

The Universal Primers and the Shotgun DNA Sequencing Method

Joachim Messing

1. Introduction

For studies in molecular biology, DNA purification has been essential, in particular for DNA sequencing, probing, and mutagenesis. The amplification of DNA in *Escherichia coli* by cloning vehicles derived from M13mp or pUC made expensive physical separation techniques such as ultracentrifugation unnecessary. Although today the polymerase chain reaction (PCR) is a valuable alternative for the amplification of small DNA pieces (*1*), it cannot substitute for the construction of libraries of DNA fragments. Therefore, *E. coli* has served not only as a vehicle to amplify DNA, but also to separate many DNA molecules of similar length and the two DNA strands simultaneously. For this purpose, a bacteriophage such as M13 can be used. The various viral *cis*- and *trans*-acting functions are critical not only for strand separation, but also to separate the single-stranded DNA from the *E. coli* cell by an active transport mechanism through the intact cell wall.

Although it may have been somewhat surprising to some how many changes in its DNA sequence the phage tolerated, manipulations of this amplification and transport system have been extended

From: *Methods in Molecular Biology*, vol. 167: *DNA Sequencing Protocols*, 2nd ed.
Edited by: C. A. Graham and A. J. M. Hill © Humana Press Inc., Totowa, NJ

today, even to the viral coat proteins for the production of epitope libraries (2). Much of the work is now more than a decade old, but experience has confirmed the usefulness of some simple biological paradigms. Techniques that were new and limiting 20 years ago included automated oligonucleotide synthesis and the use of thermostable enzymes, which add a critical dimension to molecular biology today. Neither necessarily replaces the previous techniques, but they create greater flexibility, enormously accelerate our investigations, and even make certain analyses possible for the first time.

2. Emergence of DNA Synthesis Applications

In 1974, at one of the first meetings on the use of restriction endonucleases in molecular biology DNA synthesis and sequencing methods had broken new grounds. Work on the chemical synthesis of a *tRNA* gene was presented, and the initial work on sequencing phage Φ X174 using restriction fragments as primers for the so-called plus-minus method was discussed. At that time, oligonucleotide synthesis required a major effort and could not be applied easily in general. Restriction fragments offered an alternative. They could be used as primers for DNA synthesis *in vitro* and for marker-rescue experiments to link genetic and physical maps of viruses such as SV40, both forerunners for DNA sequencing and site-directed mutagenesis.

There were several reasons to use Φ X174 as the first model in developing DNA-sequencing techniques and determining the sequence of an entire autonomous genome. First, it was one of the smallest DNA viruses; it is even smaller than M13. Second, the mature virus consists of single-stranded DNA, eliminating the need to separate the two strands of DNA for template preparation; this is even more critical if one wishes to use double-stranded restriction fragments as primers. Third, a restriction map was superimposed on the genetic map by marker-rescue experiments (3). Physical mapping still serves today as a precondition for sequencing other genomes. Restriction sites were critical as signposts along the thousands of nucleotides and provided the means to dissect the double-

stranded replicative form or RF of Φ X174 in small but ordered pieces that permitted the DNA sequencing effort to proceed in a walking manner along the genome. Today the restriction map can be replaced by any DNA sequence (e.g., STS) because the synthesis of oligonucleotides is so rapid that we can use the DNA sequence that was just read from a sequencing gel to design and produce an oligonucleotide to extend the sequencing gel further in the 5' direction. Therefore, the use of oligonucleotides instead of restriction fragments in such a primer walking method would have enormously accelerated the Φ X174 project.

At the Cleveland Conference on Macromolecules in 1981, after a talk that I had given, the replacement of shotgun sequencing by such a method was suggested by Marvin Caruthers, who, as a pioneer in DNA synthesis and its automation (4), saw a perfect match of this emerging technology with DNA sequencing. Another expert in the chemical synthesis of DNA, Michael Smith, recognized the potential of the chemistry as a DNA mutagen (5). Design of oligonucleotides as mutagens, however, requires the knowledge of the target DNA sequence. Therefore, the location of amber mutations in Φ X174 genome and the use of restriction fragments to rescue them was critical not only to delineate a physical and genetic map for DNA sequencing, but also to develop oligonucleotide site-directed mutagenesis (3). It is clear that today's protein engineering had its roots right there with the right people at the right time. It was also possible to show that oligonucleotides eliminated the need for strand separation for DNA sequencing (6). Even double-stranded DNA sequencing with universal primers became easier with the development of the pUC plasmids (7). However, application of oligonucleotide site-directed mutagenesis to cloned DNA was also greatly accelerated by the development of single-stranded DNA cloning vectors (8,9).

