Progress in Molecular and Subcellular Biology

6

Contents

Springer-Verlag
Berlin · Heidelberg · New York

6 Progress in Molecular and Subcellular Biology

With Contributions by
F. C. Cabello · S. N. Cohen · B. M. Cox
J. F. P. Dixon · L. Ebringer · A. Goldstein
F. E. Hahn · S. L. Hajduk · H. Kleinkauf
H. Koischwitz · R. L. O'Brien · J. W. Parker
K. N. Timmis

Edited by: F. E. Hahn, H. Kersten
W. Kersten, W. Szybalski

Advisors: T. T. Puck, G. F. Springer
K. Wallenfels

Managing Editor: F. E. Hahn

Springer-Verlag Berlin Heidelberg New York 1978

Professor Fred E. Hahn, Ph. D.
Division of Biochemistry
Walter Reed Army Institute of Research
Washington, D. C. 20012, USA

With 65 Figures

ISBN 3-540-08588-2 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-08588-2 Springer-Verlag New York Heidelberg Berlin

Library of Congress Catalog Card Number 75-79748.

Printed in Germany.

Offsetprinting: Beltz Offsetdruck, Hemsbach
Bookbinding: Konrad Triltsch, Graphischer Betrieb, Würzburg
2131/3130-543210

List of Contributors

Cabello, Felipe, C., Department of Microbiology, New York Medical College, Valhalla, New York 10595, USA

Cohen, Stanley N., Department of Medicine/S-155, Stanford University School of Medicine, Stanford, California 94305, USA

Cox, Brian M., Addiction Research Foundation, 701 Welch Road, Palo Alto, California 94304, USA

Dixon, James F.P., Department of Pathology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, USA

Ebringer, Libor, Department of Microbiology, Comenius University, Páričkova 3, 80100 Bratislava, Czechoslovakia

Goldstein, Avram, Addiction Research Foundation, 701 Welch Road, Palo Alto, California 94304, USA

Hahn, Fred E., Division of Biochemistry, Walter Reed Army Institute of Research, Washington, D.C. 20012, USA

Hajduk, Stephen L., Department of Zoology, University of Georgia, Athens, Georgia 30602, USA

Kleinkauf, Horst, Max-Volmer-Institut für Physikalische Chemie und Molekularbiologie, Technische Universität Berlin, Franklinstraße 29, 1000 Berlin 10

Koischwitz, Hans, Max-Volmer-Institut für Physikalische Chemie und Molekularbiologie, Abtl. Biochemie, Technische Universität Berlin, Franklinstraße 29, 1000 Berlin 10

O'Brien, Richard L., Department of Pathology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, USA

Parker, John W., Department of Pathology, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, California 90033, USA

Timmis, Kenneth N., Max-Planck-Institut für Molekulare Genetik, Ihnestraße 63-73, 1000 Berlin 33

Genes Within Genes

Fred E. Hahn

In the Editorial to the first volume of this Progress series (Hahn, 1969), a proposition of Stent (1968) was discussed that molecular biology was essentially of the past and that all that remained for the field in the academic phase that it had attained was "the need to iron out the details." Subsequently, an apocryphal *bon mot* was widely circulated that a molecular biologist was a former *Escherichia coli* geneticist in search of a new field of study.

Kornberg (1976) observed: "Ten years ago, fashionable biochemists were examining .... the molecular details of gene expression in replication, transcription and translation. There was emphasis on using bacteria and their phages, the simplest and most rewarding systems for basic biomedical and genetic studies. Then interest shifted to bigger and more complex things: ..... animal cells were in, and bacteria were out."

Kornberg then described one of his "most useful" research decisions, *viz.*, to abandon work on *E.coli* DNA replication and, instead, turn to the biochemistry of much simpler bacteriophage DNA. Toward the end of his writing, he singled out a major concern: "research is diverted by fashion and by funding pressures to a premature attack on overly complex problems."

