A Bacteriophage T4 *in vitro* System to Clone Long DNA Molecules

Final report

6/1/90 – 1/31/96

Venigalla B. Rao
Department of Biology
The Catholic University of America
Washington, D.C. 20064

September 1997

RECEIVED

SEP 29 1997

OSTI

**PREPARED FOR THE U.S. DEPARTMENT OF ENERGY
UNDER GRANT NUMBER DE-FG05-90ER60996**

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *28*

MASTER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Three major objectives were proposed in the original proposal: i) Development of a bacteriophage T4 *in vitro* system, and necessary techniques to clone long segments of foreign DNA; ii) Development of a giant prohead DNA packaging system that could be potentially used to clone even a megabase size DNA; and iii) Development of techniques to rapidly map the cloned DNA inserts. A summary of the progress achieved with respect to each of these objectives is described below.

1. Development of the T4 system to package foreign DNA

When this project was initiated, we had already isolated two of the most important components of the bacteriophage T4 *in vitro* packaging system. These were, i) the packaging competent proheads, and ii) the soluble packaging proteins gp16 and gp17; the proheads were purified from the packaging deficient *17am18amrII* mutant extracts; the packaging genes *16* and *17* were cloned under the control of the strong λ P_L promoter, and the packaging proteins 16 and 17 were overexpressed and purified (the packaging proteins are also referred to as the terminase proteins, as they also generate the termini of the packaged DNA molecule). We therefore focussed our efforts on the development of techniques that are necessary to use this system to clone 150 kb size genomic DNA fragments. First, we developed techniques to isolate size fractionated 150 kb DNA fragments. The whole process of generating 150 kb size partial DNA fragments could be divided into three important steps: i) preparation of very high mol. wt. genomic DNA from cells, ii) generation of *Sau3AI* partial digests, and iii) size fractionation of the DNA partial digest to isolate 150 kb DNA for cloning. Working with the high mol. wt. DNA is a highly arduous and often frustrating process, due to the relative ease with which it can be sheared during manipulations. Considerable time was devoted to develop protocols and optimize conditions for generating the size fractionated DNA.

2. Cloning the size fractionated DNA

Vectors used for cloning: We used two vectors, AD10 and AD10SacBII for cloning the size fractionated DNA. Both these vectors were constructed, and were kindly provided by the late Dr. Nat Sternberg of Du Pont. Both the vectors have a unique *Bam*HI site for insertion of *Sau3AI* genomic DNA fragments. In the Ad-10 vector, the *Bam*HI site is located within the *tet*^R gene. Therefore, insertion into this site would result in the inactivation of the *tet* gene. This feature was used to select for the insert-containing recombinant clones; for selection of recombinant clones, the transductants were first plated on LB-kanamycin (LB-kan) plates, and *kan*^R transductants were tested for *tet*S by replica plating on a LB-tet plate. In the case of AD10SacBII vector, the *Bam*HI cloning site is within a *sacB* cassette. *SacB* gene was derived from *B. amyloliquefaciens*, and codes for an enzyme that converts sucrose to levan which accumulates in the periplasmic space of *E. coli*, leading to cell death. Thus, *E. coli* cells die in media containing sucrose, but grow normally in the absence of sucrose. Insertion into the *Bam*HI site results in the loss of expression of the *sacB* gene, and therefore, the recombinants grow normally in the presence of sucrose. This feature was used for direct positive selection of the recombinants by simply plating the transductants on a LB-kan-sucrose plate. For cloning the foreign DNA, the vector was linearized with *Bam*HI, and was then dephosphorylated with alkaline

phosphatase to prevent vector to vector ligation. The size fractionated DNA was then ligated to the vector DNA to form vector to insert concatemers. Prior to cloning experiments, both the vector and insert DNA were rigorously tested for ligation efficiency, and for the formation of appropriate ligation products.

Packaging and transduction: The ligated DNA (10 μ l) was directly added to the packaging reaction mixture (total volume:100 μ l), and the reactions were performed according to the published protocols. Basically, the packaging system consisted of purified proheads, partially purified packaging/terminase proteins gp16 and gp17, a freshly prepared head-minus, terminase-minus, *16am17am23amrII* mutant extract that provides neck proteins, tails, and tail fibers, that are required for the complete assembly of an infectious phage particle. The reaction mixtures also contained 5 mM ATP, 10% PEG and 100 mM NaCl for optimal packaging efficiencies. The reactions were stopped by the addition of DNAase (10 μ g/ml). The packaged DNA was then transduced into *E. coli* by adding an aliquot of the packaging mixture to a fresh *E. coli* culture expressing Cre, and incubating it at 37°C for 30-45 min. The entire culture was then spread either on a LB-kan plate when AD10 vector was used, or on a LB-kan-sucrose plate when AD10SacBII vector was used. The colonies were allowed to develop at 37°C for 16-20 hrs. In the case of Ad10 vector, *kan*^R colonies were replica-plated on LB-tet plates, and the insert-containing tet^S colonies were selected. In the case of AD10SacBII vector, the selection was directly done by spreading the culture on a LB-kan-sucrose plate; almost all the colonies that grew had inserts.

