-free environment or elimination of RNase activity is essential during the entire process. Because endogenous RNase activity varies tremendously from one biological source to the next, the degree to which action must be taken to inhibit RNase activity is a direct function of cell or tissue type. Knowledge of the extent of intrinsic RNase activity is derived from the literature and personal experience. Moreover, the method selected for controlling RNase activity must be compatible with the cell lysis procedure. It is also important to note that the method of nuclease inhibition must support the integrity of the RNA throughout the subsequent fractionation or purification steps, some of which can be quite time-consuming. In addition, the reagents used to inhibit RNase activity must be easily removed from purified RNA preparations so as not to interfere with subsequent manipulations. One such strategy is addition of RNase inhibitors in the lysis buffer, reaction buffers, and in the standard preparation of reagents intended for RNA work and storage. At the time of extraction, RNase inhibitors are most often added to relatively gentle lysis buffers when subcellular organelles (especially nuclei) have to be purified intact. Unfortunately, the inhibitors do not offer perfect protection because all of the available inhibitors inhibit only some part of the existing RNases. Consequently, these substances cannot be handled without working cleanly. RNase inhibitors are also used to protect mRNA in cDNA synthesis, and *in vitro* transcription/translation system. Note that RNase inhibitors are heat-sensitive and are stored at –20°C. The RNase inhibitors may be composed of mixtures of proteins that inhibit a more or less broad spectrum of RNases. There are two types of RNase inhibitors, *viz*., specific and nonspecific inhibitors. The examples of these two types are discussed below:

• **Vanadyl ribonucleoside complexes** Amongst the category of specific RNase inhibitors are vanadyl ribonucleoside (VDR) complexes, protein inhibitors, and macaloid. VDR complexes were developed in the mid-1970s as a means of controlling RNase activity when using relatively gentle methods to support cellular lysis. VDR consists of complexes formed between the oxovanadium ion and any or all of the four ribonucleosides in which vanadium takes the place of phosphate. These complexes then function as transition state analogs that bind to many RNases and inhibit their activity almost completely. The four vanadyl ribonucleoside complexes are added to intact cells and used at a concentration of 10 mM during all stages of RNA extraction and purification. The resulting mRNA is isolated in a form that can be directly translated in frog oocytes and can be used as a template in some *in vitro* enzymatic reactions (e.g., reverse transcription of mRNA). However, vanadyl ribonucleoside complexes strongly inhibit translation of mRNA in cell-free systems and must be removed from the mRNA by multiple extractions with phenol (equilibrated with 0.01 M Tris–HCl, pH 7.8) containing 0.1% 8-hydroxyquinoline. Vanadyl ribonucleoside complexes are available from several commercial suppliers. In the absence of VDR, RNase-mediated cleavage of the phosphodiester backbone of RNA results in the transient formation of a dicyclic transition state intermediate, which is subsequently opened up by reaction with a water molecule. VDR is not widely used any longer due to associated drawbacks, *viz*., (i) RNA isolation compounds that are more efficient inhibitors of RNase have been developed; (ii) VDR in trace quantities can inhibit *in vitro* translation of purified mRNA; and (iii) VDR inhibits reverse transcriptase and hence contraindicated for RT-PCR and the numerous permutations thereof. Thus, if purified RNA is to be subjected to either of these applications, the use of VDR is not recommended. 8-Hydroxyquinoline chelates heavy metals, making it very useful for removing VDR complexes from cell lysates. The color imparted by the 8-hydroxyquinoline changes from yellow to dark green upon binding VDR. When the phenol phase of the extraction buffer remains yellow, it indicates that all VDR has been removed .

• **Human placental inhibitor** Many RNases bind tightly ($K_i \approx 3 \times 10^{10}$) to a protein isolated from human placenta forming equimolar, noncovalent complexes that are enzymatically inactive. *In vivo*, the protein is probably an inhibitor of angiogenin, an angiogenic factor whose amino acid sequence and predicted tertiary structure are similar to that of pancreatic RNase. The inhibitor, which is sold by several manufacturers, is stored at –20°C in a solution containing 50% glycerol and 5 mM DTT. Preparations of the human placental inhibitor that have been frozen and thawed several times or stored under oxidizing conditions should not be used, as these treatments may denature the protein and release bound RNases. The inhibitor is therefore not used when denaturing agents are used to lyse mammalian cells in the initial stages of extraction of RNA. However, it should be included when more gentle methods of lysis are used and should be present at all stages during the subsequent purification of RNA. Fresh inhibitor should be added several times during the purification process, as it is removed by extraction with phenol. The inhibitor requires sulfydryl reagents for maximal activity and does not interfere with reverse transcription or cell-free translation.

• **Macaloid** Macaloid is clay that has been known for many years to adsorb RNase. The clay is prepared as slurry that is used at a final concentration of 0.015% w/v in buffers used to lyse cells. The clay, together with its adsorbed RNase, is removed by centrifugation at some stage during the purification of RNA (e.g., after extraction with phenol).

• **Diethyl pyrocarbonate** The most common example of nonspecific RNase inhibitor is DEPC. It is a highly reactive alkylating agent, which acts as a potent and efficient RNase inhibitor. It destroys the enzymatic activity of RNase chiefly by ethoxyformylation of histidyl groups. DEPC is commonly used to chemically inactivate trace amounts of RNases that may contaminate solutions, glassware, and plasticware to be used for the preparation of nuclear RNA. Note that endogenous RNases in cell suspensions or lysates are not inhibited by addition of DEPC. After treatment of glassware, plasticware, and solutions, DEPC has to be removed before use in RNA isolation procedure. This is because in addition to reacting with histidine residues in proteins, DEPC also carbethoxylates single stranded nucleic acids, and has strong affinity for adenosine nucleotides. It forms alkali-labile adduct with the imidazole ring N7 of unpaired purine, resulting in cleavage of the glycosidic bond and generation of an alkali-labile abasic site. Even trace amounts of residual DEPC result in chemical modification

of adenine, thereby changing the physical properties of RNA and compromising its utility for *in vitro* translation and other applications, including standard blot analysis and PCR. However, the ability of DEPC-treated RNA to form DNA:RNA hybrids is not seriously affected unless a large fraction of the purine residues have been modified. Note that due to ability of DEPC to modify unpaired adenine, DEPC is also carcinogenic and should be handled with extreme care. After treatment of glassware, plasticware, and solutions with DEPC for appropriate time period, it is removed readily by thermal degradation (autoclaving), leading to its decomposition into CO_2 and ethanol, both of which are quite volatile under the conditions of autoclaving. This is, however, problematic, because small amounts of these products lead to an increase in ionic strength, and lower the pH of unbuffered solutions. Samples of DEPC that are free of nucleophiles (e.g., water and ethanol) are perfectly stable, but even small amounts of these solvents can cause complete conversion of DEPC to diethylcarbonate. For this reason, DEPC should be protected against moisture. It should be stored under small aliquots in dry conditions and bottle should be allowed to reach temperature before being opened.

- **Other examples** Other examples of RNase inhibitors are heparin, iodoacetate, dextran sulfate, polyvinyl sulfate, and cationic surfactant, etc. Many RNase inhibitors are available commercially, for example, RNAsin from Promega, RNase Block from Stratagene, RNaseZap from Ambion Inc., Ribonuclease inhibitor from Clonetech, etc.

Sources: Farrell Jr RE (2005) *RNA Methodologies–A Laboratory Guide for Isolation and Characterization*, Elsevier Academic, New York, Appendix F; Tait RC (1997) *An Introduction to Molecular Biology*, Horizon Scientific Press; Sambrook J and Russel DW (2001) *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 7.2–7.12.

Classification of Cell Lysis Buffers Based on differences in membrane solubilization and RNase inhibition properties, lysis buffers typically fall into two categories:

Lysis buffers containing plasma membrane and subcellular organelle membrane solubilizing agents This category of lysis buffers includes harsh chaotropic agents, denaturing agents, or surfactants, which rapidly disrupt the plasma membrane as well as subcellular organelle membranes of the cells, solubilize their components, and simultaneously inactivate RNase (i.e., addition of exogenous inhibitors of RNase is not required). Examples included in this category are GIT, guanidinium chloride, SDS, sarkosyl, urea, water-saturated phenol, or chloroform. Thus the extract obtained after direct cell lysis using such agents is a representative of final accumulation of RNA in the cell, as it includes the RNA from both the nucleus and cytoplasm. It is also known as steady state RNA.