Despite all the advantages of choosing Φ X174 as a model system, Sanger's group nearly picked a different single-stranded DNA phage, fd. In principle, *E. coli* has two different types of single-stranded DNA phage, represented by Φ X174 and fd. The first is packaged into an icosahedral head; it kills and lyses the host cell,

but does not require F pili, which are receptor sites on the surface of the cell wall encoded by F factors. Its host range is restricted to *E. coli* C. Phage such as fd can infect only male-specific *E. coli* producing pili at their surface that are packaged in a filamentous coat and discharged from the cell without lysis; infected cells can continue to divide. These differences are critical, but there was another reason for choosing fd originally. The major coat protein encoded by gene *VIII* of the phage, a very small but very abundant protein, had been sequenced by protein sequencing methods. Therefore it seemed obvious, particularly to someone who had pioneered protein sequencing, to use protein sequence to check the DNA sequence. The protein sequence allowed the design and synthesis of an oligonucleotide that would prime in vitro DNA synthesis within the coat protein gene. Furthermore, the derived DNA sequence had to match the protein sequence. Of course, the codon redundancy of many amino acids made it difficult to design a unique primer, and it might not have been too surprising that the approach did not lead to the correct DNA sequence (*10*). It turned out later that this was owing not to the choice of codons, but to a mistake in the protein sequence. Still, cloning based on protein sequence information again has its roots right there.

3. Replication Systems and M13

In 1974, there was a great interest in understanding how DNA is replicated. Viruses and plasmids were used to identify components of the cell machinery required for the initiation and elongation process of DNA replication. These studies were also pursued at the “Abteilung” of Virology at the Max Planck Institute of Biochemistry in Munich. Eleven years earlier, Hofschneider isolated a filamentous phage from the Munich sewers that he named after a series of phage with the initial M (*11*). Number 13 was the one that was studied most. Looking for a different research topic than DNA replication, I explored the possibility of combining M13 phage production with the in vitro DNA synthesis-based method of DNA sequencing.

Although this might have been obvious to those familiar with phage replication, innovative methods were needed for adaptation to DNA cloning techniques. The walking method for sequencing Φ X174 was the strategy used at that time, and I thought that it would be difficult to clone large fragments into M13 (although the author's record was approx 40 kb) and that a walking method might therefore have a limited use. Logically, the only alternative to the walking method was the use of shotgun cloning and a universal primer. The replicative form of M13 could be used to clone DNA fragments of a size slightly larger than necessary for single sequencing reactions, and a universal sequence near the cloning site would be used as a primer. This would shift the work from preparing primers to preparing templates, which has become essential for high-throughput DNA sequencing (12). If they were numerous, cloning was much faster than any biochemical technique. In view of these considerations, a plan took shape to construct in vitro recombinants of phage M13 using a different method from any other cloning methods. One might recall that in vitro recombinants were usually based on drug-resistance markers. This led to the development of plasmid vectors with unique cloning sites that were scattered all over the plasmid genome (13). However, such an approach is not compatible with the life cycle of M13.

4. New Mutants of Filamentous Phage

Both Schaller's and Zinder's laboratory considered, and later used, transposons to develop f1 and fd transducing phage (14,15). However, one could predict that such a course of experiments, although useful for plasmid cloning vehicles, would be less useful for M13. It appeared plausible, and such an experiment could demonstrate that, in contrast to Φ X174, filamentous phage can accommodate additional DNA by extending the filamentous coat; infected cells can be treated like plasmid-containing cells. To some degree this had already been proven since phage mutants of more than unit length had been described (16). Another advantage of transposon mutagenesis was that insertion mutants would be naturally selected.

Selection of an insertion site in the phage genome was one of the greatest obstacles from the beginning. Although plasmids and bacteriophage λ were natural transducing elements, filamentous phage had never been shown to have this property, and it was not obvious whether insertion mutants would be viable. This was difficult particularly because it was already known that amber mutants of most viral genes not only cause abortive infection, but also lead to killing of the host, which does not occur when these mutations are suppressed. Therefore, insertion mutants that are treated like plasmids would kill the cell.

