DEVELOPMENT OF PROCEDURES

FOR DNA SEQUENCING OF THE ATTACHMENT SITE

OF LAMBDA BACTERIOPHAGE

BY

MARGARET A. HADDOW

A Thesis

submitted to

The Faculty of Graduate Studies and Research

University of Manitoba

In partial fulfillment

of the requirements for the degree

Master of Science

1978

DEVELOPMENT OF PROCEDURES FOR DNA SEQUENCING OF THE ATTACHMENT SITE OF LAMDA BACTERIOPHAGE

BY

MARGARET A. HADDOW

A dissertation submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

TO MY FAMILY

ABSTRACT

ABSTRACT

In order to sequence the attachment site of lambda virus two different protocols were tried. The first approach to sequencing λ att was to form heteroduplexes of λc_1 and λgal strains which would produce a single stranded loop encompassing the phage attachment region. This area would be subjected to digestion by S1 nuclease and the cleavage product isolated. The DNA fragment would then be reannealed to the appropriate $\lambda c_1 L$ strand to provide the necessary template for DNA polymerase I repair. The repair product would be analysed to establish the sequence. The second approach to sequencing the attachment site involved circularizing the heteroduplex product. The circular DNA was then treated with S1 nuclease, and the linear product labelled with polynucleotide kinase.

The growth of phage strains was essential to all work attempted. The λcIS7 strain was readily grown in lysogenic form and is harvested from the bacterial host after induction.

Growth and plating conditions for the phage were optimized. The plating efficiency was dependant on the host, the use of Difco yeast extract, the multiplicity of infection and the preadsorption of the phage to the host. Following these precautions yields of 10^{13}-10^{15} phage/ml could be obtained for λc_1 phage.

The low yield of λgal strains was not resolved by these procedures. Initial work was done on λgal which was used to infect a susceptible host. A second strain was also studied. This strain was used in a similar manner to λcIS7. The use of higher Mg^{++} ion concentration seemed to improve the yield somewhat. The use of pancreatic DNase and chloroform resolved problems with phage loss.

Despite these modifications to the harvesting procedure, the yield was only 10^{11} phage/ml.

To overcome the problem of growing λgal phage, experiments were developed to obtain a lysogen. Several procedures were tried which were unsuccessful. In the course of these experiments it was determined that the addition of 0.8% bile salts would differentiate between λc_1 and λgal.

The purification of S1 nuclease was studied. The assay procedure was dependant on the substrate size and structure. Experimental work with alkaline gradients demonstrated that a DEAE column purified the nuclease sufficiently without reducing the yield excessively.

The specificity of the enzyme was apparently associated with physiological conditions rather than purification conditions. Enzyme studies on specificity were related to the substrate used and the concentration involved. Using denatured DNA, the optimum substrate concentration was 160 μg/ml. In the presence of low concentrations of native DNA (less than 15 μg/ μl S1) some cleavage did occur. This cleavage could be overcome by the addition of small amounts of denatured carrier DNA.

The separated strands of DNA required for annealing were subject to degradation after a period of six weeks.

The isolated strands could be annealed to produce heteroduplexes. After analysis of the heteroduplexes on alkaline sucrose gradients, the pH used in incubating the DNA was determined to be critical.

The heteroduplex DNA and nuclease product had to be recovered from the gradients. It was found that the yield of DNA was low.

Pretreatment of the polyallomer tubes with a mixture of DNA-EDTA-PPi or the use of higher concentrations of sample alleviated the problem somewhat. The denaturation conditions prior to loading the gradients could also be modified to produce clearer peaks.

It was necessary to determine the precise cleavage site of the S1. Initial experiments with various templates indicated that the S1 produced butt ended cleavage.

The final aspect studied was the repair reaction as catalysed by DNA polymerase I. The isolated S1 cleaved fragments annealed to a $\lambda c_1 L$ strand were repaired and the repair product isolated. The 3' end was analysed and the preliminary indication was an adenylate residue. Further work was not done with this system.

Using the second approach to sequencing involved circularizing the heteroduplex product. The procedure to form circles was studied. In general the yield was 10 -30% of the sample treated. Using air to concentrate the DNA and a longer milder ligase treatment gave the best results.

Alkaline gradients were used to analyse the circular product. The peaks for the circles were very low and the profile indicated a mixed linear and circular population. Consequently agarose gels were used to isolate and analyse the circles. Ethidium bromide was used to visualize the DNA samples. Conditions were modified until the circular and linear DNA could be resolved.