Guanidinium salts function by destroying the 3-D structure of proteins and convert most proteins to a randomly coiled state. The chloride salt of guanidinium is a strong inhibitor of RNase, but it is not powerful enough as a denaturant to allow the extraction of intact RNA from tissues rich in RNase. On the other hand, GIT is a stronger chaotropic agent and contains potent cationic and anionic groups that form strong H-bonds. It is used in the presence of a reducing agent (e.g., β-ME) to break intramolecular protein disulfide bonds and in the presence of a detergent such as sarkosyl to disrupt hydrophobic interactions. Note that the efficiency of protein denaturation (including disruption of RNase) is enhanced by the inclusion of β-ME in the lysis buffer. Another common reducing reagent, DTT, is not used in combination with guanidinium salts because of their chemical reactivity. After cell lysis, partitioning of RNA, DNA, and protein in the resulting lysate is required. This is achieved in three ways: (i) Isopycnic ultracentrifugation using CsCl or CsTFA; (ii) Acid–phenol treatment; and (iii) Application of monophasic reagent containing acidified phenol, guanidinium or ammonium thiocyanate, and a phenol solubilizer (Figure 1.7). In the first and original method, guanidinium lysate is subjected to time-consuming isopycnic ultracentrifugation through CsCl/CsTFA. The original method has later been modified to eliminate ultracentrifugation step, for example, by inclusion of LiCl in the procedure or by the application of hot phenol (note that the hot phenol procedure results in high yields of RNA within hours). The second method involves guanidinium-based whole cell lysis followed by acid–phenol extraction. This method offers certain benefits, such as reduction in the length of RNA isolation step, collection of increased number of samples at a time, reduction in RNA loss, and purification of RNA from smaller sample sizes. Although such extraction buffers are easily prepared in the molecular biology laboratory by mixing water-saturated phenol with an acidic solution of sodium acetate, premixed monophasic formulations of phenol and guanidinium thiocyanate (e.g., TRIzol and TRI reagent) are readily available commercially. On phase separation, RNA is retained in the aqueous phase, while DNA and proteins partition into the interface and organic phase. Then RNA is recovered by precipitation with isopropanol and collected by centrifugation. In these procedures, RNA is efficiently isolated from as little as 1 mg of tissue or 10^6 cells usually in less than 1 hour. The single-step method yields the entire spectrum of RNA molecules, including small (4–5S) RNAs. The yield of RNA depends on the type of tissue used for isolation. Typically, 100–500 μg of total RNA is obtained from 100 mg of muscle and up to 800 μg of RNA from 100 mg of liver tissue. The yield of total RNA from cultured cells is in the range of 50–80 μg (fibroblasts, lymphocytes), or 100–120 μg (epithelial

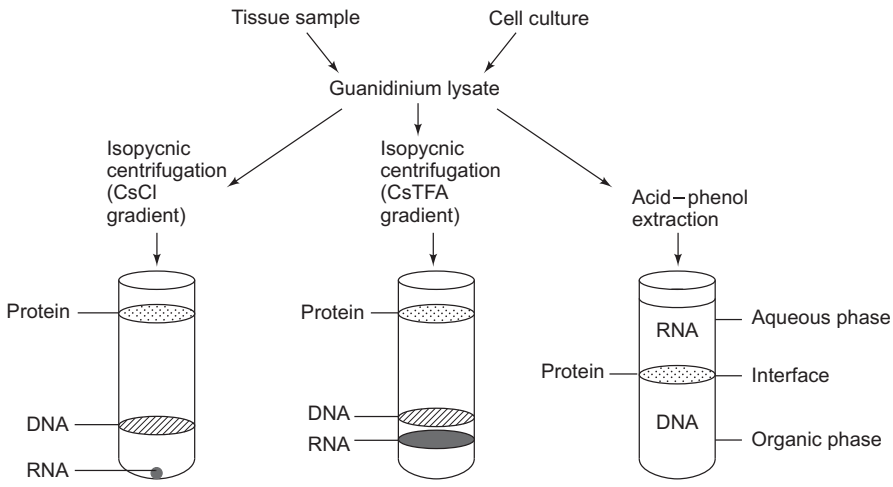


Figure 1.7 Methods for fractionation of total cellular RNA following cellular disruption with guanidinium buffer

cells) per 10^7 cells. The third method involves application of monophasic reagent to optimize the speed and extent of RNase inactivation (for details see Section 1.3.1). Several commercial RNA isolation kits are based on these monophasic lysis reagent, which are commercially available with a variety of trade names, for example, TRIzol reagent (Life Technologies), TRI reagent (Molecular Research Center), Isogen (Nippon Gene, Toyama, Japan), RNAzol and RNA-Stat-60 (Tel-Test Inc.) that allow partitioning of RNA in the aqueous phase from where it is precipitated with isopropanol and is further purified by chromatography on oligo (dT)-cellulose columns, and/or used for northern blot hybridization, reverse transcription, or RT-PCR. The yield of total RNA depends on the tissue or cell source, but it is generally 4–7 μg/mg starting tissue or 5–10 μg/10^6 cells.

Lysis buffers containing only plasma membrane solubilizing agents The abundance of specific steady-state RNA species may certainly be interpreted as an indication of how specific genes are modulated, but it does not furnish any information either about the rate at which these RNA molecules are transcribed or about their stability (or half-life) in the cell. This is the major disadvantage of evaluating the steady-state RNA alone. It is certainly clear that the cellular biochemistry responds to environmental change not only by modification of the rate of transcription but also through the processing efficiency of precursor RNA, the efficiency of nucleocytoplasmic transport, the stability of RNA in the cytoplasm, and the translatability of salient messages. In contrast to the protocols for the analysis of steady-state transcripts, transcription rate studies require the isolation of intact nuclei that are able to support elongation of, and label incorporation into, RNA molecules whose transcription was initiated at the time of cellular lysis. This family of techniques is known collectively as the nuclear runoff assay. Hence, another category of cellular lysis involves gentle solubilization of plasma membrane, while maintaining nuclear integrity using agents such as hypotonic Nonidet P-40 (NP-40) or Igepal CA-630 (Sigma) lysis buffers. Intact nuclei, other organelles, and cellular debris are then removed from the lysate by differential centrifugation. The reliability of this approach is often dependent on the inclusion of nuclease inhibitors in the lysis buffer and careful attention to the handling and storage of RNA so purified. Note that NP-40-based method is more appropriate for RNase-poor tissues or cells because the inactivation of the RNase is less efficient.

Removal of Contaminating Biomolecules

After successful cell lysis, cellular extract is obtained that contains biomolecules depending on the species under consideration. However, for various molecular biology studies, it is of paramount importance to isolate high-quality, high-purity, undegraded, or intact RNA. Hence, next step is the purification of RNA and the following strategies are adopted for removal of contaminating proteins, DNA, and carbohydrates.

Deproteination For complete removal of proteins from cellular lysate, RNA isolation procedure also includes a step for deproteination. Removal of proteins may be accomplished by any one or the combination of following procedures: (i) Digestion of the sample with the enzyme proteinase K; (ii) Repeated extraction with mixtures of organic solvents such as phenol and chloroform; (iii) Salting out of proteins; and (iv) Solubilization in guanidinium buffers (for details see Section 1.3.1). Note that although phenol denatures proteins efficiently, it does not completely inhibit RNase activity and it is a solvent for RNA molecules that contain long tracts of poly (A).

Prevention of Genomic DNA Contamination The complete separation of DNA from RNA preparation is technically challenging because both these nucleic acids share structural similarities and essentially all existing methods for isolating total RNA copurify genomic DNA, which interferes with sensitive detection methods. The elimination of DNA from nucleic acid preparation requires an additional step, which is however not always entirely effective. A common approach to eliminate DNA contamination is to treat the nucleic acid preparation with RNase-free DNase I (bovine pancreas), which is a double strand-specific endonuclease. After DNase digestion, RNA is reprecipitated with EtOH as described for precipitation of DNA in Section 1.3.1. Another method for recovery of RNA from a lysate that also contains DNA involves organic extraction at pH below 6.0. This drives DNA to the aqueous–organic interface, while RNA enters the aqueous phase. A yet another method is based on the application of guanidinium-based monophasic reagent. Besides, LiCl precipitation efficiently precipitates RNA, but not DNA, protein, or carbohydrate.