Thus, it was predictable that insertion mutants had to be restricted to noncoding regions. Therefore, a decision was made to use a restriction enzyme that recognized at least two different sites in the intergenic region of the RF. Rather than asking whether the intergenic region contains a target site for transposons, it was decided to look at the restriction map, which showed that it was possible to obtain at least two different insertion mutants within the intergenic region. The only difficulty in such an experiment was to find conditions where the restriction enzyme would cut RF only once but at any of the possible target sites, so that a population of unit length RF could be ligated to the appropriate marker DNA fragment.

However, there was another reason not to use drug-resistance markers. Infected cells still divide, but very slowly. Therefore, selection takes much longer than with plasmids, but it makes it very easy to distinguish infected from noninfected cells on a bacterial lawn. A single infection grown on a bacterial lawn forms a turbid plaque. If bacterial cells are transfected by the calcium chloride technique of Mandel and Higa (17), a transformed cell can be recognized as a plaque. Hence, no selection technique is necessary. Still, how would one be able to distinguish between wild-type M13 and M13 insertion mutants? Although it was quite plausible to think of the histochemical screen used for bacteriophage λ *plac* by Malamy et al. (18), the *lacZ* gene would have been a large insertion. However, rather than using entire genes as markers, one could clone only the portion encoding the amino (N)-terminal and the repressible control region and provide the rest *in trans* by the host of the phage.

This became clear when Landy et al. (19) wrote on the purification of a 800 bp *Hind*II fragment from λ plac capable of α -complementation in a cell-free transcription–translation system.

An informal sequence of this fragment showed that it was 789 bp long and included the first 146 codons of the *lacZ* gene, but it was necessary to assemble many components and purify several restriction endonucleases. Work began after some strains and purified *lac* repressor were traded: This allowed the purification of the 789 bp *Hind*II fragment out of about 50 other restriction fragments by simply filtering the DNA–*lac* repressor complex through a nitrocellulose filter. After adding isopropylthiogalactoside (IPTG), it was possible to recover the DNA in solution. Using DNA binding proteins for purifying and cloning promoter regions has become a very important technique today. Another well-established procedure, the ligation of restriction fragments via blunt ends, was also untested when this experiment was ready. This explains why only two transformants were obtained (20)—one of them was saved and named M13mpl (mp for Max Planck Institute, M for Munich). Electron microscopy proved that added DNA was packaged as a filamentous phage and produced as single-stranded DNA (21,22).

Now the path took a more formal shape. Not only would the histochemical screen work by detecting a blue among colorless plaques, but it could also be reversed. One uncertainty was how to introduce new restriction sites in the right region. Such a site had to be unique for M13mpl and positioned not somewhere in the viral genome, but in the *lacZ* region, so that insertion mutants would not give rise to blue plaques. Inspection of the sequence showed that there were not many sequences in the N-terminal region that could be converted in a single step into a unique restriction site. Attempts to use *Eco*RI linkers that became available at the time to “marker rescue” them did not succeed, probably because they were too short. Without somebody to synthesize a customized oligonucleotide homologous to the *lac* region, it was fruitless to continue this approach. As an alternative, a chemical mutagen seemed to be more practical. It was known that methylated G could mispair with uracil or thymine. Therefore, by methylating the single-stranded M13

The first ten codons of the *lacZ* gene

1	2	3	4	5	6	7	8	9	10	M13mp1
ATG	ACC	ATG	ATT	ACG	<u>GAT</u>	<u>TCA</u>	CTG	GCC	GTC	(+ or viral strand)
					GG ^m AT	TC				(+ or viral strand)
					CT	TA	AG			(- or complementary strand)

1	2	3	4	5	6	7	8	9	10	M13mp2
ATG	ACC	ATG	ATT	ACG	<u>AAT</u>	<u>TCA</u>	CTG	GCC	GTC	(+ or viral strand)
					<u>EcoRI</u>					

Fig. 1. Creation of an *EcoRI* site by chemical mutagenesis. By screening the nucleotides of the first 10 codons of the *lacZ* gene, we found that the sequence GGATTC could be converted into either an *EcoRI* site GAATTC or a *BamHI* site GGATCC by a single base change. Because a *BamHI* site was already present in gene III, but no *EcoRI* site in M13mp1, and I had an ample supply of *EcoRI* enzyme purified in my laboratory, we decided to select the GAATTC site, which also changed codon GAT for aspartic acid to AAT for asparagine (23).