Recent advances in the biochemical genetics of phage ϕX174, however, indicate that molecular biology may have entered into another fundamental pioneering phase through the study of a small and simple object and that some generally held basic assumptions about the nature of the genetic code may need to be revised.

Gamow (1954) suggested the first model of a genetic code with overlapping triplet codons and also presaged the possibility of codon ambiguity, meaning that several different triplets might code for one and the same amino acid. These ideas were focused upon by Crick in 1955 in a paper entitled On Degenerate Templates and the Adaptor Hypothesis. Unfortunately, Crick's manuscript was not published but only circulated among a small number of friends; only the adaptor hypothesis was eventually published as a discussion remark (Crick, 1957).

The designation of the code as being "degenerate" is perhaps an unfortunate choice of expression not only from the cryptographic viewpoint but because of its evolutionary connotation. Nevertheless, of the 4^3=64 possible triplets, three are stop signals (UAA, UAG, and UGA), and the codons AUG (for methionine), or less frequently GUG (for valine), are parts of a more complicated initiation signal in the translation of the code into protein. This leaves 61 different codons to specify 20 amino acids. Different codons which specify the same amino acid are called synonyms (leucine and arginine, for example, each are specified by 6 different synonyms). The biological significance or purpose of the extensive use of synonyms in the code has gone unrecognized although it explains the existence of organisms whose DNA composition ranges from 30 to 70 per cent [G+C].

It now turns out that the abundance of synonyms renders it possible to write structural genes which are entirely contained in the nucleotide sequence of larger structural genes. Two such genes are read in different reading frames or "phases" during phenotypic expression (Barrell et al., 1976; Sanger et al., 1977).

The genome of the small *E.coli* virus ϕX174 consists of one single strand of DNA of a length of 5375 nucleotides; the complete nucleotide sequence has been determined by Sanger and his associates (1977). A genome of this size has a maximal coding capacity for proteins of an aggregate molecular weight of approximately 200,000 daltons. However the nine gene products of ϕX174 DNA have a combined molecular weight of 250,000. How is this excess explained?

The nine ϕX174 genes carry the designations A, B, C, D, E, J, F, G and H. The 260 nucleotides of the B-gene are totally contained within the 1546 nucleotides of the A-gene (Sanger et al., 1977) and the 273 nucleotides of the E-gene are totally contained within the 456 nucleotides of the D-gene (Barrell et al., 1976). This is accomplished by the placement of synonymous triplets whose sequence can be read in two reading frames or "phases", each messenger making sense and giving rise to translation into a biologically functional protein.

Up to that time, it was generally assumed that individual genes in genomes are contiguous and separated by termination and initiation "punctuation" and that a shift in the reading frame (as in frameshift mutations) would, from its locus on down, cause a nonsensical transcription and, hence, translation into non-functional protein. The translation of a message was thought to be non-overlapping, the correct reading frame was set at a defined starting point and the message then sequentially read off, groups of three letters at a time. For the gene E which has been shown (Barrell et al., 1976) to lie completely within the nucleotide sequence of gene D it is established that it is read in a different "phase" or frame which is displaced one nucleotide to the right, i.e., in the direction of reading. The possibility of complete structural gene overlap is under consideration for additional bacterial viruses (Lewin, 1976).

These recent results of fundamental import bear out the view of Kornberg (1976) that the study of "tiny bacterial viruses proved to be uniquely useful beyond my expectations." For it must be noted that these results were obtained with a phage for which there existed a complete genetic analysis; a complete sequence analysis of its genome, and complete amino acid sequence analyses for several of its gene products.

The central question remains, of course, if this type of compact genetic organization is merely a peculiarity of certain bacterial viruses or is of wider biological distribution and significance. It will be difficult and, currently, is impossible to obtain conclusive experimental evidence for or against such a genetic organization in bacteria whose genomes and number of gene products are orders of magnitude greater than those of the small viruses.