Elimination of Polysaccharides and Polyphenols Although osmotic cell lysis described above is one of the gentlest methods for cell disruption, it does not work well when carbohydrate-rich cell walls are present, as is the case with certain bacteria, fungi, and plant cells. These polysaccharides or polyphenolics form irreversible complexes with RNA during tissue extraction and coprecipitate during subsequent alcohol precipitation steps. Depending on the nature and quantity of these contaminants, the resulting alcohol precipitates can be viscous or gelatinous and difficult to dissolve. Moreover, an RNA solution contaminated with polysaccharides and/or polyphenols absorbs strongly at 230 nm and hence prevents an accurate quantitation of RNA preparation. Furthermore, RNA contaminated with polysaccharides or polyphenols is not suitable for cDNA synthesis, RT-PCR amplification, *in vitro* translation, or northern analysis. However, the problems associated with polysaccharides and polyphenolics can be rectified by including additional steps in the extraction and purification procedure, for example, treatment with hot phenol/SDS or hot borate.

Recovery of RNA

Recovery of RNA is the final step in RNA purification schemes. This is done by any of the following procedures:

Alcohol–Salt Precipitation Followed by Centrifugation The most versatile method for concentrating RNA is precipitation using various combinations of salt and alcohol. This is because nucleic acids and the salt that drives their precipitation form complexes with greatly reduced solubility in EtOH (or isopropanol). Thus, RNA is efficiently precipitated with 2.5–3.0 volumes of EtOH from solutions containing Na^+, K^+, Li^+, or NH_4^+ ions (Table 1.1). Similar to concentration of genomic DNA, carriers may be added for ease in visualization. Precipitated RNA is recovered by centrifugation at high speed, and the resulting pellet is washed with 70% EtOH and air-dried.

The rate of precipitation is temperature-dependent. Unlike the dramatic precipitation of genomic DNA, the precipitation of RNA is much slower, often requiring longer incubation periods at –20°C to ensure complete recovery, especially when using the NP-40 method. The choice among salts is determined on the basis of downstream application of isolated RNA as well as genome size or the complexity of organism. Since the potassium salt of dodecyl sulfate is extremely insoluble, potassium acetate is avoided if the precipitated RNA is to be dissolved in buffers containing SDS, for example, buffers that are used for chromatography on oligo (dT)-cellulose. For the same reason, potassium acetate is avoided if the RNA is already dissolved in a buffer containing SDS. LiCl is frequently used when high concentrations of EtOH are required for precipitation (i.e., for precipitating RNA). LiCl is very soluble in ethanolic solutions and is not coprecipitated with the nucleic acid. Small RNAs (tRNAs and 5S rRNAs) are soluble in solutions of high ionic strength (without EtOH), whereas large RNAs are not. Because of this difference in solubility, precipitation in high concentrations of LiCl can be used to purify large RNAs. LiCl precipitation offers major advantages over other RNA precipitation methods. It is the most appropriate method for the recovery of high molecular weight RNA to be used in cDNA library construction. Note that large RNAs (e.g., rRNAs and mRNAs) are insoluble in solutions of high ionic strength and can be removed by centrifugation, while small (<200 nucleotides) RNAs, for example, tRNA or 5S rRNA, are soluble or are not effectively precipitated. In other words, LiCl precipitation should be avoided if small RNAs are desired. It does not efficiently precipitate DNA, proteins, or carbohydrates. It is the method of choice for removing inhibitors of translation or cDNA synthesis from RNA preparations. It also provides a simple and rapid method for recovering RNA from *in vitro* transcription reactions. Precipitation of RNA with LiCl is also helpful in removing glycoproteins and yolky components from the preparations. Moreover, RNAs precipitated by this method give more accurate values when quantitated by UV spectrophotometry since LiCl is very effective at removing free nucleotides. However, LiCl is avoided when the RNA is to be used for cell-free translation or reverse transcription. This is because Li^+ ions inhibit initiation of protein synthesis in most cell-free systems and suppress the activity of RNA-dependent DNA polymerase.

Single-step Resin Binding Method An alternative to the alcohol precipitation is the single-step resin binding method, which employs RNA-binding resins [e.g., RNA Track™ resin, (Biotecx), and RNA MATRIX™ (BIO 101)]. The

method leads to enhancement of the purification of total RNA, and avoids conventional lengthy precipitation methods of EtOH or isopropanol. RNA purified by this method is free from impurities such as traces of guanidinium salts, phenol, etc., which might interfere with subsequent RNA applications. This method requires no additional carrier protein in the case of low-yield RNA samples such as serum. This method also eliminates the requirement for density gradient centrifugation after guanidinium-based lysis. Although the yield of isolated RNA by this procedure is 10–20% lower than the conventional alcohol precipitation methods, the purity of RNA is improved.

Silica-based Column Purification Another modified single-step approach to quickly isolate milligram quantities of pure total RNA with undetectable levels of contaminating DNA is silica-based column purification. This solid-phase purification method relies on the fact that the nucleic acids bind (adsorb) to the solid phase (silica) depending on the pH and the salt content of the buffer. Similar to RNA-binding resins, the application of silica filters precludes the requirement for density gradient centrifugation. Moreover, high-quality total RNA is isolated with minimal effort from a wide variety of samples. One such example is the PureYield™ RNA midiprep system (Promega Corp.), which uses PureYield silica-membrane technology to isolate intact total RNA ranging in size from less than 100 bases to greater than 20 kb. Up to 300 mg of plant or animal tissue, 5×10^7 tissue culture cells, 1×10^{10} bacterial cells, 5×10^8 yeast cells, or 20 ml of blood can be processed per column. A disadvantage of this technology is that an RNA column is typically unsuitable for purification of short (<100 nucleotides) RNA, such as small interfering RNA (siRNA) and microRNA (miRNA).

Glass Fiber-based RNA Purification Glass fiber-based RNA purification step is done either as part of the RNA isolation strategy or as an additional clean-up step after RNA isolation. This purification technique increases the purity of RNA samples and dramatically reduces the amount of 5S rRNA and tRNA in samples. Moreover, RNA yields are higher than those obtained with simple organic extraction methods. This strategy is included in RiboPure™-bacteria kit (Ambion). This kit combines an efficient glass bead and RNAWIZ™-mediated disruption step followed by a glass filter-based RNA purification for high yields of exceptionally pure bacterial RNA. The kit also includes DNA-free™ reagents for quick and simple DNase treatment of samples without organic extraction, alcohol precipitation, or column purification.

Isopycnic Ultracentrifugation The application of guanidinium-based lysis mandates a procedure for partitioning of RNA, DNA, and protein in the resulting lysate. This is done by various methods, one of which is to subject guanidinium lysates to isopycnic ultracentrifugation using CsCl/CsTFA. In CsCl density gradient centrifugation, only ultrapure nuclease-free preparations are used because CsCl has only a limited ability to inhibit RNase activity. If necessary, impure, solid CsCl may be baked at 200°C for 6–8 hours to remove residual RNase activity prior to exposure to RNA. CsTFA is a highly soluble salt that solubilizes and dissociates proteins from nucleic acids without the use of detergents. CsTFA is an excellent inhibitor of RNase and its use precludes removal of proteins from a sample by more traditional methods (e.g., phenol:chloroform, proteinase K). As compared to CsCl, CsTFA is more chaotropic, inhibits RNase to a greater extent, and shows greater solubility in EtOH, which expedites its removal following isopycnic ultracentrifugation of RNA. RNA can be banded or pelleted in CsTFA as desired, because solution densities up to 2.6 g/ml (or 2.6 g/cm^3) are possible using this reagent. Because the buoyant density of RNA in CsCl (1.8 g/ml) is much greater than that of other cellular components, rRNAs and mRNAs migrate to the bottom of the tube during ultracentrifugation. As long as the step gradients are not overloaded, proteins remain in the guanidinium lysate, while DNA floats on the CsCl cushion. This method is useful when ultrapure RNA is required and the RNA extracted from cultured cells and most tissues can be directly used in further experiments without any further treatment. Depending on the cell type, typical RNA yields range from 50 to 75 μg/10^6 cells. The method is especially useful for kinetic studies where samples have to be taken at several different time points because different samples can be stored in guanidinium buffer for several days before ultracentrifugation. On the other hand, this method is cumbersome and involves extensive labor and expensive instrumentation (ultracentrifuge with a rotor), generally limiting the number of samples that can be easily isolated simultaneously. Furthermore, small RNAs (e.g., 5S rRNA and tRNAs), which do not sediment efficiently during centrifugation through CsCl, should not be prepared by this method. Moreover, running gradients is generally a time-consuming procedure (two days for RNA purification) that is usually not preferred for mainstream molecular biology applications. Currently, several methods have been developed where ultracentrifugation time with CsCl has been reduced from 12–18 hours to 3 hours utilizing a Beckman TL-100 ultracentrifuge and TLS-55 rotor.