DNA with nitrosomethylurea, a mutation could be introduced into the minus strand and the subsequent RF molecules (Fig. 1).

Unfortunately, there was no good genetic selection for this procedure. It would require brute force methods of enriching *EcoRI*-sensitive RF from a transformed phage library by gel electrophoresis of linear vs circular molecules. Still, it was difficult to believe at the time when a mutant M13 RF was isolated that was not only sensitive to *EcoRI*, but also had exactly the predicted base change in codon 5 of the *lacZ* gene (23). The same mutagenesis led to two more *EcoRI* mutants and a mutant RF that was resistant to *BamHI*, a site within gene III. At that time, there was still much concern that many mutations might not be tolerated because of changes in the protein sequence or the secondary structure of RNA, but this *BamHI* site served later on as a cloning site for epitope libraries (2). Still, changes in the *lac* DNA should probably occur at a higher frequency than in the viral DNA. Furthermore, the N-terminus of the *lacZ* gene appeared to be more flexible because it was demonstrated that fusion proteins retained β -galactosidase function. On the other hand, the

tremendous selection power for suppressor mutations of phage insertion mutants remained elusive for many years. In other words, any mutation that was introduced in one location of the viral genome could potentially be compensated for by a second mutation somewhere else. The primary mutant might give a low titer, but because of the growth advantage a suppressor mutant would take over rapidly. An example of such a case is M13mpl. Dotto and Zinder (24) showed that insertion mutants at the *mpl* *Hae*III site gave a low titer phenotype. Since M13mpl gave a normal titer, they searched for a suppressor mutation. Codon 40 in gene *II* of M13mpl indeed was changed.

5. Polycloning Sites, a Universal Primer, and Biosafety

Using chemical mutagens other restriction sites were eliminated. The reason for this becomes clear by considering the elements of the enzymatic DNA sequencing technique (25). The *Eco*RI site in M13mp2 allowed cloning by screening for colorless plaques and prepare a template for sequencing the inserted DNA. Still, adjacent *lac* DNA needed to be purified as a primer, and initially a small restriction fragment was subcloned into a plasmid for primer purification (26). However, such a primer fragment needed to be denatured because it was double-stranded, and to be cut off after the sequencing reaction to produce a shift of the 3' end in the sequencing gel.

Although such a protocol could still be improved on, a more serious obstacle arose suddenly from the concern over the biological containment of M13 recombinant DNA. The NIH Recombinant Advisory Committee or RAC thought that the conjugation proficient *E. coli* host strains could lead to the spread of M13 infection and pose a risk in using M13 as a cloning vehicle. On the other hand, using one of the *traD* or *traI* mutants reduces conjugation by a factor of 10^6 , leaving the infection of M13 unabated. Because the F factor carrying the *traD* mutation was wild-type *lac*, a histochemical screen with the mp vectors would not be possible and one would have to return to drug-resistant-type M13 vectors. The scientific reasoning of RAC is hard to understand. First, nobody argued against

A

Polylinker of pUC4K:

Kan^r
 EcoRI-BamHI-SalI-PstI-SalI-BamHI-EcoRI
 AccI AccI
 HincII HincII

- 1) Partial digest of target DNA with *Mbo*I, *Hpa*II, and *Taq*I for instance.
- 2) Cut pUC4K with *Bam*HI or *Acc*I to completion.
- 3) Clone Kan^r into target DNA.
- 4) Cut target DNA with *Pst*I and select for Kan^s.
- 5) Insertions are either insertions of eight (*Bam*HI) or four (*Acc*I) amino acids.

Fig. 2. Symmetric (**A**) and asymmetric (**B**) (*opposite page*) polylinkers. Two other “tricks” were used in the construction of polylinkers. One type of polylinker was symmetrical, in which all sites except the central one occurred twice. By cloning a drug-resistance marker into the central site, the polylinker could be used in a linker scanning method of coding regions (7) and (35). The other type of polylinker was a pair, where two vectors contain an array of sites only once, but each of them in the opposite orientation. Cloned DNA no longer could be cloned out with a single enzyme as in the first type, but DNA could be cloned by using two different sites at the same time. This had the advantage that the orientation of a cloned fragment could be determined. By using a vector pair, both orientations can be obtained with the same pair of restriction cuts and therefore each strand of a restriction fragment could become the viral strand of M13 and available as a template for sequencing. Furthermore, by using two restriction enzymes that produce 3' and 5' overhangs, one can use it either for cloning oligonucleotide libraries or to generate unidirectional deletions with exonuclease III.