Two arguments can be advanced which favor the idea of a more general occurrence of genetic overlap. In evolutionary terms, prokaryotes (and probably their parasites) are much older than eukaryotes. It would be paradoxical if the obvious advantages of high genetic information density should have been lost during evolution in favor of a strictly sequential arrangement of individual genes.

More important, the high content of synonyms in the codon catalogue may serve the *purpose* of organizing overlapping structural genes. No other

persuasive purpose of synonymy has been discovered or proposed. Since the code is univeral, it is difficult to evade the speculation that the potential and advantage of the overlapping type of cryptography may be equally universal.

References

Barrell, B.G., Air, G.M., Hutchinson, C.A. III: Overlapping genes in bacteriophage ϕX174. Nature 264, 34 (1976)
Crick, F.H.C.: Discussion. Biochem. Soc. Symp. 14, 25. Cambridge Univ. Press, 1957
Gamow, G.: Possible relation between deoxyribonucleic acid and protein structures. Nature 173, 318 (1954)
Hahn, F.E.: On molecular biology. In: Progress in Molecular and Subcellular Biology, Vol. 1, p. 1. Berlin-Heidelberg-New York: Springer 1969
Kornberg, A.: Research, the lifeline of medicine. N. Engl. J. Med. 294, 1212 (1976)
Lewin, B.: DNA sequences coding for more than one protein. Nature 264, 11 (1976)
Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson, C.A. III, Slocombe, P.M., Smith, M.: Nucleotide sequence of bacteriophage ϕX174 DNA. Nature 265, 687 (1977)
Stent, G.S.: That was the molecular biology that was. Science 160, 390 (1968)

Contents

DNA Cloning and the Analysis of Plasmid Structure and Function

K. N. Timmis, S. N. Cohen, and F. C. Cabello

A. Introduction

Genetic recombination is the fundamental biological process of exchange of genetic information between different chromosomes. Its constant activity during evolution has enabled the large number of spontaneously occuring genetic changes that improve the fitness of individuals for their particular environments to accumulate in the most successful members of each biological species. Recombination has provided much of the foundation for the science of genetics; it has been exploited to great social benefit by plant and animal breeders and to great scientific benefit by classical and molecular geneticists. Although there is evidence that recombination between DNA segments that have little or no ancestral relationship can occur under some circumstances (Cohen, 1976; Starlinger and Saedler, 1976), "ordinary" or "generalized" recombinational events commonly involve the reciprocal exchange of genetic material and require DNA sequence homology in the region of exchange. Thus, recombination in the laboratory between unrelated species of organisms having little DNA sequence homology is ordinarily not feasible. However, it has long been apparent that great benefits could be derived from intergeneric, as well as intrageneric, genetic manipulations.

Genetic recombination consists essentially of the breakage and joining of DNA molecules. Recent developments now permit DNA obtained from a wide variety of prokaryotic and eukaryotic sources to be cut in vitro at precisely defined locations and the DNA fragments thus generated to be coupled enzymatically to a self-replicating genetic element, known as a cloning vector or vehicle (either a plasmid or bacterial virus genome). Hybrid molecules generated in this fashion are introduced into *Escherichia coli*, where they are perpetuated and can be studied (Cohen, 1975). The host *E.coli* cells containing a hybrid molecule thus can serve as "cellular factories" for producing large amounts of the cloned DNA (and in some instances, the gene products specified by the cloned DNA) and, in addition, can serve as a well defined genetic background against which to study the expression of the cloned DNA fragment. The technologies that have been developed to permit the in vitro cloning of individual fragments of foreign DNA are collectively termed "DNA cloning", "molecular cloning", "gene cloning", "gene manipulation", and "genetic engineering".

The potential applications in the biochemical sciences to obtain basic information about fundamental biological processes, and in the applied sciences to obtain a variety of biological products that are of medical, agricultural, and commercial importance, and that are otherwise expensive or unobtainable in large quantity, indicate that DNA cloning methods represent a tool of extraordinary usefulness (Ashby Report, 1975; Curtiss III, 1976; Cohen, 1977).