RNA Isolation Kits A large number of kits are now commercially available from various manufacturers that allow rapid and easy isolation of high-quality RNA from different sources in good yields. These kits work by combining one or the other lysis, decontaminating, and recovery procedures described in Section 1.3.2.

Other Methods Alternatives to alcohol precipitation for concentration and desalting of RNA solutions are ultrafiltration and dialysis (for details see Section 1.3.1). Similar to DNA,

magnetic bead technology and anion exchange chromatography can also be employed for the recovery of RNA (for details see Section 1.3.1).

Storage of RNA

The last step in every RNA isolation protocol, whether for total RNA or mRNA preparation, is to resuspend the purified RNA pellet. After painstakingly preparing an RNA sample, it is crucial that RNA be suspended and stored in a safe, RNase-free solution. Hence, for further use, RNA is stored in any of the following ways: (i) RNA precipitate is dissolved in deionized formamide and stored at –20°C. Note that formamide provides a chemically stable environment that protects RNA against degradation by RNases. Purified, salt-free RNA dissolves quickly in formamide up to a concentration of 4 mg/ml. At such concentrations, RNA samples are analyzed directly by gel electrophoresis and used for RT-PCR or RNase protection, saving time and avoiding potential degradation. When required, RNA is recovered by precipitation with four volumes of EtOH or by diluting the formamide four-fold with 0.2 M NaCl and then adding the conventional two volumes of EtOH. Note for long-term storage, RNA is dissolved in highly purified 100% formamide or a commercially available stabilized form of formamide known as FORMAzol (Molecular Research Center), in which RNA is stable for up to 2 years at –20°C. For northern blots, the RNA dissolved in formamide can be used directly although it must previously be precipitated using four volumes of EtOH for a cDNA synthesis reaction; (ii) RNA precipitate is dissolved in aqueous buffer, which minimizes hydrolysis of RNA (note that divalent cations catalyze the base hydrolysis of RNA) and is compatible with all of the common downstream applications of RNA. One such example is that of aqueous buffer containing 0.1–0.5% SDS in TE buffer (pH 7.6), followed by storage at –80°C. For downstream applications, SDS is removed by chloroform extraction and EtOH precipitation. Alternative storage solutions are TE buffer (pH 7.0)/RNA storage solution [Ambion; 1 mM sodium citrate (pH 6.4 ± 0.2); presence of low pH and sodium citrate as buffering and chelating agent minimizes base hydrolysis of RNA]/RNAsecure™ resuspension solution [Ambion; RNA pellet is resuspended in the RNAsecure™ resuspension solution and heated to 60°C for 10 minutes to inactivate RNases. A unique feature of RNAsecure™ is that reheating after the initial treatment reactivates the RNase-destroying agent to eliminate any new contaminants]/DEPC–water containing 0.1 mM Na_2EDTA (pH 7.5); and (iii) RNA precipitate can be stored as a suspension at –20°C in EtOH.

Once total RNA is isolated from the given source, mRNA is purified from the preparation, which is then used for cDNA library construction (for details see Chapter 5) or other applications.

Analysis of RNA Preparation

Before proceeding for downstream applications, the RNA preparation is analyzed for its integrity, purity, and concentration. The following methods find application in analysis of RNA preparation.

Denaturing Agarose Gel Electrophoresis The quality of total RNA preparation (i.e., rRNAs, mRNAs, and tRNAs) is verified by electrophoresis on denaturing (formaldehyde or glyoxal containing) agarose gel (for details see Chapter 6). Running a gel is also the best diagnostic tool to assess the integrity of RNA preparation, and the probable utility of RNA sample in downstream applications. Because rRNA is the most prevalent RNA (80–85% of cellular RNA), an aliquot of total cellular or total cytoplasmic RNA should electrophoretically resolve into two very distinct, easily observable bands (28S and 18S rRNAs in eukaryotic RNA preparation, 23S and 16S rRNAs in prokaryotic RNA preparation) present within the smear of mRNAs. The appearance of these discrete bands is a convincing evidence of the integrity or intactness of the sample. Samples in which there has been any degree of degradation usually fail to manifest the characteristic formation of rRNA bands and light smearing of the mRNA. Note that in high-quality mammalian RNA, the ratio of 28S:18S eukaryotic rRNAs (judged by the intensity of bands through EtBr staining) should be ~2:1, however, this ratio approaches 1 as one moves down the evolutionary ladder. This ratio gives information about the integrity of preparation and a ratio of 2:1 indicates that no gross degradation of RNA has occurred. In RNA samples that have been degraded, this ratio is reversed, as the 28S rRNA is characteristically degraded to 18S-like species. In addition to the 28S and 18S rRNAs, an intact RNA sample manifests its mRNA component as a significantly lighter smear that appears slightly above, between, and below the rRNAs. This is the normal appearance of cellular mRNA because of its extremely heterogeneous nature and because the mRNA is usually less than 3% of the total mass of RNA in the cell. To achieve maximum resolution of larger molecular weight RNAs, electrophoresis is often allowed to continue to the extent that the small 5S and 5.8S rRNA and tRNA species run off the distal edge of the gel and into the running buffer. This is acceptable because the 5S and 5.8S rRNA species are too small to be useful as molecular weight markers for cellular RNA because these generally migrate through the gel along with the tRNA at the leading edge of the electrophoretic separation in the 300–400 base range. An agarose gel showing pure preparation of total cell RNA and mRNA from eukaryotes is depicted in Figure 1.8. The electrophoretic profile of plant RNA from green tissue is often different from that observed on electrophoresis of animal cells. In addition to the large and small rRNAs, an abundant number of other bands are visualized on staining. These are chloroplast transcripts, the presence of which is taken as

an indicator of the integrity of the sample. Another bit of information that can be conveyed by examining an aliquot of RNA comes from the appearance of the well into which the sample is loaded. Fluorescence coming from within the well suggests that genomic DNA is present in the sample. Enormous fragments of chromosomal DNA, generated by shearing forces associated with the mechanics of RNA isolation, are unable to enter the gel during the course of the electrophoresis.

UV Spectrophotometry Similar to DNA, the concentration and purity of RNA preparations are easily determined by UV absorption spectrophotometry (for details see Section 1.3.1). As A_{260} of 1 corresponds to 40 μg/ml for RNA, the concentration of a RNA solution is easily calculated using following equation:

$$\text{Concentration of RNA } (\mu g/ml) = A_{260} \times \text{dilution factor} \times 40$$

For determination of the concentration of RNA, only OD values between 0.1 and 1 are significant because values between 0.02 and 0.1 may be associated with substantial errors. As UV spectrophotometry does not distinguish between pure samples of RNA and DNA-tainted samples, normalization based on A_{260} may be compromised because of the contribution to the total mass of the sample made by genomic DNA. Pure preparations of RNA have A_{260}/A_{280} value of 2.0. If there is significant contamination with protein or phenol, the A_{260}/A_{280} will be less than 2.0. The contamination of phenol may, however, be remedied by a final chloroform extraction. On the other hand, a low ratio may indicate incomplete solubilization of the RNA pellet.

Colorimetric Analysis by Orcinol Method Orcinol reaction confirms the presence of RNA in the sample. The principle behind the procedure is that when an RNA solution is heated in the presence of concentrated HCl, furfural is formed, which when reacted with orcinol in the presence of $FeCl_3$ as a catalyst forms a green-colored complex with a sharp absorption maximum at 665 nm. Note that only the purine nucleotides give any significant reaction. RNA concentration is determined by measuring the intensity of absorbance of the green-colored solution at 665 nm with a spectrophotometer and comparing with a standard curve of known RNA concentrations.

1.3.3 Construction of Recombinant DNA Molecule

After isolation and purification of DNA fragment containing the gene to be cloned and the vector DNA, the next requirement in a gene cloning experiment is construction of the recombinant DNA molecule. This involves cleavage of a vector as well as DNA to be cloned followed by their joining together in a controlled manner.

