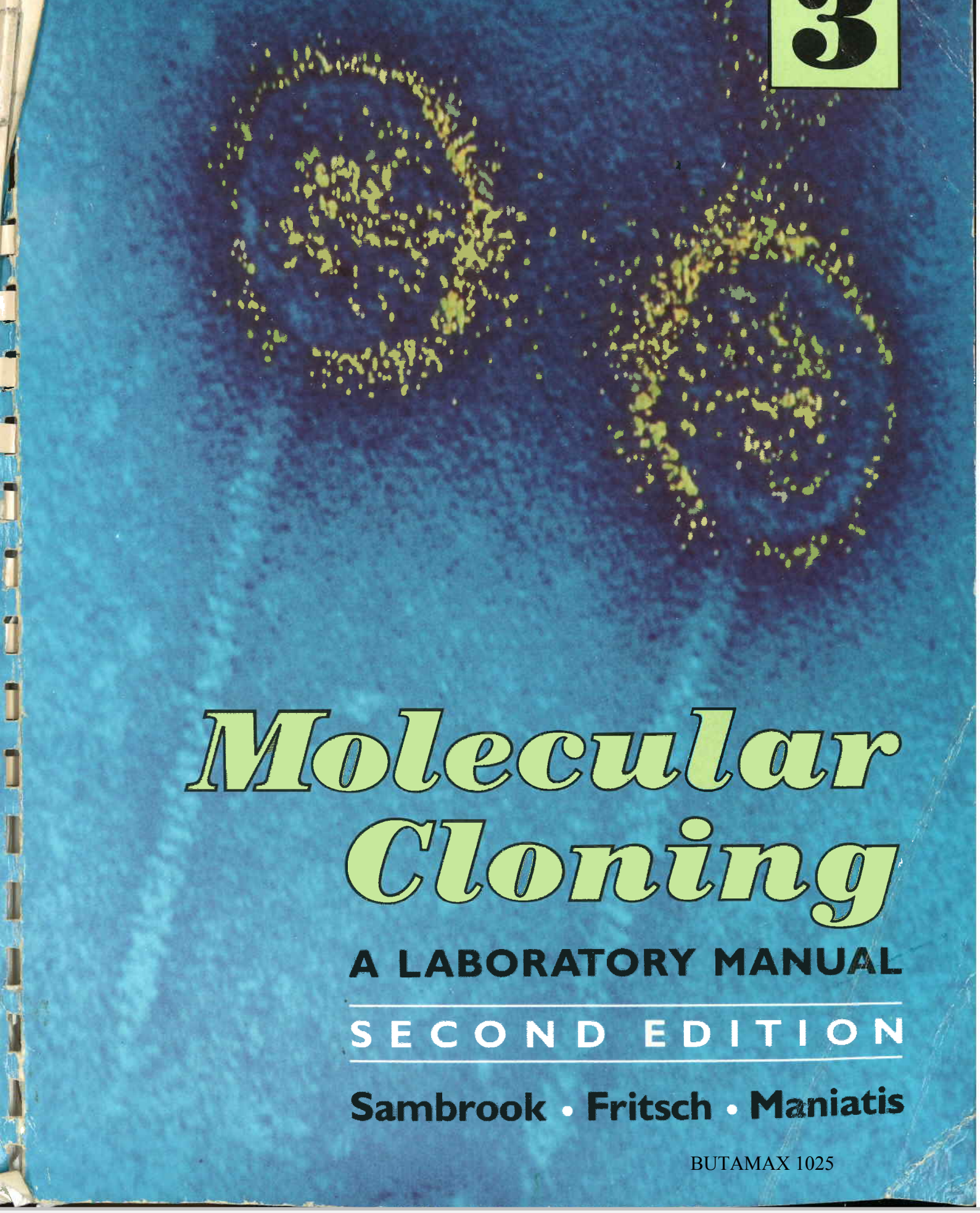

3

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

Sambrook • Fritsch • Maniatis

BUTAMAX 1025

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

Molecular Cloning

A LABORATORY MANUAL

SECOND EDITION

All rights reserved
© 1989 by Cold Spring Harbor Laboratory Press
Printed in the United States of America

9 8 7 6 5 4 3 2

Book and cover design by Emily Harste

Cover: The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. (Thomas R. Broker, Louise T. Chow, and James I. Garrels)

Cataloging in Publications data

Sambrook, Joseph
Molecular cloning : a laboratory manual / E.F. Fritsch, T. Maniatis—2nd ed.
p. cm.
Bibliography: p.
Includes index.
ISBN 0-87969-309-6
1. Molecular cloning—Laboratory manuals. 2. Eukaryotic cells--Laboratory manuals. I. Fritsch, Edward F. II. Maniatis, Thomas III. Title.
QH442.2.M26 1987
574.87'3224—dc19 87-35464

Researchers using the procedures of this manual do so at their own risk. Cold Spring Harbor Laboratory makes no representations or warranties with respect to the material set forth in this manual and has no liability in connection with the use of these materials.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Cold Spring Harbor Laboratory Press for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of $0.10 per page is paid directly to CCC, 21 Congress St., Salem MA 01970. [0-87969-309-6/89 $00 + $0.10] This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724. Phone: 1-800-843-4388. In New York (516)367-8423.