The spectacular advances in the investigation of the structure and function of prokaryotic genes and operons, for example in the bacterio-

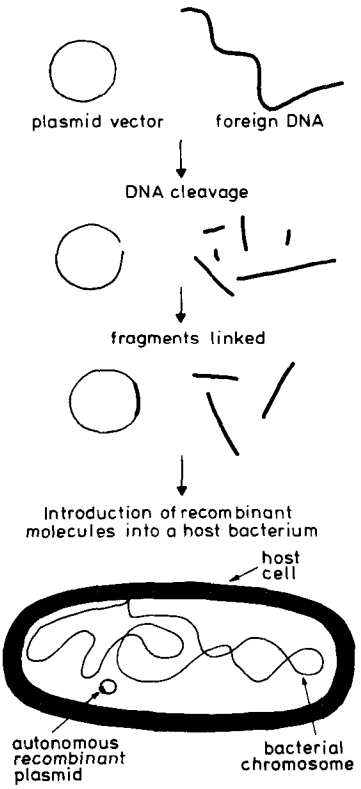


Fig. 1. The DNA cloning procedure

phage *lambda* and the *lac* operon, occurred largely because the DNA segments containing the genes under investigation were either part of a small autonomous replicon (*lambda*) or were easily sequestered on to such replicons (e.g., lambdoid phages and the sex factor F). In this state of reduced genetic complexity, genes could be investigated in a variety of ways in systems greatly depressed in non-specific genetic and biochemical background activity (see various articles in "The Lactose Operon", eds. Beckwith and Zipser, 1970, and "The Bacteriophage Lambda", ed. Hershey, 1971). The enormous genetic complexity of eukaryotic cells (some of which are five orders of magnitude more complex than the bacteriophage λ) is thus the main obstacle to a molecular analysis of gene structure, function, and regulation in higher organisms. Hence the development of methods to sequester defined eukaryotic DNA sequences on to small prokaryotic or eukaryotic autonomous extrachromosomal elements represents a major breakthrough in eukaryote molecular biology. As a consequence of this and other newly developed technologies, an understanding of the basic regulatory mechanisms controlling eukaryotic gene expression is expected to be greatly facilitated. Such an understanding is an absolute requirement for the elucidation of the biochemical bases of a number of pathological conditions, including tumorogenesis.

Cloned DNA sequences from prokaryotic and eukaryotic sources may be easily isolated in quantity in a degree of purity that was previously unattainable and that is suitable for structural studies, such as physical mapping and DNA sequencing. Furthermore, the gene products coded

by cloned fragments, at least those from many prokaryotic sources, may be manufactured in large amounts by the host cell. This latter aspect of molecular cloning has obvious significance for the production of biological products important in medicine, agriculture, and industry. A few selected examples of potential benefits of gene cloning experiments are given below:

1. The use of non-pathogenic host bacteria containing genes for antigens characteristic of specific agents of disease should greatly improve the effectiveness and safety of vaccine production.

2. Pollution of the environment by oil wastes may be combated with bacteria specifically armed with DNA sequences enabling them to utilize these products as sole carbon sources for growth. Indeed single cell protein might be manufactured from such wastes.

3. Provided that problems of transcription and translation of some types of foreign DNA in *E.coli* are overcome, the exciting possibility of manufacturing medically important products like insulin and other hormones, interferon, enzymes, and antibiotics, in convenient bacterial systems may soon be realized.

4. Food plants and animals, or the microbes that associate with them, may be genetically manipulated in order to increase the yield or the quality of the food material produced. Currently receiving great attention is the possibility that nitrogen fixation genes might be introduced into bacteria that can colonize the roots of nonlegumenous plants.