Cleavage of Vector DNA and Isolation of DNA Fragment from Genomic DNA

Each vector molecule is cut at a single position (e.g., in case of plasmids, λ insertional vectors, etc.) or at two different positions (e.g., in case of λ replacement vectors or artificial chromosomes, etc.) with restriction enzyme(s) and at exactly the same position(s) all the time. The large DNA molecule (genomic DNA) is cleaved or broken in precise and repro-

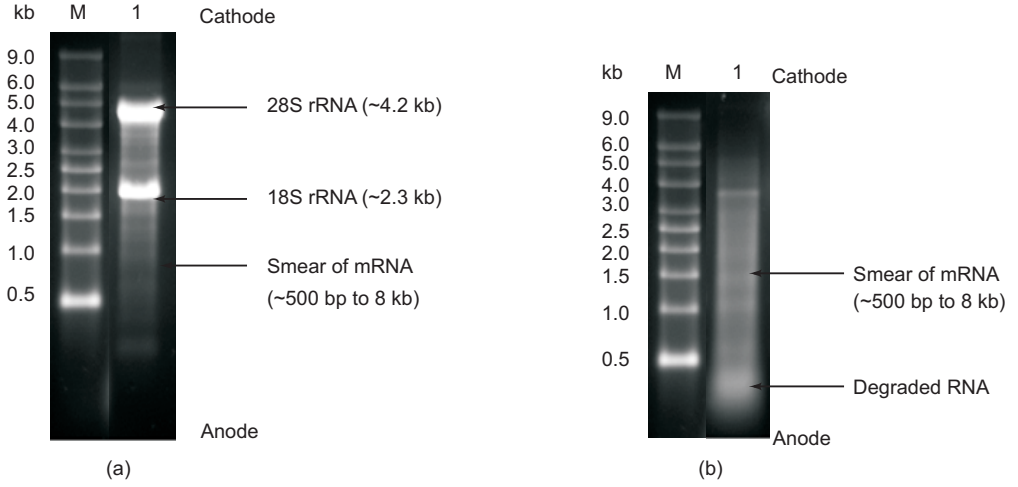


Figure 1.8 Electrophoresis of RNA on denaturing agarose gel; (a) Total RNA preparation [Lane M : RNA size marker; Lane 1 : Total cell RNA in eukaryotes; 18S and 28S rRNAs form clearly visible discrete bands at positions around 2.3 and 4.2 kb, respectively; Different mRNAs appear as smear between ~500 base and 8 kb, with the bulk mRNA lying between 1.5 kb and 2 kb. In prokaryotic total RNA preparation (not shown in figure) 16S and 23S rRNAs form two discrete bands at positions around 1.5 kb and 2.9 kb, respectively within the mRNA smear]. (b) Pure mRNA preparation [Lane M : RNA size marker; Lane 1 : Different mRNAs form smear between ~500 base and 8 kb; Degraded RNA appears as a lower molecular weight smear; Note that a pure mRNA preparation lacks two discrete bands of rRNAs.]

ducible manner. The purpose is either to isolate an intact single gene or to break down the large DNA into fragments small enough to be carried by the vector, which exhibits a preference for DNA fragments of a particular size range. Fragmentation is typically achieved by enzymatic or mechanical methods and the extent of treatment governs the average size of fragments produced. The most commonly used methods for fragmentation of genomic DNA are summarized below and also listed in Table 1.2.

Digestion with Restriction Enzymes Restriction digestion using site-specific restriction endonucleases is routinely used for obtaining DNA fragments in a precise and reproducible manner (for details see Chapter 3). The procedure is simple, reliable, and generates specific sized fragments.

Cleavage by Nonspecific Endonucleases Another enzymatic method for fragmentation of DNA is nonspecific endonuclease-catalyzed cleavage. This method gives nonuniform and random fragments.

Mechanical Shearing The long, thin duplex DNA molecules are sufficiently rigid to be very easily broken by shear forces in solution. By high-speed stirring in a blender, controlled shearing can be achieved. Typically, high molecular weight DNA is sheared to fragments with a mean size of ~8 kbp by stirring at 1,500 rev/min for 30 min. Similarly, passage through the orifice of a 28-gauge hypodermic needle leads to mechanical shearing of high molecular weight DNA (Table 1.2). The method is cheap, easy, and requires only small amounts of DNA. Both these procedures induce random breaks with respect to DNA sequence so that each time a DNA sample is treated, a different set of fragments are generated. Moreover, shearing leads to raggedness of ends. The termini consist of short, single stranded regions, which may have to be taken into account in subsequent joining procedures. These methods are thus useful for generating random, overlapping fragments of genomic DNA for use in chromosome walking experiments (for details see Chapter 19).

Automated Shearing Using High-Performance Liquid Chromatography (HPLC) During the last few years, a method for hydrodynamic shearing, initially based on the use of HPLC and called the 'point-sink' flow system, has become increasingly refined and finally automated. In the 'point-sink' flow system, an HPLC pump is used to apply pressure to the DNA sample, thereby forcing it through tubing of very small diameter. In the automated process known as HydroShear (commercially available from Gene Machines), a sample of DNA is repeatedly passed through a small hole until the

Table 1.2 Hydrodynamic Shearing Methods Used for DNA Fragmentation

Method	Advantages and disadvantages
Sonication	Easy and quick method of fragmentation; Requires sophisticated instrument; Requires relatively large amounts of DNA (10–100 µg); Fragments of DNA distributed over a broad range of sizes; Only a small fraction of the fragments are of a length suitable for cloning and sequencing; Requires ligation of DNA before sonication and end-repair afterward; Hydroxyl radicals generated during cavitation may damage DNA.
Nebulization	Easy and quick method of fragmentation; Requires sophisticated instrument; Requires only small amounts of DNA (0.5–5 µg), and large volumes of DNA solution; No preference for AT-rich region; Size of fragments easily controlled by altering the pressure of the gas blowing through the nebulizer; Fragments of DNA distributed over a narrow range of sizes 900–1,330 bp); Requires ligation of DNA before nebulization and end repair afterward.
Passage through the orifice of a 28-gauge hypodermic needle	Cheap method of fragmentation that does not require any sophisticated instrumentation; The method is easy and quick to perform; Requires only small amounts of DNA; Fragments are a little larger (1.5–2.0 kbp) than required for shotgun sequencing; Requires ligation of DNA before cleavage and end repair afterward.
Circulation through an HPLC pump	Requires expensive apparatus; Requires 1–100 µg of DNA; Fragments of DNA distributed over a narrow range of sizes that can be adjusted by changing the flow rate; Ligation of DNA required; End repair of fragments before cloning not necessary.

sample is fragmented to products of a certain size (Table 1.2). The final size distribution is determined by both the flow rate of the sample and the size of the opening. These parameters are controlled and monitored by the automated system. At any given setting, DNA fragments larger than a certain length are broken, whereas shorter fragments are unaffected by passage through the opening. The resulting sheared products therefore have a narrow size distribution. Typically 90% of the sheared DNA falls within a two-fold size range of the target length. Libraries constructed from these DNA fragments are likely to be of higher quality than those made using one of the old mechanical shearing methods. These certainly contain clones of more uniform size and possibly may be more comprehensive in their coverage of the genome.

Sonication In this method, isolated DNA is subjected to hydrodynamic shearing by exposure to brief pulses of sound waves (Table 1.2). A diagram showing functioning of sonicator is depicted in Figure 1.9. Most sonicators shear DNA to a size of 300–500 bp and sonication is continued until the entire population of DNA fragments has been reduced to this size. However, the yield of subclones is usually greater if sonication is stopped when the fragments of target DNA first reach a size of ~700 bp. Excessive sonication for long time periods makes it extremely difficult to clone the sonicated DNA, perhaps due to damage caused by free radicals generated by cavitation.

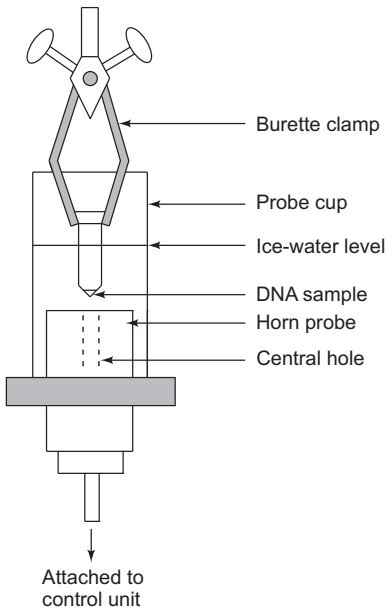


Figure 1.9 Cup Horn sonicator for random fragmentation of DNA [The cup horn attachment for the Heat Systems sonicator is depicted with a sample tube in place. The cup horn unit, which contains a large horn probe, is attached to the sonicator control unit and filled with ice water before the sample is sonicated. The sample tube is held in place from above by using a burette clamp and a ring stand. Alternatively, a tube holder can be fabricated from 1/4 inch plastic and used to hold up to eight tubes for simultaneous processing.]