Using disc gels on a preparative basis, the bands containing circular DNA were cut out and eluted. The recovery from the eluted sample was very low, about 20%. Further work was not done due to the poor yields.

ACKNOWLEDGEMENTS

The author is indebted to Dr. Peter Loewen, Associate Professor of the Department of Microbiology for his guidance and interest throughout the course of this investigation.

The author is thankful to Ms. M. L. Duckworth for valuable advice and Mrs. S. Kelly for typing of this thesis.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure | Page

LIST OF TABLES

LIST OF ILLUSTRATIONS

ABBREVIATIONS:

MOI	Multiplicity of infection i.e., No. of phage/bacteria.
λc_1	$\lambda c_1 857 Sam7$
λgal	$\lambda gal8c_1857$
λgal S7	$\lambda gal8c_1857Sam7$
DNA	Deoxyribonucleic acid
λ	Lambda virus
NaOH	Sodium hydroxide
$MgSO_4$	Magnesium sulphate
Mg^{++}	Magnesium divalent ion
$ZnSO_4$	Zinc sulphate
Zn^{++}	Zinc divalent ion
CsCl	Cesium chloride
HCl	Hydrochloric acid
tris	tris(hydroxymethyl)aminomethane
EDTA	Ethylene diamine tetra acetate disodium salt
SM	Diluent buffer
SM-Mg	Diluent buffer with high magnesium concentration
^oC	Degrees Centrigrade
poly UG	Copolymer of uracil and guanine
S1	S1 nuclease (EC 3.1.4.21)
kinase	polynucleotide kinase (EC 2.7.1.78)
pol 1	DNA dependant DNA polymerase 1
panc. DNase	Pancreatic DNase

ABBREVIATIONS:

LB	Luria Broth
gal	Galactose utilizing mutant
phage	Bacteriophage
McConkey	Selective and differential medium for specific sugars such as lactose or galactose
$CHCl_3$	Chloroform
CSH	Cold Spring Harbor

. . . "little did I know what lay afore me" . . .

abridged...from David Copperfield

ch. 40 by Charles Dickens

HISTORICAL

In approaching the problem of sequencing the λ attachment site, two areas should be discussed. First, a discussion of how lambda virus lysogenizes its host (in particular the role of the attachment site) and second, a discussion of the DNA sequencing methods available for study of DNA fragments.

A I Introduction

The bacteriophage λ has been the object of intensive study due to, among other reasons, its ability to lysogenize its host, Escherichia coli.

Lysogeny involves the decision not to lyse the host cell as determined by environmental influences [1,2] and the genetic incorporation of the virus genome into the host DNA. The phenomenon of lysogeny has been the subject of many excellent reviews.[3,4,5,6,7]

Campbell[2] proposed a model for lysogeny on which the linear phage DNA first formed a circle by joining ends. The incorporation of the phage into the bacteria involved insertion of the circular λ phage into the circular bacterial chromosome by crossing-over at a specific site. This reciprocal recombination event could account for the difference between vegetative and prophage maps, as well as the appearance of transducing phage.[2]

II Control of Recombination

Recombination in phage is under the control of various systems. Generalized recombination can occur throughout the phage genome. Signer and Weil[8] studied the phage recombination system called *red* through the use of point and deletion mutants. They determined that the *red* system was linked to vegetative phage reproduction and that it can be replaced by the bacterial recombination system *rec*. The majority of recombination seemed to be a product of the *red* rather than the *rec* system. Further work demonstrated that the *red* and *rec* systems were largely independent of each other.

Complementation studies and deletion mutants allowed the *red* system to be analyzed further. Two genes were mapped, *red* α and *red* β which produce respectively λ exonuclease and β protein.[10] The β protein interacts with λ exonuclease to increase its affinity for DNA. Recombination is envisioned to involve some sort of breakage and reunion model, not necessarily reciprocal.[11]

Some workers suggested the role of λ exonuclease in recombination may involve exposure of single stranded regions, which subsequently form a heteroduplex. This is followed by cleavage of redundant areas leaving a nick which is sealed by polynucleotide ligase.[12]