It would not be appropriate for us to discuss the potential benefits of gene cloning methods without mentioning the concerns that have been raised about possible hazards of some experiments that use these methods (see Cohen, 1977). Organisms appear to have evolved biological mechanisms that originally limited genetic exchange with unrelated species. Although the significance of these mechanisms is not yet elucidated, their functions are assumed to be of some evolutionary importance. DNA cloning in vitro permits the investigator to construct molecular chimeras by the fusion of DNA segments derived from organisms that are not known to be otherwise capable of exchanging genetic information. Since the properties of certain gene combinations made in this way may not be entirely predictable, the investigator is now required by national research organizations to carefully evaluate potential biohazards of projected molecular cloning experiments and subsequently to perform such experiments under appropriate laboratory conditions of containment similar to those used for work with organisms known to be hazardous (NIH Guidelines, 1976; Williams Report, 1976).

In this review we will focus our discussion on how the molecular cloning technology can be fruitfully exploited to investigate plasmid DNA structure and function. Because the problems of transcription/translation of cloned plasmid DNA segments are minor compared with those associated with cloned eukaryotic DNA segments, significant advances promoted by the use of the cloning technology have already taken place in the plasmid field in the four years since its first description in the literature (Cohen et al., 1973). However, most of the cloning strategies that have been employed for investigation of plasmid DNA structure and function are appropriate for the study of larger, more complex replicons such as chromosomes. We therefore feel that this is an appropriate point in time at which to review recent advances in the plasmid field resulting from molecular cloning experiments and to underscore those experimental approaches that may find a wider use in the study of more complex genetic systems. Although it will be neces-

sary for us to discuss relevant aspects of restriction enzymes, plasmids, and the cloning technology, these topics will not be covered exhaustively here. For other pertinent reviews of these subjects the order is: Arber, 1974 (restriction and modification); Nathans and Smith, 1974; Roberts, 1976; Roberts, 1977 (restriction enzymes); Helinski, 1973; Falkow, 1975 (plasmids); Cohen, 1975; Murray, 1976; and Collins, in press 1977 (gene cloning).

B. Restriction Endonucleases

Restriction endonucleases are site-specific endodeoxyribonucleases that cleave double-stranded and in some instances single-stranded DNA. Although the first site-specific endodeoxyribonucleases to be characterized were known to function as enzymatic blocks or restriction barriers for the prevention of invasion of the host cell by foreign DNA, i.e., were components of restriction-modification systems and hence were called restriction endonucleases, there is no evidence that many of the more recently described endonucleases are also bona fide components of restriction-modification systems. Nevertheless, for convenience we will continue to refer to site-specific endodeoxyribonucleases as restriction endonucleases, as suggested by Roberts (1976). All restriction endonucleases *recognize* specific DNA sequences; some, but not all, may also *cleave* at specific sequences. Class II enzymes are of the former type and generate specific DNA fragments usually, but not always, by cleavage at a sequence within the endonuclease recognition sequence. Class I enzymes are of the latter type and appear to cleave DNA randomly, there by generating heterogeneous DNA products. The ability of class II enzymes to cleave long complex molecules of DNA at precise locations and the availability of a wide range of enzymes having different recognition sequences has revolutionized the investigation of the structure and function of genetic material.

Table 1 lists the recognition sequences and cleavage sites of a selection of currently available and widely used restriction endonucleases. It may be noted that endonucleases differ from one another not only in the composition and sequence of bases within the recognition sequence, but also in the length of the sequence and in the type of cleavage effected. In general, an enzyme that recognizes a tetranucleotide sequence will cleave a given DNA molecule more frequently than will an enzyme that recognizes a hexanucleotide sequence. However, the relative frequency of different restriction site sequences and their distribution within a DNA molecule varies from one DNA species to another (contrast the number of cleavage sites in SV40 DNA for the enzymes Alu1 and TaqI, enzymes which recognize mirror image sequences). Thus, the number and location of restriction endonuclease cleavage sites on any given DNA molecule depends upon the sequence of the nucleotides recognized by the enzyme, the length of the sequence, and its frequency and distribution in the substrate DNA molecule.