Agrobacterium tumefaciens as a plant transformation vector, although it was conjugation proficient and easily could spread in the environment. Second, F pili were never made under stress or anaerobic conditions, something that was already known as “phenocopies.” Conjugation in the human gut was in any case nearly zero. Third, M13 infection *per se* reduces conjugation by a factor of 10⁶. Leaving these arguments unanswered, a new series of *E. coli* strains (JM series) all carrying the M15 deletion on the F' *traD*36

B Polylinker of M13mp8 and M13mp9

EcoRI-SmaI-BamHI-SalI-PstI-HindIII (mp8)
 XmaI AccI
 HincII

HindIII-PstI-SalI-BamHI-SmaI-EcoRI (mp9)
 AccI XmaI
 HincII

5'----->3' GATCXXXXXXXXXXXXXXXXXXTGCA
 GAATTCCTCGGG <---- exoIII GCCAAGCTT
 CTTAAGGCCCTAG <-- exoVII----> ACGTCGGTTCGAA
 BamHI PstI

- 1) The small internal fragment is lost and recircularization (eliminating the background of recircularized vector) is only possible by providing the target DNA with the correct sticky ends.
- 2) It can only be cloned in one orientation, for the other orientation the other vector has to be used.
- 3) The notch can be used to clone oligonucleotide libraries with the first and the last four nucleotides in common.
- 4) Since exonuclease III requires double-stranded DNA and exonuclease VII single-stranded DNA, one can be used to attack the 3' recessed end of the *Bam*HI site, and the other both, the single-stranded ends of *Bam*HI and the *Pst*I sites.

Fig. 2B.

episome was constructed. Because it was a concern of NIH, a publication circulated to all potential users by NIH was used to describe the construction of the JM series and list the primer plasmids and the different M13 vectors, even already one with a *Hind*III site flanked by *Eco*RI sites that could produce still a blue plaque (27).

Although sequencing of eukaryotic DNA by M13 cloning was now possible, preparation of the primer from the plasmid was still cumbersome. It was clear that an oligonucleotide was needed to replace the restriction fragment as a universal primer. Although commercially custom-synthesized oligonucleotides were still very

expensive, the emergence of new chemical DNA synthesis methods such as the “triesther method” allowed the demonstration of a short oligonucleotide as a universal primer (28). However, oligonucleotide synthesis had other applications as well.

In 1978, another interesting observation was made, namely that inframe insertions of linkers in the *EcoRI* site could still give a positive color reaction. One of these isolates, M13mp5, already listed in the NIH Bulletin described earlier (27), could be used to clone both *EcoRI* and *HindIII* fragments at the same site and with the same primer for sequencing. The utility of creating cloning sites on top of each other was based on the universal primer concept, but in turn caused the development of multiple cloning sites (MCS) or poly-linkers that are found in all cloning vehicles today and provide many additional uses (Fig. 2). Therefore, work began on the synthesis of an oligonucleotide that could be inserted into the *EcoRI* site and generate restriction sites recognized by six base pair cutters such as *Bam*HI, *Acc*I, *Sma*I, or *Hind*III useful for cloning either blunt-ended fragments or fragments with sticky ends produced by four base cutters such as *Sau*3A, *Taq*I, and *Hpa*II (Fig. 3) (consistent with a DNA shotgun sequencing approach). This also required a renewed chemical mutagenesis to eliminate the *Acc*I and the *Hinc*II sites naturally occurring in M13. All single mutations were combined by marker rescue to give rise to M13mp7. Another important change in the enzymatic DNA sequencing protocol was required, facilitating the handling of many templates at the same time. Although the original sequencing method was carried out in sealed glass capillaries (25), reactions were now carried out in Eppendorf tubes (28).