Analysis of *E. coli* and human clones: We first standardized the techniques described above using *E. coli* genomic DNA since it could be conveniently prepared in virtually unlimited quantities. Then, we used human genomic DNA for constructing libraries. To analyze the clones, individual colonies were inoculated into 10 ml LB-kan medium, and were grown till the culture became turbid. To this, isopropyl- β -D-thiogalactopyranoside (IPTG), the *lac* operon inducer, was added to 1 mM final concentration, and the cultures were grown overnight at 37°C. The recombinant plasmids are maintained normally as a unique single copy plasmid under the control of phage P1 plasmid replication system. Both the vectors also have an engineered P1 lytic replicon controlled by *lac* promoter. This replicon is normally repressed in the *lacI*^r background, but can be activated by the addition of IPTG to the cultures; this increases the copy number of the recombinant plasmid to about 20 copies per cell. Plasmid DNA was prepared from the IPTG-induced cultures by a modified alkaline lysis procedure. The DNA was digested with *Xho*I and *Bg*III, and the restriction fragments were resolved by 0.8% agarose gel electrophoresis. The vectors have the *Bam*HI cloning site either within a 7 kb fragment as in the case of AD10, or within a 10 kb fragment as in the case of AD10SacBII; insertion into this site would result in the disappearance of this band, while the rest of the three fragments having a size of 3.9, 1.8 or 0.7 kb should remain as common fragments in all the clones. Consistent with this, only this fragment disappeared in all the clones, and new fragments of varying length appeared in the positive colonies analyzed. The remaining three fragments of the P1 domain were seen as common fragments among all the. As would be expected, >90% of the *tet*^S or *suc*^R colonies had inserts. The sizes of the cloned inserts were estimated by adding the sizes of the individual restriction fragments in each lane calculated from their relative mobility values.

In some cases, since there were more than one fragment in the same position, the negatives of the photographs were scanned to determine the number of copies of restriction fragments in such bands. The calculated sizes of the inserts are most likely underestimates, since fragments less than 2 kb were difficult to visualize by ethidium bromide staining under the gel conditions used. The sizes of the inserts ranged from 60 kb to 130 kb.

The efficiencies of cloning in these experiments were on the order of 10^3 clones per μg of recombinant DNA. Among the clones analyzed, we observed variability in the size of the cloned insert and also a bias towards cloning smaller than full length inserts. Similar observations were also made with the phage P1 system using the cloning. This was mainly due to two reasons: i) this cloning strategy, which generates concatemeric DNA, for packaging, inherently favors cloning of smaller than full length; and ii) the size fractionated DNA used in these experiments was not of high quality; the DNA, as analyzed by FIGE, showed maximal intensity at about 60 kb, and a smear below and above up to about 150 kb. These cloning experiments were done, though we were aware of the above considerations, mainly to address the question of how well the T4 system can be used to clone foreign DNA. For construction of genomic libraries, we need to generate higher mol. wt. size-fractionated DNA.

3. Purification and characterization of giant proheads

A very attractive feature of the T4 system is that very large DNA clones, even as big as a megabase size, could be generated using certain mutants in the major capsid gene 23. These phages differ in the so called 'Q number' of the elongated icosahedral capsid of T4, and vary anywhere between 2 to 12 times the length of the wild type capsid. These mutants produce a mixture of smaller (petite), longer (giant), and normal length phage. Because of this feature, these mutants are referred to as *ptg* (petite giant) mutants. We crossed the packaging defective *17am18amrII* mutations into the giant mutant *ptg191C*. The resulting mutant, upon infection of the suppressor minus *E. coli* P301, produced empty giant proheads as visualized by electron microscopy (EM). The total proheads were purified by differential centrifugation, and EM of the purified proheads showed that the giant proheads constituted roughly 20% of the total proheads; rest of the prohead preparation was constituted by petites and wild type proheads. In order to test whether the giant proheads can package DNA *in vitro*, we developed a protocol to isolate pure giants that are essentially free of wild type and petite proheads. The mixture of proheads obtained after differential centrifugation were loaded onto a sucrose gradient (10-40% linear gradients) and were sedimented at 22,000 rpm for 3 hrs in a SW40.1 rotor at 4°C. Fractions of about 500 μl were collected and were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The data showed that the proheads were resolved into two peaks. As would be expected from the size of the proheads, EM of the sucrose gradient fractions showed that the bottom peak contained giant proheads of varying lengths, and the top peak contained wild type and petite proheads. The giant and wild type prohead fractions were pooled separately, and were directly loaded onto a DEAE-Sephacel column. After thoroughly washing the column, the proheads were eluted with 0.2 M NaCl, and the eluted fractions were analyzed by SDS-PAGE and EM. The data showed that giant proheads prepared by this protocol were very pure, and were essentially