Some restriction endonucleases cleave both DNA strands at a single site on the molecule (i.e., cleave both phosphodiester linkages between two nucleotide pairs), thus generating DNA fragments having "flush" or "blunt ends", whereas others cleave one DNA strand several nucleotides away from the cleavage on the opposing strand, thereby generating "cohesive" ends. Restriction endonucleases that generate fragments with cohesive ends can produce 5' extensions (EcoRI) or 3' extensions (PstI) composed of a dinucleotide (TaqI), a trinucleotide (HinfI), a tetranucleotide (HindIII), or a pentanucleotide (EcoRII). As can be seen in Table 1, in some cases the same DNA sequence can be

Table 1. Selected restriction endonucleases: recognition sequences and sites of cleavage[a]

	Endonucleases that generate fragments with cohesive termini			Endonucleases that generate fragments with flush termini		
Hexanucleotide recognition	Bam HI	G↓GATCC	(5;1)			
	Bgl II	A↓GATCT	(5;0)	Hpa I	GTT↓AAC	(11;5)
	Eco RI	G↓AATTC	(5;1)	Sma I	CCC↓GGG	(3;0)
	Hind III	A↓AGCTT	(6;6)			
	Pst I	CTGCA↓G	(18;2)			
	Xma I	C↓CCGGG	(3;0)			
	Hae II	PuGCGC↓Py	(>30;1)			
Tetranucleotide recognition	Hha I	GCG↓C	(>50;2)	Alu I	AG↓CT	(>50;32)
	Hpa II	C↓CGG	(>50;1)	Hae III	GG↓CC	(>50;18)
	Mbo I	↓GATC	(>50;8)			
	Taq I	T↓CGA	(>50;1)			

[a]Compiled from the review by Roberts, 1976. For convenience, only one strand of the recognition sequence is shown. The left hand end of each sequence is the 5' end. The arrow indicates the site of cleavage of the phosphodiester linkage. The two numbers shown in brackets after each recognition sequence represent the numbers of cleavage sites present in lambda and SV40 DNAs respectively.

recognized by two or more enzymes that generate DNA fragments with different termini (e.g., SmaI and XmaI), whereas in other cases cleavage at different recognition sequences can generate DNA fragments with identical termini (i.e., BamH1 and BglII; Roberts, 1976a).

Because restriction endonuclease-generated DNA fragments with identical cohesive ends readily anneal with one another they can be efficiently joined together by the action of DNA ligase. This particular property of DNA fragments generated by the EcoRI restriction endonuclease greatly facilitated the early DNA cloning experiments.

C. DNA Cloning

I. In Vitro Recombination

1. Generation of DNA Fragments

To permit insertion of foreign DNA at an appropriate location in the vector, the vector is cleaved at a unique location that is not in or near functions that are essential for its effective use (see below). Multiple cleavages are sometimes made to remove non-essential segments of vector DNA in order to permit packaging of the maximum amount of foreign DNA in bacteriophage vectors (Murray and Murray, 1974; Thomas et al., 1974). However, because of the specificity of cleavage required, vector cleavage is always accomplished by means of a restric-

tion endonuclease. It is usually convenient to generate vector DNA having cohesive termini that can subsequently anneal with similar termini on foreign DNA fragments that are to be cloned.

Cleavage of foreign DNA to be cloned is also most conveniently accomplished by the action of one or more restriction endonucleases. Fragments thus generated may possess either cohesive ends or flush ends, or one of each type, according to the enzyme (s) used. Ordinarily, enzymatic digestion of the DNA is allowed to proceed to completion, i.e., all sites on the DNA molecule that are susceptible to cleavage by the enzyme (s) are cleaved. However, in those instances where all available enzymes cleave within the required DNA fragment, or where the cloning of a series of overlapping fragments of DNA is desired, it is necessary to obtain partial endonuclease digestion products of the DNA (Skurray et al., 1976), or to generate random fragments, for example, by mechanical shearing (Clarke and Carbon, 1975; see, however, Backman et al., 1976).