6. Shotgun Sequencing in Practice

This was just the system, but did shotgun sequencing succeed? Initially, lack of funds and proper laboratory facilities made a demonstration impossible. Financial support finally arrived from the USDA, and the entire genome (8031 bp) of a plant virus, cauliflower mosaic virus (CaMV), was determined by DNA shotgun sequenc-

Unique vector sequence	Compatible target sites
G'GATC C	N'GATC N
C CTAG'G	N CTAG'N
<u>Bam</u> HI	<u>Sau</u> 3A, <u>Mbo</u> I
GT'CG AC	NT'CG AN
CA GC'TG	NA GC'TN
<u>Acc</u> I	<u>Taq</u> I
	NC'CG GN
	NG GC'CN
	<u>Hpa</u> II
GTC'GAC	NNN'NNN
CAG'CTG	NNN'NNN
<u>Hinc</u> II	Restriction enzymes like
	<u>Alu</u> I, <u>Hae</u> III, etc.;
	<u>Bal</u> 31 or <u>Exo</u> III/V;
	sheared and repaired DNA

Fig. 3. Sticky and blunt-end cloning of small fragments into unique cloning sites. By designing a unique sequence for the M13mp vectors that were recognized by restriction enzymes that could cut a hexanucleotide sequence in various ways by either producing sticky ends of four or two bases, or blunt ends, the variety of DNA fragments that could be cloned next to the universal primers were endless. Note that the sequence GTCGAC was recognized by *Sal*I, *Acc*I, and *Hinc*II, each producing different ends. In our sequencing project with cauliflower mosaic virus, we generated small DNA fragments for shotgun DNA sequencing by cleaving CAMV with *Eco*RI*, *Mbo*I, *Hpa*II, *Taq*I, *Hinc*II, *Hae*III, and *Alu*I (29). Later, we used DNase I (36), sonication (31), and a combination of exonuclease III and VII to generate blunt ends (37).

ing for less than \$3 a base. This was accomplished in a record time of 3 mo and finally published 1 yr later (29).

During the same time, work was progressing in sequencing human mitochondrial DNA twice the length of CaMV, but came to a grinding halt because the British guidelines emulated the NIH guidelines limiting the use of M13 to hosts with conjugation-reduced F factors. Because the M13mp2 vector and the universal

primer were already used in this project, both the JM strains and the newly developed M13mp7 for blunt-end cloning presented a major turning point for this work. Still, it seems hard to believe that a manuscript describing the concept of M13 shotgun DNA sequencing was rejected in the course of a review by PNAS as trivial at the time. However, it became clear that a few publications could not prove what many laboratories subsequently have proven. Therefore, it was critical not to have restricted the dissemination of the M13/pUC system (e.g., patent restrictions), but to have used the entire scientific community as a laboratory at large. This indeed became reality when this work became the most frequently cited work from 1981 to 1990 (30).

Becoming overwhelmed by requests for strains and protocols, the newly developing reagent companies were turned to for help. Their educational and service role immensely helped to disseminate the knowledge needed to train students and investigators in academia and industry in M13/pUC cloning, sequencing, and site-directed mutagenesis (31). Along the way, the first Apple-based software on shotgun sequencing (32,33) in addition to an undergraduate textbook (34). A good overview of M13mp, pUC vectors, and helper phage has also appeared (20).

Clearly, there are many new innovations that followed and made DNA sequencing what it is today. The substitution of X-ray films by lasers in conjunction with fluorescently labeled nucleotides allowed the automation of base calling (38). The throughput of templates per gel received improvement of fourfold, when all four chain terminators were synthesized with different dyes and could be separated in a single lane of a polyacrylamide gel (39). Subsequently, tracking software for these gels improved as well so that today reactions from 96 templates can be analyzed simultaneously (e.g., ABI 377, Applied Biosystems, Foster City, CA). Preparing gels and the running time of each gel would, however, limit the throughput of megabases of DNA sequences. The recent replacement of gels by capillaries yet provides another boost to the throughput of templates for sequencing. A major time-consuming step has been the loading of sequencing reactions on a gel. However, capillaries are loaded

with a robot arm from an array of reactions in 96-well microtiter plates. Because separation of DNA chains in capillaries is much faster than through gels, a station with four plates can keep the machine running in an uninterrupted fashion (e.g., ABI 3700, Applied Biosystems, Foster City, CA).