free of smaller heads. This protocol is also highly reproducible, and EM of a number of batches showed that the giant proheads constituted about 90% of the total proheads.

The giant proheads were then tested for their competence to package DNA *in vitro*. We used two substrates, the purified phage T4 DNA, and the concatemerized AD10 DNA, to test the packaging ability of the giant proheads. The packaging activity data show two important results: i) the giant proheads could package both T4 and foreign DNA *in vitro*; and ii) the packaging efficiencies of the giant proheads, calculated as activity per prohead particle, are approximately the same as those of the wild type proheads. We have not yet tested these proheads for cloning several hundred kb size genomic DNA because of non-availability of the high. mol. wt. Size-fractionated DNA.

4. Characterization of a novel terminase activity involved in phage T4 DNA packaging

An important step in the dsDNA packaging pathway is the generation of termini by cleavage of the concatemeric DNA by a protein complex referred to as the terminase complex. In phage T4, the terminase complex is constituted by the 18 kd gp16 and the 69 kd gp17. The same complex attaches the cleaved DNA to the empty prohead and is believed to be directly involved in DNA translocation. Understanding the requirements for DNA cutting by T4 terminase is probably the most critical step in developing strategies for efficient cloning of foreign DNA. By developing an *in vivo* expression-based terminase assay, we have characterized the T4 terminase activity. These experiments demonstrated that the cleavage activity is associated with the large terminase subunit gp17, but not with the small terminase subunit gp16. Gp17, consistent with the circularly permuted ends of phage T4 DNA, cleaved DNA in a nonspecific manner generating random ends. Interestingly, the terminase cutting occurred preferentially on a DNA substrate that is transcriptionally active. This feature may be incorporated into the design of next generation cloning vectors.

5. Purification of large quantities of the T4 DNA packaging proteins gp16 and gp17

One of the major objectives in the later parts of this project was to construct recombinants that will overexpress gp16 and gp17. The current λ -P_L based constructs do not overexpress as abundantly as the more recent T7-based vectors, and in addition, require 42°C temperature shift for induction which may inactivate the proteins. We have constructed a series of recombinants using the phage T7 pET vector system which overexpress gp 16 and/or gp17 under various genetic backgrounds. We have also simplified the purification procedure using which we can now purify milligram quantities of the packaging proteins in 3-4 days. Using these highly concentrated proteins, we can now generate *in vitro* DNA packaging efficiencies that are 1-2 orders of magnitude greater than the λ P_L-overexpressed proteins. We are currently testing these proteins for cloning of large DNA fragments, and if the cloning efficiencies are also enhanced by 1-2 orders of magnitudes, it is possible to construct genomic libraries of complex genomes readily using small amounts of size-fractionated DNA.

6. Development of strategies to map the cloned inserts:

Two mapping strategies were developed that would be very useful for characterization of T4 clones as well as clones derived from the other large DNA cloning systems such as YACs and P1 libraries. These strategies are: i) an exonuclease III (ExoIII) mapping technique to generate restriction maps of the cloned insert; ii) a rapid polymerase chain reaction (PCR)-directed sequencing technique to determine the end sequences of the cloned insert.

The ExoIII mapping strategy involves linearization of the recombinant plasmid with the infrequent cutting *NotI* site, followed by treatment with Exo III, for successively hydrolyzing the nucleotides from the 3'-recessive ends. Under appropriate conditions of ExoIII digestion, a unique single strand of about 40 bases will be created at each of the 5'-ends. The DNA will be then annealed with a ³²P-labeled oligonucleotide that is complementary to the vector sequence adjacent to the *NotI* site. As a result, one of the single stranded ends will be specifically labeled with the oligonucleotide. The labeled DNA is then partially digested with a six base-cutting restriction enzyme, and the partial DNA fragments are resolved by FIGE. Of the numerous partial DNA bands that will be generated, only few will have the label at the common end. The agarose gel is then dried and exposed to an X-ray film. The restriction map of the cloned DNA can be readily constructed by simply reading the fragment sizes off the autoradiogram starting from the smallest fragment.