BUTAMAX 1025

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

TATA box and the *upstream promoter elements*. The TATA box, located 25–30 bp upstream of the transcription initiation site, is thought to be involved in directing RNA polymerase II to begin RNA synthesis at the correct site. In contrast, the upstream promoter elements determine the rate at which transcription is initiated. These elements can act regardless of their orientation, but they must be located within 100 to 200 bp upstream of the TATA box. *Enhancer elements* can stimulate transcription up to 1000-fold from linked homologous or heterologous promoters. However, unlike upstream promoter elements, enhancers are active when placed downstream from the transcription initiation site or at considerable distances from the promoter. Many enhancers of cellular genes work exclusively in a particular tissue or cell type (for review, see Voss et al. 1986; Maniatis et al. 1987). In addition, some enhancers become active only under specific conditions that are generated by the presence of an inducer, such as a hormone or metal ion (for review, see Sassone-Corsi and Borrelli 1986; Maniatis et al. 1987). Because of these differences in the specificities of cellular enhancers, the choice of promoter and enhancer elements to be incorporated into a eukaryotic expression vector will be determined by the cell type(s) in which the recombinant gene is to be expressed. Conversely, the use of a prefabricated vector containing a specific promoter and cellular enhancer may severely limit the cell types in which expression can be obtained.

Many enhancer elements derived from viruses have a broader host range and are active in a variety of tissues, although significant quantitative differences are observed among different cell types. For example, the SV40 early gene enhancer is promiscuously active in many cell types derived from a variety of mammalian species, and vectors incorporating this enhancer have consequently been widely used (Dijkema et al. 1985). Two other enhancer/promoter combinations that are active in a broad range of cells are derived from the long terminal repeat (LTR) of the Rous sarcoma virus genome (Gorman et al. 1982b) and from human cytomegalovirus (Boshart et al. 1985).

TERMINATION AND POLYADENYLATION SIGNALS

During the expression of eukaryotic genes, RNA polymerase II transcribes through the site where polyadenylation will occur. Consequently, the 3′ terminus of the mature mRNA is formed by site-specific posttranscriptional cleavage and polyadenylation (for review, see Birnstiel et al. 1985; Proudfoot and Whitelaw 1988; Proudfoot 1989). Although discrete sites for the termination of the primary transcript have not yet been identified, general regions of DNA a few hundred nucleotides in length and downstream from the polyadenylation site have been identified where transcription randomly terminates.

Two distinct sequence elements are required for accurate and efficient polyadenylation: (1) GU- or U-rich sequences located downstream from the polyadenylation site and (2) a highly conserved sequence of six nucleotides, AAUAAA, located 11–30 nucleotides upstream, which is necessary but not sufficient for posttranscriptional cleavage and polyadenylation (for review, see Mason et al. 1986; Proudfoot and Whitelaw 1988). The practical implication of these observations is that sequences downstream from the polyadenyl-

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

ation site must be included in eukaryotic expression vectors to ensure efficient polyadenylation of the mRNA of interest. Although a full-length cDNA clone may encode the conserved AAUAAA sequence and a tract of poly(A), these endogenous elements are not by themselves sufficient to guarantee polyadenylation. The downstream GU- or U-rich sequences necessary for cleavage and polyadenylation must therefore be incorporated into the vector. The most frequently utilized signals are those derived from SV40; a 237-bp *Bam*HI-*Bcl*I restriction fragment contains the cleavage/polyadenylation signals from both the early and the late transcription units. These signals are positioned in opposite orientations, one on each DNA strand, and both sets of signals have been shown to be extremely efficient for the processing of hybrid mRNAs. Less frequently, polyadenylation signals have been provided by fusing a full-length cloned cDNA onto a partial genomic copy of a gene already resident in an expression vector (O'Hare et al. 1981; Kaufman et al. 1986b).

Sequences within the 3′ noncoding regions of eukaryotic genes may play a role in mRNA stability. For example, the presence of an AU-rich sequence, derived originally from the 3′ noncoding region of granulocyte-macrophage colony-stimulating factor (GM-CSF), has been shown to destabilize mRNAs transcribed from mammalian expression vectors (Shaw and Kamen 1986). Although similar motifs have been found in analogous locations within mRNAs encoding a variety of growth factors and oncogenes, relatively little is known about the way they function. To obtain maximal expression of a cloned gene, it may therefore be necessary to remove the nucleotide sequences 3′ of the termination codon.