Calling bases from an autoradiogram has provided an important quality control that was lost when it was automated. Therefore, it was important to develop new computer software to control the accuracy of sequence information directly from the fluorograms produced by the DNA sequencing machines instead of their output of sequences (40,41). Without such a screen, the assembly of overlapping sequence reads becomes inefficient. Moreover, the faster throughput of sequencing reaction from shotgun libraries presented a major challenge to assemble the complete sequence from a set of overlapping fragments. The assembly of the 1.83 Mb *Haemophilus influenzae* Rd genome from more than 10^4 sequence reads has provided a good test case (12). Fragments of 1.6–2 kb were cloned into pUC18 instead of M13mp18. As shown previously, it permits the sequencing of an insert from both ends with universal primers (7). The advantage of having sequences from both ends aid in the computational assembly of fragments because their distances are known.

While all these improvements have accelerated the output of sequence information, one original feature has become even more prominent. The shotgun sequencing method sought to replace physical mapping by sequencing. Restriction maps, the hallmark of physical maps, could easily be reconstructed from the DNA sequence. By sequencing, restriction mapping became instant. For sequencing genomes larger than bacteria, however, bacterial artificial chromosomes (BACs) have now been used to sequence those genomes in increments. The entire human genome has been fragmented into fragments of 150–200 kb and cloned into BAC vectors (42). Each BAC is sequenced by the shotgun method and used to form contiguous sequences with overlapping BACs. Physical mapping methods can be used to position members of the BAC library on the intact genome, but this is very costly and a bottleneck to high-throughput sequencing. Instead, sequencing the ends of each BAC clone cre-

ates a database of sequence-tagged sites (STS) that can be used to develop a minimal tiling path (43). Once a BAC is sequenced, it is the sequence that is used to find the neighboring clone through its STS and not through the physical map. The ultimate application of shotgun DNA sequencing, however, has come recently as a proposal to drop the intermediate step of cloning the human DNA fragments into a BAC vector and to clone them as small fragments directly into sequencing templates (44). Although it remains to be seen if an assembly of sequence reads can be made with a 3 billion base pair genome, both approaches to sequence the human genome end up sequencing inserts with universal primers as we did with an 8-kb viral genome project 20 years ago (29).

Acknowledgments

Most of the initial work on M13 and pUC vectors was supported by the Deutsche Forschungs Gemeinschaft, the Department of Agriculture Grant No. 5901-9-0386, and the Department of Energy and AC02-81ER10901, respectively.

References

1. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
2. Scott, J. and Smith, G. (1990) Searching for peptide ligands with an epitope library. *Science* **249**, 386–390.
3. Edgell, M. H., Hutchison, C. A., III, and Sclair, M. (1972) Specific endonuclease R fragments of bacteriophage Φ X174 deoxyribonucleic acid. *J. Virol.* **9**, 574–582.
4. Beaucage, S. L. and Caruthers, M. H. (1981) Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862.
5. Hutchison, C. A., III, Phillips, S., Edgell, M. H., Gillam, S., Jahnke, P., and Smith, M. (1978) Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560.

6. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., and Hall, B. D. (1979) Sequence of the gene for iso-1-cytochrome C in *Saccharomyces cerevisiae*. *Cell* **16**, 753–761.
7. Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259–268.
8. Zoller, M. and Smith, M. (1982) Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res.* **10**, 6487–6500.
9. Norrander, J., Kempe, T., and Messing, J. (1983) Improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene* **26**, 101–106.
10. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H., and Fischer, H. (1973) Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage f1 DNA. *Proc. Natl. Acad. Sci. USA* **70**, 1209–1213.
11. Hofschneider, P. H. (1963) Untersuchungen über “kleine” *E. coli* K12 Bacteriophagen M12, M13, und M20. *Z. Naturforschg.* **18b**, 203–205.
12. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
13. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. V., Heynecker, H. L., Boyer, H. W., Crosa, J. W., and Falkow, S. (1977) Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**, 95–113.
14. Vovis, G. F. and Ohsumi, M. (1978) The filamentous phages as transducing particles, in *The Single-Stranded DNA Phages* (Denhardt, D. T., Dressler, D., and Ray, D. S., eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 445–448.
15. Herrmann, R., Neugebauer, K., Zentgraf, H., and Schaller, H. (1978) Transposition of a DNA sequence determining kanamycin resistance into the single-stranded genome of bacteriophage fd. *Mol. Gen. Genet.* **159**, 171.
16. Salivar, W. O., Henry, T. J., and Pratt, D. (1967) Purification and properties of diploid particles of coliphage M13. *Virology* **32**, 41–51.
17. Mandel, M. and Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 159–162.