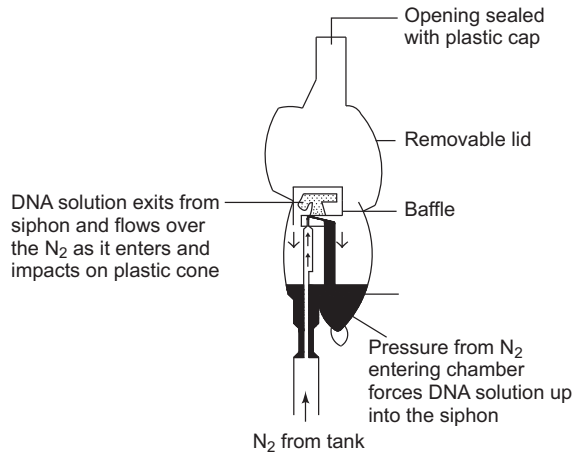


Figure 1.10 Functioning of nebulizer [A viscous DNA solution containing glycerol is placed in the nebulizer, which is attached to a nitrogen tank. Pressure from the nitrogen entering the chamber siphons the DNA solution from the bottom of the chamber to the top. The solution exits the siphon and impacts on a small plastic cone suspended near the top of the chamber, thus shearing the DNA.]

Nebulization Nebulization, a form of hydrodynamic shearing, is another way to get DNA fragments from the large genomic DNA (Table 1.2). It is performed by collecting the fine mist created by forcing DNA in solution through a small hole in the nebulizer unit (e.g., CA-209 from CIS-US Inc.) (Figure 1.10). The speed of passage of DNA solution through the hole, the viscosity of the solution, and the temperature regulate the size of the fragments.

Other Methods of Obtaining DNA Fragment to be Cloned If the gene of interest is small and its sequences are known, it can be synthesized chemically for cloning in expression vector (for details see Chapter 4). Similarly, if the gene sequences from the same or related sources are known, DNA fragment may be obtained by PCR amplification using gene-specific primers and the resulting PCR product may be cloned directly into an expression vector (for details see Chapter 7). If the interest lies in gaining information about the expressed sequences or splice sites or developmentally regulated and tissue-specific sequences, mRNA molecules purified from the total cell RNA may be reverse transcribed to cDNA molecules, which when cloned into vector molecules form a cDNA library (for details see Chapters 5 and 15).

Although each approach is reasonably successful for generating a range of fragments from a large contiguous segment of DNA, each has its particular limitation. As physical methods used for shearing DNA are independent of sequence

composition, these methods typically result in more uniform and random disruption of the target DNA than enzymatic methods. In particular, methods involving hydrodynamic shearing due to physical stress induced by sonication or nebulization produce collections of appropriately random fragments. The variety in lengths of these fragments is quite large and their application usually requires a subsequent size selection step to narrow the range of fragments to be acceptable for cloning or sequencing. However, libraries constructed from sonicated or hydrodynamically sheared DNA, although imperfect, are certainly workable. The only way to generate precise and defined fragments is to cleave with restriction enzymes.

Joining of DNA Fragments to Vector DNA Molecule

Once the desired DNA fragment has been isolated, further manipulations require them to be inserted into a cloning vector (for details see Chapter 13). This is because fragments of DNA are not replicons and in the absence of replication, these are diluted out of their host cells. It should be noted that even if a DNA molecule contains an origin of replication, this might not function in a foreign host cell. If fragments of DNA are not replicated, the obvious solution is to attach them to a suitable replicon. Such replicons are known as vectors (for details see Section 1.2). Thus isolated DNA fragments are joined with linearized or cleaved vector DNA molecules in a DNA ligase-catalyzed reaction, resulting in the formation of recombinant DNA molecules. The problems that may occur due to vector reconstruction or DNA fragment dimer formation are reduced at this stage by dephosphorylation (for details see Chapter 2) and the DNA fragments of appropriate size are obtained by size fractionation techniques (for details see Chapter 6). Short synthetic oligonucleotides, *viz*. linkers and adaptors, may be used to increase the versatility of ligation reaction (for details see Chapter 13). Other strategies for joining of DNA fragments with vector DNA molecule include homopolymer tailing and action of *Vaccinia* DNA topoisomerase (for details see Chapter 13).

1.3.4 Introduction of Recombinant DNA Molecule into Host Cell

The next step in a gene cloning experiment is to introduce the recombinant DNA molecule into living cells, usually bacterium, or any other living cell. Once introduced into host cell, with the multiplication of the host, a large number of recombinant DNA molecules may be produced from a limited amount of starting material. When the host cell divides, copies of the recombinant DNA molecules are passed to the progeny, and further vector replication takes place. After a number of cell divisions, a large number of genetically identical host cells (clones) are produced, which appear as a colony or plaque depending upon the vector molecules. Each cell in the clone contains one or more copies of a recombinant DNA molecule. This stage marks the 'cloning' of the gene. Different techniques used for introduction of recombinant DNA molecules into host cell include transformation, transfection, or *in vitro* packaging followed by natural infection by the assembled infectious phage (for details see Chapters 9 and 14).

1.3.5 Selection of Recombinants and Screening of Desired Clone

To isolate the clone of interest or to identify the clone containing the gene of interest, the next step in a gene cloning experiment is the selection and screening of recombinants or transformants. Screening of desired clone is almost invariably done at the plating out stage. Thus, by plating on selection medium (based on the product of selectable marker gene), recombinants are either distinguished from nonrecombinants by color reaction or are the only hosts capable of growth on that medium (for details see Chapter 16). Once it is established which colonies/plaques are recombinants, the next step is to search for the desired clone. Although there are many different procedures by which the desired clone is obtained, all are variations of two basic strategies. (i) Direct selection for the desired gene; and (ii) Identification of the desired clone from a gene library.

Direct Selection for the Desired Gene

This strategy is used when the cloning experiment is designed in such a way that only cells containing the desired recombinant DNA molecule divide and the clone of interest is automatically selected, i.e., the clones obtained are the clones containing the required gene. In general terms, this is the preferred method, as it is quick and usually unambiguous.

Identification of the Desired Clone from a Gene Library

Direct selection is not applicable to all genes and hence techniques for clone identification are very important. Moreover, during the ligation reaction, several different recombinant DNA molecules are produced, all containing different pieces of DNA and there is no selection for an individual fragment. Consequently, a variety of recombinant clones are obtained after transformation and plating out (Figure 1.11). Somehow the correct one must be identified. For this purpose, first a clone library representing all or most of the genes present in the cell is constructed (for details see Chapter 15), which is followed by analysis of the individual clones to identify the desired one from a mixture of lots of different clones. This strategy requires extensive screening because even the simplest organisms, such as *E. coli*, contain several thousand genes and a restriction digest of total cell DNA produces not only the fragment carrying the desired gene, but also many other fragments