SPLICING SIGNALS

The DNA sequences coding for a eukaryotic protein are rarely contiguous; usually, they are separated in the genome by intervening noncoding sequences that may vary in size from tens to many thousands of nucleotides. Following polyadenylation of the primary transcript, the introns are removed by splicing to generate the mature mRNA, which is then transported from the nucleus to the cytoplasm (for review, see Nevins 1983; Green 1986; Padgett et al. 1986; Krainer and Maniatis 1988).

The minimal sequences required for splicing of mRNA are located at the 5′ and 3′ boundaries of the intron. Comparison of a large number of these sequences has led to the identification of consensus sequences in which the first two and the last two nucleotides of the intron are essentially invariant:

A G : **G U** (A) A G U . . . intron . . . (U/C) N_{11} C **A G** : G
5′ splice site 3′ splice site

The development of in vitro splicing systems has led to the elucidation of much of the biochemistry of the splicing reaction, but the processes that guarantee correct matching of 5′ and 3′ splice sites are not yet understood. The fact that hybrid pre-mRNAs containing 5′ and 3′ splice sites derived from different introns can be accurately spliced (Chu and Sharp 1981) indicates the importance of the conserved consensus sequences in this process. However, these sequences cannot be the sole determinants of splice-site selection, since identical, but ordinarily inactive, consensus sequences can be

BUTAMAX 1025

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

developed that express the Tn*5* *neo*^{r} gene under the control of SV40 regulatory elements (Chia et al. 1982; Southern and Berg 1982; Okayama and Berg 1983; Van Doren et al. 1984). Vectors such as pSV2-*neo* (Southern and Berg 1982) and pRSVneo (Figure 16.1C), which have been widely used in cotransformation experiments, contain a version of the Tn*5* *neo*^{r} gene that retains prokaryotic promoter sequences between the eukaryotic promoter and the APH coding sequences. This configuration yields a vector that can confer antibiotic resistance upon both prokaryotic and eukaryotic cells. However, perhaps because the bacterial promoter contributes several upstream AUG codons, the efficiency of translation of APH mRNAs synthesized from these vectors is comparatively low in mammalian cells (Chen and Okayama 1987). Vectors such as pko-neo (Figure 16.1D) (Van Doren et al. 1984) and pcDneo (Okayama and Berg 1983; Chen and Okayama 1987), which lack prokaryotic promoter sequences, are therefore preferred.

- *Hygromycin B phosphotransferase*. The *E. coli* gene encoding hygromycin B phosphotransferase (Gritz and Davies 1983) can be used as a dominant selectable marker in much the same way as the APH gene. When the hygromycin B phosphotransferase gene (*hyg*) is introduced into mammalian cells on an appropriate expression vector (e.g., pHyg, Figure 16.1E) (Sugden et al. 1985), the transfected cells become resistant to the antibiotic hygromycin. Resistance to neomycin and to hygromycin can be selected for independently and simultaneously in cell lines that have been transfected with both genes. Thus, two different vectors can be introduced into one cell line, either simultaneously or sequentially.

- *Xanthine-guanine phosphoribosyl transferase*. The *gpt* gene of *E. coli* encodes the enzyme xanthine-guanine phosphoribosyl transferase (XGPRT), which is the bacterial analog of the mammalian enzyme hypoxanthine-guanine phosphoribosyl transferase (HGPRT). Whereas only hypoxanthine and guanine are substrates for HGPRT, XGPRT will also efficiently convert xanthine into XMP, which is a precursor of GMP. The bacterial *gpt* gene has been cloned and expressed in mammalian cells under the control of an SV40 promoter (Mulligan and Berg 1980, 1981a,b) (see, e.g., Figure 16.1F). Vectors expressing XGPRT restore the ability of mammalian cells lacking HGPRT activity to grow in HAT medium (Szybalska and Szybalski 1962; Littlefield 1964, 1966).

  Of much greater general use is the application of the *gpt* gene as a dominant selection system, which can be applied to any type of cell (Mulligan and Berg 1981a,b). Vectors expressing XGPRT confer upon wild-type mammalian cells the ability to grow in medium containing adenine, xanthine, and the inhibitor mycophenolic acid. Mycophenolic acid blocks the conversion of IMP into XMP and inhibits the de novo synthesis of GMP. The selection can be made more efficient by the addition of aminopterin, which blocks the endogenous pathway of purine biosynthesis.

- *CAD*. A single protein, CAD, possesses the first three enzymatic activities of de novo uridine biosynthesis (carbamyl phosphate synthetase, aspartate transcarbamylase, and dihydroorotase). Transfection of vectors expressing the CAD protein from Syrian hamsters into CAD-deficient (UrdA) mutants of CHO cells allows selection of CAD$^+$ transfectants that are able to grow in the absence of uridine (Robert de Saint Vincent et al. 1981).

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.