If a fragment of DNA to be cloned has been identified, it may be possible to purify this fragment prior to cloning, if it has a physical property (e.g., size or buoyant density) that distinguishes it from other fragments. For example, it was possible to purify an EcoRI endonuclease-generated fragment of DNA coding for ampicillin (Ap) resistance derived from the plasmid pSC122 by equilibrium centrifugation in caesium chloride (Timmis et al., 1975). This Ap DNA fragment banded in the gradient according to its buoyant density of 1.692 g/cc, whereas the other DNA fragment of the pSC122 plasmid banded at a more dense part of the gradient owing to its buoyant density of 1.710 g/cc. Similarly, DNA fragments of different sizes may be purified by electrophoresis through agarose gels (Lovett and Helinski, 1976). The recent description of a method (R loop formation) to form stable hybrids of RNA and duplex DNA should facilitate the purification of DNA fragments containing sequences for which a purified complementary RNA probe is available (Thomas et al., 1976). Prior purification of the fragment to be cloned greatly facilitates the subsequent cloning procedures.

2. Joining of DNA Fragments

The in vitro covalent linkage of fragments of foreign DNA to a cloning vector is achieved by treatment with *E.coli* or T4 DNA ligase to form phosphodiester linkages between the DNA fragments. The most convenient method currently employed is to ligate foreign DNA and vector fragments having identical cohesive ends (Mertz and Davis, 1972; Sgaramella, 1972; see Fig. 2). Under appropriate reaction conditions, the cohesive ends anneal and permit efficient covalent linkage by DNA ligase. While cohesive termini are convenient for some experiments, they are not absolutely required for the linkage of DNA fragments. Although the termini of flush ended DNA fragments cannot anneal they nevertheless can be joined by the action of T4 DNA ligase if the ligase and DNA fragment ends are present at high concentration in the ligation reaction (Sgaramella et al., 1970; Sgaramella and Khorana, 1972).

The recovery of a cloned DNA fragment from a hybrid molecule, subsequent to construction, is an important consideration in determining the strategy of a cloning experiment. Fragment recovery is always possible if the cloning vector and the foreign DNA fragment have been generated by the same endonuclease, since ligation of two ends produced by a single enzyme will regenerate the original endonuclease recognition sequence. If the cloning vector and foreign DNA are cleaved by different restriction enzymes and subsequently ligated together, in most in-

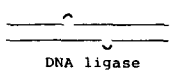
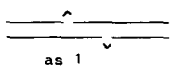
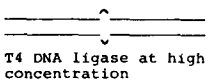
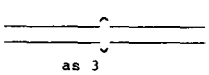
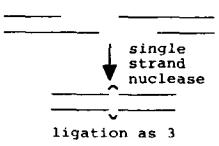
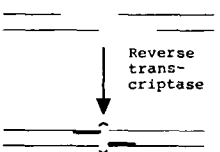
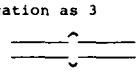
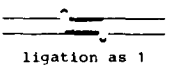
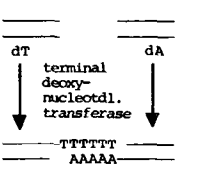
	Recognition Sequence	Type of Ends	Joining	Recovery of Cloned Fragment
1.	Identical	Identical, cohesive	 DNA ligase	Yes; with original cloning enzyme (e.g. EcoRI)
2.	Different	Identical, cohesive	 as 1	Not with cloning enzymes. Hybrid recognition site generated (e.g. BamH1, BglII)
3.	Identical	Identical, flush	 T4 DNA ligase at high concentration	Yes; with original cloning enzyme (e.g. SmaI)
4.	Different	Identical, or different, flush	 as 3	Not with cloning enzymes. Hybrid recognition site generated (e.g. AluI, SmaI)
5.	Different	Different, cohesive	a)  single strand nuclease ligation as 3	Not usually possible; no specific sequence at joint
			b)  Reverse trans-criptase ligation as 3	Yes, if one of the original recognition sequences is regenerated (e.g. EcoRI, XmaI); with one cloning enzyme
6.	Identical, or different	Different, one cohesive, one flush	Convert cohesive ends to flush as in 5, ligation as 3 	as 5
7.	Use of *adaptor fragments*		Ligation to *adaptor fragments*  ligation as 1	Yes, with enzyme that cleaves *adaptor fragment*
8.	Homopolymer tailing		 ligation unnecessary	Yes, if one of the original recognition sequences is regenerated (e.g. if SmaI or PstI cleaved DNA is tailed with dG)