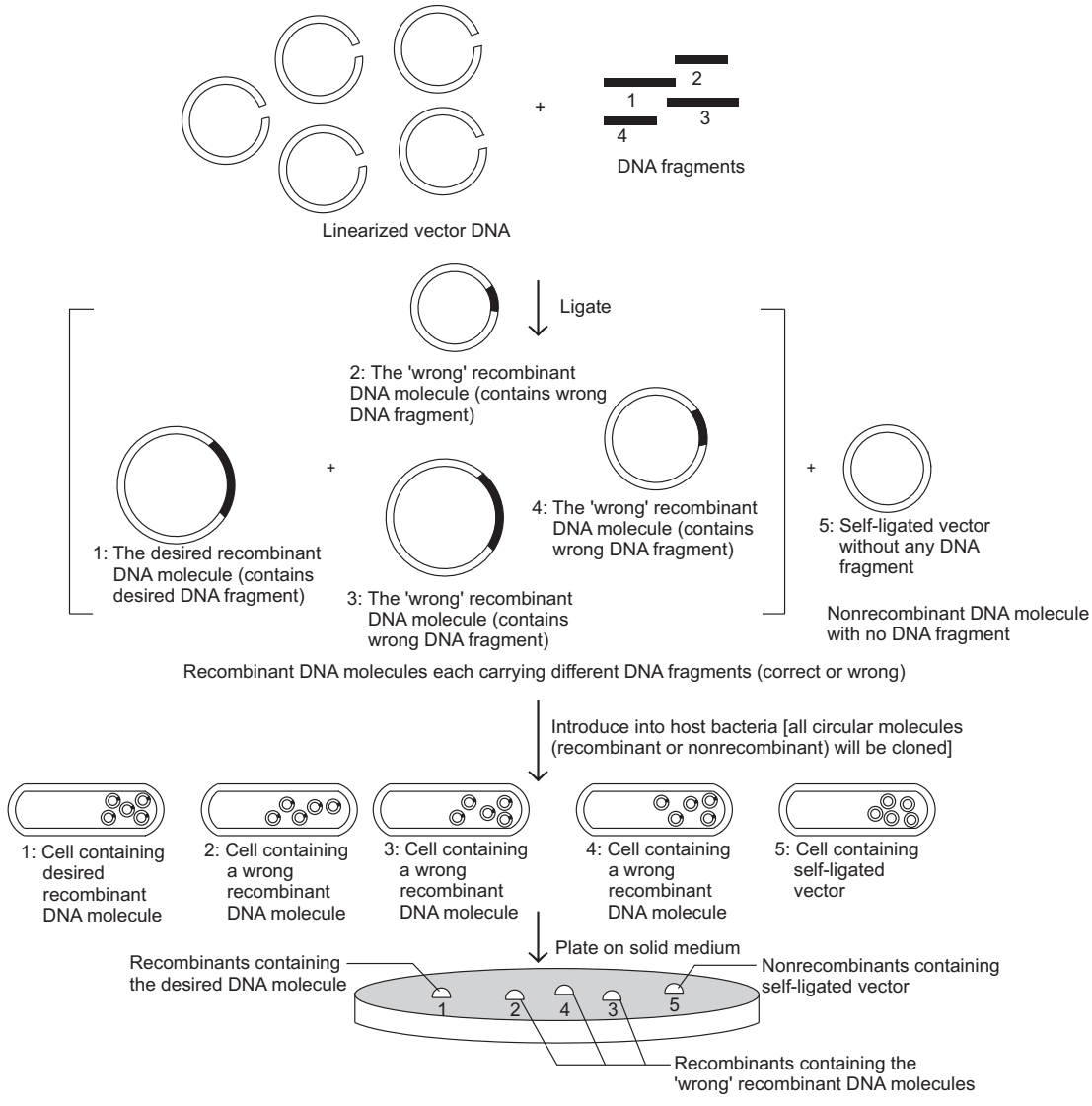


Figure 1.11 Cloning allows purification of individual fragments of DNA generated in the first step of cloning experiment [The DNA fragment to be cloned is one member of a mixture of many different fragments, each carrying a different gene or part of a gene. This mixture could indeed be the entire genetic complement of an organism, a human, for instance. Each of these fragments becomes inserted into a different vector molecule to produce a family of recombinant DNA molecules, one of which carries the gene of interest. Usually only one recombinant DNA molecule is transported into any single host cell so that the final set of clones contain multiple copies of just one DNA molecule. The gene is now separated away from all the other genes in the original mixture, and its specific features can be studied in detail. Cloning is analogous to purification. From a mixture of different molecules, clones containing copies of just one molecule can be obtained.]

carrying other genes. Moreover, there is no guarantee that the entire gene is present on a single clone. Thus, once a suitable library is prepared, a number of procedures based either on direct identification of the correct recombinant DNA molecule or detection of translation product of the cloned gene are employed to identify the desired clone (for details see Chapter 16). The DNA-based techniques are usually easier and involve hybridization probing, i.e., hybridization analysis by using either a radioactively or fluorescently labeled DNA probe complementary or partially complementary to a region of the gene sequence. On the other hand, other methods rely on the expression of coding sequences of the clones in the library and

identification of the protein product by its activity or by antigen antibody reaction. Once identified, a cloned gene is analyzed by restriction mapping and ultimately by DNA sequencing. The sequence may also be analyzed by comparison with other known sequences from databases and the complete sequence of the protein product determined. The sequence is then available for manipulation in any of the diverse applications of DNA cloning (for details see Chapter 20).

1.3.6 Storage of Desired Bacterial Clones

Once the clone of interest is identified, it should be stored for long periods. As bacteria do not survive more than a week in a liquid culture, their storage is not done in broth. On an agar plate, the bacteria can survive up to 4 weeks and this works best when the plate is sealed with parafilm to prevent drying. For long-term storage, however, this is not of any use. Storage at room temperature in agar vials is better. This approach deals with agar stabs (1.5–5.0 ml), which are prepared by filling two-third of the glass or plastic vials with sterile stab agar (LB medium with 0.6 w/v agar; 10 mg/l of cysteine may be added to increase the survival time of bacteria). This agar stab is inoculated with the desired bacterial clone by repeatedly pushing the sterile inoculation needle or toothpick and the stab is incubated for about 8 hours at 37°C and stored tightly closed in a dark and dry place at room temperature. The bacterial clones survive in this manner for years although 3–6 months is more realistic. To revive the bacteria, a sterile inoculating loop is stuck into the stab agar and smeared onto an LB agar plate and the bacteria are allowed to grow overnight at 37°C. To preserve bacteria for a very long time, glycerin or dimethyl sulfoxide (DMSO) stocks are prepared by adding 1 ml of fresh culture with 1 ml of glycerin solution (65% v/v glycerin; 0.1 M $MgSO_4$; 25 mM Tris-Cl, pH 8.0) or 1 ml of DMSO solution (7% v/v DMSO). These stocks are stored at –20°C or –70°C. Note that the latter temperature increases the survival time. The advantage of DMSO is that it is easy to pipette as compared to glycerin. The bacteria are revived from these stocks by scratching the surfaces with a sterile toothpick or inoculation loop, inoculating an LB plate and growing overnight at 37°C. Another method for long-term storage of bacteria is freeze-drying or its low-cost version, vacuum drying. In this procedure, the bacteria are frozen, followed by removal of water by means of a vacuum, and collection of all the remnants in glass tubules in an anaerobic state. Depending on the mode of storage and the specific microorganisms, the survival rate can be up to 30 years. For revival of bacteria, the ampoule is broken carefully and the powdered bacteria are used for inoculation into a fresh LB medium. Besides storing bacteria, the most common and reliable method is to store bacterial DNA stocks at –20°C. As long as the DNA is free from DNases, it is unlikely to degrade. The concentration should be more than 0.1 µg/ml. Revival is somewhat laborious than with bacterial stocks because the DNA has to be transformed into bacteria.

1.4 SUBCLONING

The simple transfer of a cloned fragment of DNA from one vector to another is termed as subcloning. It serves to illustrate many of the routine techniques involved in gene cloning, for example, for investigation of a short region of a large cloned fragment in more detail or to transfer a gene to a vector designed to express it in a particular species or for transferring in M13 vectors after manipulation to get single stranded DNA (ss DNA) for further application in DNA sequencing and site-directed mutagenesis experiments. Note that phagemids possess the advantages of both plasmids and filamentous coliphages (e.g., M13) and hence cloning in phagemids does not require any subcloning to get ss DNA.

1.5 ADVANTAGES OF GENE OR cDNA CLONING

Amongst the wide-ranging benefits of gene cloning or cDNA cloning, often given together under the term genetic engineering, are the following:

To Get Homogeneous Preparations of any Desired DNA Cloning was a significant breakthrough in molecular biology because it became possible to obtain homogeneous preparations of any desired DNA molecule in amounts suitable for laboratory-scale experiments.

To Get High Yields of Recombinant DNA The vector not only provides the replicon function, but it also permits the easy bulk preparation of the foreign DNA sequence free from host cell DNA (Figure 1.12). Several micrograms of recombinant DNA molecules are usually prepared from a single bacterial colony, representing a 1,000-fold increase over the starting amount. If the colony is used as a source of inoculum for a liquid culture, the resulting cells may provide milligrams of DNA, a million-fold increase in yield.