We have systematically worked out the conditions for mapping the cloned DNA using a variety of standard DNAs for which restriction maps are available, such as pUC18 (2 kb), M13-g20-RF (9 kb) λ -*XhoI* fragment (35 kb), and of DNAs for which restriction maps are not available, such as AD10 (28 kb), and pNad (110 kb) (this is a clone derived from a P1 library of *E. coli* genome). We have optimized this technique in such a way that the entire strategy up to the stage of gel electrophoresis could be completed within 3-4 hours. We have also generated complete restriction maps of pUC18, λ -*XhoI* fragment (35 kb), AD10 (28 kb) and a partial restriction map of pNAD (110 kb) using this strategy.

We also developed a rapid PCR-directed sequencing strategy that would be very useful for a variety of genome analysis experiments. The basic sequencing strategy consisted of two phases. In the first phase, the target DNA was amplified by symmetric PCR with low concentrations of dNTPs and primers. In the second phase, a small aliquot of the PCR mixture was amplified without any purification of the template DNA, by asymmetric PCR in the presence of a 5'-³²P-labeled primer and one of the four ddNTPs. This resulted in the accumulation of single stranded DNA products that are terminated at specific points by incorporation of the appropriate ddNMP. The products were then analyzed by electrophoresis on a sequencing gel followed by autoradiography. To demonstrate the utility of this strategy, we chose the phage T4 genome as a standard DNA and optimized the conditions for generation of sequence ladders of several hundred nucleotides starting from as low as 100 DNA molecules.

This sequencing technique will be very useful for a variety of genome analysis experiments that we plan to do in the very near future: i) it will be used for generating DNA

sequence of several hundred nucleotides from the ends of the insert; in order to accomplish this, two oligonucleotides that are complementary to the vector sequences flanking the *Bam*HI cloning site, will be used in the sequencing reactions; these primers would be extended into the unknown insert sequences, and terminated at specific points to generate sequences of the ends of the cloned DNA; the new stretches of DNA sequence generated from the ends would represent sequence tagged sites (STSs) that are roughly 150 kb apart; if the clone used was a giant clone, the STSs would flank several hundred kb; ii) it would be used to screen genomic libraries using the STS data obtained as above, to identify contiguous clones. Since the technique is highly sensitive, crude genomic libraries could be used for screening; this will be followed by screening different sets of genomic library pools until a pure clone is isolated; since this technique generates DNA sequence in addition to an appropriate size amplified fragment, false positives will be eliminated; iii) in addition, this strategy will have applications in other areas of genome analysis, for example, definitive prenatal diagnosis of certain genetic diseases directly from amniotic fluid.

7. Publications

Leffers, G. and Rao, V.B. A discontinuous headful packaging model for packaging less than headful length DNA molecules by bacteriophage T4. 1996. *J. Mol. Biol.* 258:839-850.

Rao, V.B. Strategies for direct sequencing of PCR-amplified DNA. 1994. *PCR Methods and Applications* 4:S15-S23.

Bhattacharyya, S. and Rao, V.B. Structural analysis of DNA cleaved in vivo by bacteriophage T4 terminase. 1994. *Gene* 146:67-72.

Rao, V.B. Direct sequencing of polymerase chain reaction-amplified DNA. 1994. *Anal. Biochem.* 216: 1-14.

Saunders, N., Zollinger, W. and Rao, V.B. A rapid and sensitive PCR strategy employed for amplification and sequencing of porA from a single colony forming unit of Neisseria meningitidis. 1993. *Gene* 137: 153-162.

Bhattacharyya, S. and Rao, V.B. A novel endonuclease activity associated with the DNA packaging protein gp 17 of bacteriophage T4. 1993. *Virology* 196: 34-44.

Rao, V.B. and Leffers, G. Purification and characterization of giant empty proheads from packaging-defective *23ptg* mutants of bacteriophage T4. 1993. *Virology* 196: 896-899.

Rao, V.B., Thaker, V. and Black, L.W. A phage T4 in vitro packaging system for cloning long DNA molecules. 1992. *Gene* 113: 25-33.

Rao, V.B. and Saunders, N. A rapid polymerase chain reaction-directed sequencing strategy using a thermostable DNA polymerase from Thermus flavus. 1992. *Gene* 113: 17-23.