18. Malamy, M. H., Fianndt, M., and Szybalski, W. (1972) Electron microscopy of polar insertions in the *lac* operon of *Escherichia coli*. *Mol. Gen. Genet.* **119**, 207–222.
19. Landy, A., Olchowski, E., and Ross, W. (1974) Isolation of a functional *lac* regulatory region. *Mol. Gen. Genet.* **133**, 273–281.
20. Messing, J. (1991) Cloning in M13 phage or how to use biology at its best. *Gene*, **100**, 3–12.
21. Messing, J., Gronenborn, B., Müller-Hill, B., and Hofschneider, P. H. (1977) Filamentous coliphage M13 as a cloning vehicle: insertion of a *Hind*II fragment of the *lac* regulatory region in the M13 replicative form in vitro. *Proc. Natl. Acad. Sci. USA* **74**, 3642–3646.
22. Messing, J. and Gronenborn, B. (1978) The filamentous phage M13 as carrier DNA for operon fusions in vitro, in *The Single-Stranded DNA Phages* (Denhardt, D. T., Dressler, D., and Ray, D. S., eds.) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 449–453.
23. Gronenborn, B. and Messing, J. (1978) Methylation of single-stranded DNA in vitro introduces new restrictions endonuclease cleavage sites. *Nature* **272**, 375–377.
24. Dotto, G. P. and Zinder, N. D. (1984) Reduction of the minimal sequence for initiation of DNA synthesis by qualitative and quantitative changes of an initiator protein. *Nature* **311**, 279–280.
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
26. Heidecker, G., Messing, J., and Gronenborn, B. (1980) A versatile primer for DNA sequencing in the M13mp2 cloning system. *Gene* **10**, 69–73.
27. Messing, J. (1979) A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. *Recombinant DNA Tech. Bull.* NIH Publication No. 79-99, 2, No. 2, 43–48.
28. Messing, J., Crea, R., and Seeburg, P. H. (1981) A system for shotgun DNA sequencing. *Nucleic Acids Res.* **9**, 309–321.
29. Gardner, R. C., Howarth, A. J., Hahn, P. O., Brown-Leudi, M., Shepherd, R. J., and Messing, J. (1981) The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucleic Acids Res.* **9**, 2871–2888.
30. Holden, C. (1991) Briefings. *Science* **254**, 28.
31. Messing, J. (1983) New M13 vectors for cloning. *Methods Enzymol.* **101**, 20–78.

32. Larson, R. and Messing, J. (1982) Apple II software for M13 shotgun DNA sequencing. *Nucleic Acids Res.* **10**, 39–49.
33. Larson, R. and Messing, J. (1983) Apple II computer software for DNA and protein sequence data. *DNA* **2**, 31–35.
34. Hackett, P. H., Fuchs, J. A., and Messing, J. (1984) An introduction to recombinant DNA techniques, in *Basic Experiments in Gene Manipulation*, Benjamin-Cummings, Menlo Park, CA.
35. Messing, J., Vieira, J., and Gardner, R. (1982) Codon insertion mutagenesis to study functional domains of β -lactamase, in *In Vitro Mutagenesis*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 52.
36. Messing, J. and Seeburg, P. H. (1981) A strategy for high speed DNA sequencing, in *Developmental Biology Using Purified Genes* (Brown, D. and Fox, F., eds.), ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. XXIII. Academic, NY, pp. 659–669.
37. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp and pUC vectors. *Gene* **33**, 103–119.
38. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679.
39. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336–341.
40. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998) Base-calling of automated sequencer traces using phred - I - accuracy assessment. *Genome Research* **8**, 175–185.
41. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**, 186–194.
42. Osoegawa, K., Woon, P.-Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**, 1–8.
43. Venter, J. C., Smith, H. O., and Hood, L. (1996) A new strategy for genome sequencing. *Nature* **381**, 364–366.
44. Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O., and Hunkapiller, M. (1998) Shotgun sequencing of the human genome. *Science* **280**, 1540–1542.



<http://www.springer.com/978-0-89603-716-8>

DNA Sequencing Protocols

Graham, C.A.; Hill, A.J.M. (Eds.)

2001, X, 244 p. 33 illus., Hardcover

ISBN: 978-0-89603-716-8

A product of Humana Press