Fig. 2. Methods of joining DNA fragments

stances a hybrid recognition sequence will be formed that cannot be recognized by either of the original endonucleases. Unless an enzyme is available that recognizes the hybrid sequence, precise recovery of the cloned fragment will not be possible.

Occasionally, cleavage of the cloning vector and foreign DNA with different enzymes will generate DNA fragment with two different types of cohesive ends or with one type of cohesive end and one type of blunt end. In such situations, prior to ligation it is necessary either to generate appropriate cohesive ends in both types of DNA fragment or to convert both types of fragment to the flush-ended form.

The synthesis of *universal* cohesive ends on DNA fragments having any type of end (flush, or with 5' or 3' terminal extensions) is readily accomplished by means of the enzyme terminal deoxynucleotidyl transferase (Bollum, 1974). In the presence of cobalt ions this enzyme can add long homopolymer blocks composed of any one of the four deoxyribonucleotides to the 3' terminus of DNA fragments (Roychoudhury et al., 1976). Thus, a series of identical deoxyribonucleotides (e.g., dA) is added to the 3' ends of the cloning vector and a series of complementary deoxyribonucleotides (e.g., dT) is added to the 3' ends of the DNA fragments to be cloned (so-called *homopolymer tailing*; Jackson et al., 1972; Lobban and Kaiser, 1973). The two types of fragment are then mixed, allowed to anneal, and introduced into a bacterial host. Because it is not possible to synthesize homopolymer blocks of precisely defined length, DNA fragments annealed at their homopolymer tails will contain single-stranded regions at these joints. Such single-stranded regions may be repaired and covalent linkage of the fragments effected in vitro by the action of exonuclease III, DNA polymerase I and DNA ligase or in vivo by cellular repair enzymes (Jackson et al., 1972; Lobban and Kaiser, 1973; Wensink et al., 1974; Clarke and Carbon, 1975).

This particular method has certain advantages over other methods of joining DNA fragments. Firstly, circularization of single fragments, i.e., *intra*molecular joining, is absolutely precluded because any particular fragment has identical terminal homopolymer tracts. Thus, regeneration of the original cloning vehicle with the resultant elevation of "background" *non-hybrid* molecules is obviated. Secondly, because association of the long annealed homopolymer tails is quite stable, in vitro ligation of the cloning vehicle and the fragment to be cloned is unnecessary. This specific advantage is partly offset by the fact that precise cleavage of hybrid molecules at the vector:cloned fragment junction is rarely possible. If the cloned fragment can be recovered, due to the fortuitous location of appropriate restriction enzyme cleavage sites on the vector or fragment side of the homopolymer tract, either it will contain the homopolymer tract or it will be missing some of its terminal sequences. Recently, the use of single-stranded endonucleases to attack partially denatured dA-T rich segments has provided a simple method of separation of fragments that have been joined together by dA-T homopolymer tailing (Hofstetter et al., 1976). In some instances, judicious choice of the nucleotide of the homopolymer block to be added to the fragments of foreign DNA can regenerate the recognition sequence of the enzyme orginally used to obtain these DNA fragments (for instance, the addition of poly dG tails to SmaI-generated DNA fragments), and can thus permit recovery of the cloned fragment.

Conversion of DNA fragments containing 5' terminal extensions to the flush-ended form is accomplished either by removal of the extension with a single strand-specific exonuclease or by synthesis of a complementary sequence to the 5' termini (*filling-in*) by extension of the 3'