To Obtain Pure Sample of a Gene Another important application of cloning is purification of an individual gene, separated from all the other genes in the cell. To understand exactly how cloning can provide a pure sample of a gene, consider the basic experiment from Figure 1.1, drawn in a slightly different way (Figure 1.12). The manipulation that results in a recombinant DNA molecule can rarely be controlled to the extent that no other DNA molecules are present at the end of the procedure. In addition to the desired recombinant DNA molecule, the ligation mixture may contain unligated vector molecules, unligated DNA fragments, DNA fragment dimers, recircularized or self-ligated vector molecules, and recombinant DNA molecules carrying the wrong DNA insert. Though

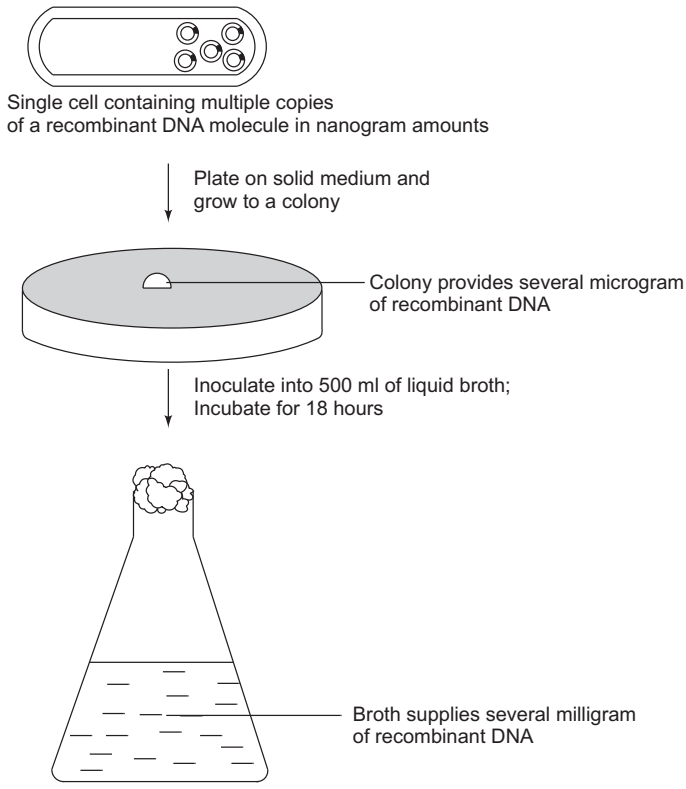


Figure 1.12 Cloning supplies large amounts of recombinant DNA

bacterial cells may take up unligated molecules or DNA fragment dimers, these rarely cause a problem because these either do not replicate or are subjected to degradation by the host enzymes (i.e., replicate only under exceptional circumstances). On the other hand, self-ligated vector molecules and incorrect recombinant vectors are replicated just as efficiently as the desired molecule. However, purification of the desired molecule can still be achieved through cloning because it is extremely unusual for any one cell to take up more than one DNA molecule. Each cell gives rise to a single colony, so each of the resulting clones consists of cells that contain the same molecule. Thus, different colonies contain different molecules, some contain the desired recombinant DNA molecule, some have wrong recombinant DNA molecule, and some contain self-ligated vector. If, somehow, the colonies containing the correct recombinant DNA molecule are identified, a pure sample of the gene of interest may be obtained.

Isolation and Manipulation of Fragments of an Organism's Genome DNA cloning facilitates the isolation and manipulation of fragments of an organism's genome by replicating them independently as part of an autonomous vector.

Isolation of Long Genes and Unknown Genes Gene cloning is the only way of isolating long genes with the application of 'chromosome walk' (for details see Chapter 19). Gene cloning is also the only procedure to isolate the genes that have never been studied before; such unknown genes cannot be isolated through PCR.

Elucidation of Gene Function, Promoter Analysis, and Identification of Mutations The task of functionally annotating genomes always lags way behind the structural annotation phase. Classically, detailed molecular analysis of proteins or other constituents of most organisms was rendered difficult or impossible by their scarcity and the consequent difficulty of their purification in large quantities. One approach is to isolate the expressed gene(s). Standard chemical or biochemical methods cannot be used to isolate a specific region of the genome for study, particularly as the required sequence of DNA is chemically identical to all the others. This is because every organism's genome is large and complex, and any sequence of interest usually occurs only once or twice per cell. Gene cloning strategies remain of value for the elucidation of gene function, for example, elucidation of expression profiles or biochemical functions of the protein encoded by the cloned gene. The technique is used for investigation of protein/enzyme/RNA function by large-scale production of normal and altered forms. Using this technique, new insights are emerging, for example, into the regulation of gene expression in cancer and development and the evolutionary

history of proteins as well as organisms. Gene cloning is also used for isolation and analysis of gene promoters and other control sequences as well as for the identification of mutations, e.g., gene defects leading to disease.

Information About Cell Type and Developmental Stage-Specific Genes, Locations of Splice Sites, and Alternatively Spliced Genes The genome sequences reveal only part of the information available for a given gene. In contrast, cDNA sequences, which are reverse transcribed from mRNA, are cloned to reveal expression profiles in different cell types, developmental stages, and in response to natural or experimentally simulated external stimuli. Additionally, for higher organisms, cDNA sequences provide useful information about splice sites, splice isoforms, and their abundance in different tissues and developmental stages.

DNA or Genome Sequencing As many genomes have yet to be mapped or sequenced, other exploding application of gene cloning is DNA sequencing and consequently derivation of protein sequence (for details see Chapter 19).

Site-directed Mutagenesis Gene cloning is also helpful in site-directed mutagenesis experiments used to alter the properties of proteins, for example, to enhance thermal tolerance and pH stability of enzymes, increase resistance of proteins to protease, or to increase specificity and catalytic efficiency of enzymes, etc. (for details see Chapter 17).

Applications Other Than Research The applications of gene cloning include large-scale commercial production of proteins and other molecules of biological importance (e.g., human insulin, growth hormone, restriction enzymes, antibiotics, biopolymers, recombinant vaccines), edible vaccines, engineering novel pathways, bioremediation, reproductive cloning, gene therapy, and engineering organisms (microorganisms, animals, and plants) with useful quality attributes (for details see Chapter 20).

Review Questions

1. Discuss the Watson–Crick double helical structure of DNA. How does DNA differ from RNA?
2. Differentiate between synthetic and complex media. Give examples of each.
3. Discuss the properties of an ideal cloning vector. Why should an ideal cloning vector be small in size? Also give some examples of cloning vectors.
4. What is biological containment and what is its significance? How is it achieved?
5. What do you understand by the term 'gene cloning'? Give its experimental details. How is the process used for gene isolation?
6. Describe in detail the procedure for the preparation of total cell DNA from a bacterial cell. How does the process differ from that of isolation of total plant DNA?
7. Describe the principle and procedure of DNA purification using guanidinium thiocyanate and silica method.
8. Explain various strategies adopted to prevent contamination of DNA preparation by proteins, polysaccharides, and RNA.
9. Describe various strategies adopted for prevention of RNase contamination during RNA isolation procedure. What is the significance of wearing gloves in the process?
10. What are the differences between guanidinium containing lysis buffers and NP-40 containing lysis buffers used for RNA isolation?
11. Discuss the roles of following in nucleic acid isolation and purification procedures: lysozyme, β-ME, guanidinium thiocyanate, SDS, Na_2EDTA, monophasic reagents, proteinase K, phenol, chloroform, isoamyl alcohol, CTAB, DNase I, RNase, DEPC, sodium acetate, LiCl, 100% EtOH, silica columns, isopycnic ultracentrifugation, and washing with 70% EtOH.
12. How is purity of DNA and RNA preparations checked by UV absorbance measurement? If absorbance analysis of a DNA preparation reveals A_{260}/A_{280} ratio of 1.5, discuss about the purity of DNA preparation.
13. How is purity and integrity of total RNA preparation assessed by agarose gel electrophoresis?
14. Explain the following:
    (a) RNA is more reactive than DNA.
    (b) *E. coli* was the first bacteria to be used in cloning experiments.
    (c) Primary cloning is preferably done in *E. coli*.
    (d) PCR can be used as an alternative to cloning.
    (e) Gene cloning is the only way for isolating long genes or those that have never been studied before.
    (f) Nucleic acids are easily precipitated with EtOH in the presence of monovalent ions such as Na^+, NH_4^+, K^+, or Li^+.
    (g) One should wear gloves while isolating RNA.
    (h) 'Boiling inactivates DNases but not RNases' or 'RNases are sturdy molecules'.
    (i) After treatment of glassware, plasticware, and solutions with DEPC for appropriate time period, it is removed readily by autoclaving.
15. Why is fragmentation of genomic DNA required for cloning? Discuss different methods for obtaining DNA fragments for cloning experiments. Also discuss the advantages and disadvantages of each procedure.
16. What do you understand by the term 'subcloning'? Explain its significance.
17. Discuss the advantages of cDNA cloning. Also enumerate various applications of gene cloning.
18. The absorbance measurement at 260 nm with 2 µl of a 10 times diluted DNA sample reveals an OD of 0.8. Calculate the concentration of DNA (in µg) in a 100 µl DNA preparation.