Preface

Since its introduction in 1977 (2, 4), the DNA sequencing process has become a routine and powerful laboratory technique, and over the years it has evolved significantly. In the early 1980s it took two people almost two and a half years to sequence the 40,000 kbp genome of bacteriophage T7 (J. Dunn, personal communication). Now the draft sequencing of a bacterial genome of 2 to 4 Mb takes a day and the sequencing of the entire human genome (2.9 billion base pairs) in a draft form was accomplished in under a year (1, 5). This feat was possible largely because of the biochemical and technological advances induced by the Human Genome Project initiative (3). Some of these advances are described in more detail in Chapters 4, 9, and 13.

Even with these unquestionable advances and the almost routine nature of the DNA sequencing process, a number of issues remain to be solved in order to close remaining gaps in any genome. The recent focus on non-standard approaches and techniques for sequencing difficult DNA templates was evident in a number of posters presented during the last three Genome Sequencing and Analysis Conferences in San Diego (2001), Boston (2002), and Savannah (2003). Hopefully, some of these posters will be further refined and published for general use. Chapter 3 describes methods of dealing with difficult templates.

In Chapter 1, controlled heat-denaturation of plasmids is described. In itself, the incorporation of this step into a DNA sequencing protocol is an effective way to overcome failures due not only to the nature of a specific template but also to a particular host strain, the presence of external DNA, or an insufficient amount of DNA. Chapter 2 concentrates on the effects of some widely used laboratory reagents on DNA sequencing.

The success of DNA sequencing is primarily dependent on the quality and quantity of the template. Some proven and newly emerging methods of preparing DNA templates are described in Chapter 7. Chapter 8

presents optimization of conditions for growing cell cultures in 96-well deep blocks to obtain the optimal amount of plasmid DNA. In recent years, a number of cloning vectors have been developed or modified to accommodate different needs such as larger insert sizes, cloning previously unclonable fragments, or cloning "without cutting," etc. Some of these vectors are described in two articles in Chapters 5 and 6. The massive amount of sequencing data and the need for easy access to and manipulation of these data have induced many small and medium-sized DNA sequencing facilities to incorporate some form of electronic management of their respective operations. One of the commercially available LIMS (Laboratory Information Management Systems) packages is described in Chapter 10. These LIMS packages are sold "as is" and quite often need modification to adapt to a particular lab setting. We also discuss one LIMS system (Chapter 11) developed in-house that is particularly geared towards a DNA finishing core facility. Recognizing the need for "base-perfect" DNA sequencing, in fact the Federal Drug Administration's requirements (if DNA is involved) when approving a new drug, we provide a guide on how to set up and run a GLP/GMP DNA sequencing laboratory (Chapter 12). We conclude this book with a chapter (Chapter 13) on the future of DNA sequencing, concentrating primarily on technological aspects and desirable new chemistries. The challenge of achieving a $1000 (or less) cost per genome has spurred many new initiatives and this chapter takes a snapshot of the cutting edge of DNA sequencing technologies at this point in time. Without doubt, novel technologies will be developed that will make the sequencing of any genome a routine, cost-effective benchtop exercise.

This book is intended to be a practical guide for those who routinely sequence DNA. We hope that it will help to refine and enhance your operation and serve as a starting point for your own inquiries into the intricacies of the DNA sequencing process. We also hope to incorporate the newest advances in DNA sequencing in future editions and would appreciate comments and suggestions for further improvements.

Note: This book is not a textbook, and it does not provide an exhaustive list of references at the end of each chapter. We apologize for any omissions.

Acknowledgments

The early 1990s witnessed an explosion of interdisciplinary scientific activities as a result of the DOE- and NIH-sponsored human genome program, and I was truly fortunate to be in the right spot at the right time. While developing hexamer-based sequencing methodology under the

incomparable and watchful Bill Studier, I learned the true meaning of "the devil is in the details." At the same time, I had the chance to interact closely with John Dunn whose creativity and enthusiasm continually pushed me to retry, retest, and question seemingly established and proven protocols. It is not too difficult to imagine that without the influence of these two great mentors this book would not have been possible. I am also thankful to Mike Blewitt, Barbara Lade, Jutta Paparelli, and Judy Romeo for all the scientific and technical help I received while working at Brookhaven National Laboratory.

After moving to Genetic Institute in Cambridge, MA (now Wyeth Research), I again found myself in an environment conducive to further inquiries into the range of issues related to DNA sequencing technology. At the center of my interest was, and continues to be, the development of methods to deal with various difficult DNA templates. I wish to thank the entire staff of the DNA sequencing group for their help in picking up "the slack" and for other technical help while I was busily testing. Special thanks goes to Lori Haines for her critical and editorial review of these many chapters. The management at Wyeth's Genomics Department deserves my gratitude for their help and encouragement while working on this project.

This book would not be what it is without the excellent chapters contributed by so many outstanding scientists. Thank you so much, and my apology for the evasive and not so encouraging answer "soon" all those times you asked when will the book be published.

I still remember the day when, with slight apprehension, I contacted Stephen Weaver, the science editor of this volume, at Jones & Bartlett Publishers. His instantaneous and enthusiastic encouragement gave me a great boost to continue working on the book. I am also grateful to the entire editorial and production staff at Jones & Bartlett, including Rebecca Seastrong, Anne Spencer, Lou Bruno, and Dean DeChambeau.

Finally, I wish to thank my wife Carla and my children Kasia, Alex, and Michael for their patience and understanding during these past long months when I was there but not really there.

Jan Kieleczawa
Wyeth Research
Cambridge, Massachusetts
June 2004

1. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
2. Maxam, A.M., and Gilbert, W. 1977. A new method of sequencing DNA. *Proc Natl Acad Sci USA* 74:560–564.
3. National Research Council. 1988. *Mapping and Sequencing the Human Genome.* National Academy Press. Washington, DC.