Another gene, γ, has also been implicated with the *red* system. Sakaki determined that the γ gene product is a protein which inhibits the host *recBC* DNase. The inhibition by γ is specific to the host *rec* system. Since the γ gene

lacks nuclease or ATPase activity, it is believed to have a regulatory role.[13] Other research indicated the γ protein facilitates the survival of irradiated lysogens. It is postulated that the protein inhibits only detrimental activity of the red β gene, thus allowing greater survival.[14]

III Integration and Excision as Site Specific Recombination

Site specific recombination involves the int system.[15,16] Mapping studies showed that integration as proposed by Campbell's model[2] occurred at a specific site, designated att.[15]

Gingery and Echols[17] studied integration deficient mutants and suggested an enzymatic role for the int gene, enhancing the structural recombination occurring at the att site. Their work was confirmed[16,18] and the results therefore support Campbell's model[2] where phage-directed site-specific recombination was postulated. Although the relationship between the generalized (red) and specific (int) recombination systems was unclear,[18] workers determined them to be independent.[15,16]

Although the nature of attachment was unclear, Echols, Gingery and Moore[15] suggested that if the int enzyme is involved in recognition of phage and bacterial DNA, there could be two sites on the protein for recognition of DNA.

In working with transducing phage, which have altered attachment sites, Guerrini[19] found that some integration and excision exchanges were not as efficient as others. He proposed that the attachment sites in bacteria were not identical to the phage attachment sites. As a result, he suggested

that int was a bifunctional protein allowing different recombinations depending on whether it was operating in an integration or excision mode. To account for the difference, the excision mode was considered to have a restricted set of exchanges possible during recombination.

The phenomena of integration and excision were complicated by the discovery of another gene, called xis.[20] Guarneros and Echols studied those phage which were unable to excise from their host after a brief heat induction. They determined that int and xis were two different genes. They complemented each other, while none of the xis mutants would complement each other. From deletion mutant studies, the xis site was mapped between int and exo (of the red system). The xis gene appeared to be functionally distinct from the int gene during the excision process. The authors suggested that xis was only involved in excision and that it could produce a protein which would restrict recombination.

In further studies, Echols[21] compared the recombination frequencies of λ with the transducing phages, λbio and λgal. He determined that the xis gene was required in addition to the int gene for excision, while only the int gene was necessary for integration to occur. The data also showed that xis was specific for those end normally present in the integrated state (Figure 1).

Both genes int and xis were found to be site specific.[21,22] This specificity of the int gene in recombination caused some workers to consider that int may be involved in recognizing the

site as well as facilitating the recombination[23].

Echols[21] suggested various modes of action for int and int complexed to xis. Free int product would favour integration of phage into the lysogenic state. Once integrated, the repressor would keep the levels of both gene products low. However, when induced, the formation of both int and xis would allow the formation of a complex and thus favour excision[23].

In order to explain the phenomenon of insertion and excision, Dove[24] proposed that the integrated state of the prophage was inherently more stable than the excised state. He suggested that the int gene product is a catalyst for both reactions, while the xis gene provides the necessary energy. The lysogenic state of λ is thus stabilized by a favourable energy situation. The xis gene product could provide energy by complexing with the int gene product[24].

The excision was determined to be reciprocal, regenerating the bacterial attachment sites. Also the excision process required a pair of attachment sites at the extremities to be effective[25].

The int protein has been purified[26,27] and the purity checked using SDS polyacrylamide gels. Ausubel[27] confirmed the identity of his extract by using different amber mutants in combination with permissive and nonpermissive hosts. He found the peak of the int protein would disappear under non-permissive conditions and reappear when a favourable host was used.

In experiments on the binding of the purified int protein, both Nash and Ausubel found only non-specific binding. Nash[26] suggested that the specific or essential factors were lost during the purification, or alternatively that the specific binding by int was masked by another protein found in the extract.

Further work[28] with an in vitro system and partially purified cell extracts defined some of the requirements for integrative recombination. It was found that spermidine, Mg^{++}ion and ATP were absolute requirements for the reaction.

Later studies[29] demonstrated that the int gene product did not regulate integration. The int gene product was determined to be a component of the recombination reaction in vivo and in vitro. Using a direct DNA binding assay, other workers determined that the int protein was involved in recognition of DNA sequences at the attachment site[30].

In studies on the excision process, Freifelder and Levine determined that lambda requires some transcription before excision can occur[52]. Gottesman and Gottesman[31] found no need for RNA or protein synthesis, using an in vitro system. There was an absolute requirement for ATP and a partial requirement for Mg^{++}ion and spermidine. Other studies by Kotewicz et al[30] suggest that the xis gene product may help mediate recognition of certain DNA sequences.

Thus the roles of int and xis in site-specific recombination are much closer to being understood.

IV Control of Lysogenic and Lytic Growth

The physical integration of the virus into the host combined with

its stable maintenance as prophage constitutes lysogeny. The decision to enter a lysogenic as opposed to a lytic cycle involves a choice and is under regulatory control. These controls have been the subject of intensive study.

When a cell is infected by λ, the phage genome circularizes and the cro and N genes are transcribed. This is known as immediate early transcription[1]. The N gene activates early gene transcription, which includes genes for replication (O and P), recombination (int) and repression (C_{II} and C_{III}). At this point the decision to enter the hytic or lysogenic cycle is made. If the lytic cycle is followed, transcription of the Q gene turns on late gene synthesis, including head and tail genes. If the lysogenic cycle is chosen, repression is established by means of the C_I gene product, known as the repressor. The achievement of repression and subsequent integration constitutes lysogeny[1] (Figure 2).

i. Effect of C_I

Steinberg and Ptashne[32], using highly purified repressor from the C_I gene and from binding studies, determined that it bound specifically to two operators, O_L and O_R, thus preventing the transcription of all other phage genes. Since different concentrations of repressor were required for binding, it was thought that the operators from the left and right differed in their affinity for repressor. The repression was shown to be a direct interaction between operator and repressor[32].

In further work with the λ repressor, Maniatis and Ptashne[33] isolated six different lengths of DNA fragments, after nuclease digestion, depending on the repressor:operator ratios. The experimental

procedure involved binding various amounts of repressor to the operator region of λ DNA, followed by nuclease digestion. The protected operator fragments were then isolated and studied. The authors concluded that the repressor in the form of either a dimer or tetramer binds preferentially to the operator region. Then additional monomer or dimer units are added sequentially in the leftward direction for O_L and rightward direction for O_R. The binding to the O_L was tighter than the binding to O_R, indicating some difference between the operators. In addition the DNA sequences of the O_R and O_L fragments have been shown to differ[33] (Figure 2b).

ii. Effect of C_{II} and C_{III}

a. Control of C_I

The repressor itself was also found to be under regulatory control[34]. Evidence from radioimmune assays showed that the repressor was produced at a linear rate after a short lag. Once the maximal level was reached, the amount of repressor was maintained but not increased. The different rates of repressor synthesis and the different binding led to the idea of the two pathways being involved in repressor control. The two promotors are called prm, promotor for repressor maintenance, and pre, promotor for repressor establishment.

Reichardt and Kaiser determined that pre is controlled by the C_{II} and C_{III} gene products, while prm is controlled by the repressor itself. The pre was mapped at the structural gene Cy[34]. However, studies with a λsar mutant which suppressed Cy mutants indicated another promotor, designated pro, was involved[35]. Since pro and pre

are functionally identical, it was not clear whether the pro promotor was simply a different map site or an alternate promotor[36]. The prm was mapped beside the righthand operator O_R where the repressor binds[34].

Echols and Green[37] confirmed the idea of diffusible products from C_{II} and C_{III} binding at pre, thus promoting the C_I repressor synthesis. It was found that the C_{II} gene product was stringently required, which led Reichardt to propose that C_{II} could bind polymerase or DNA. He suggested the function of C_{III} was to stabilize C_{II} or to modify a host protein[38].

The self-regulatory role of C_I, exerting positive control over its own synthesis through the prm site, in addition to negative control over other λ genes was also confirmed[39,40].

b. Control of int

The isolation of a mutant, int-c, which constitutively produced int gene product, led to the postulation of another promotor p_I[41]. Normal transcription begins from the leftward promotor p_L, generating int gene product. In fact some evidence was found that the p_I promotor was located within the xis gene[42].

Further studies indicated that C_I and int are regulated coordinately. It was proposed that the C_{II} and C_{III} gene products activated p_I as well as pre[43]. Other workers confirmed the positive regulation of int by C_{II} and C_{III}[30,44]. This control over int and not xis allows for a further differentiation between lysis and lysogeny[44,45].

Thus C_{II} and C_{III} gene products act bifunctionally as regulators with positive control of C_I and int negative control of late genes[45] (Figure 2c).