Cloning and Sequencing the Genes Encoding the Light Harvesting (LHII) Polypeptides of *Rhodopseudomonas acidophila* Strain 7050.

by

Ronald Christopher Mackenzie

A thesis submitted for the degree of Doctor of Philosophy.

University of Glasgow.
Department of Botany.
May 1990

DECLARATION

This thesis is an original composition which describes work performed entirely by myself unless otherwise cited or acknowledged. Its contents have not previously been submitted for any other degree. The research for this thesis was performed between October 1985 and June 1989.

Signed..

R. C. Mackenzie

Date: July 1st 1990.

ABBREVIATIONS

Unless otherwise stated S.I. units have been used in this thesis.

A_x	Absorbance at wavelength x
Å	Angstrom
Bchl	Bacteriochlorophyll
bp	Base Pair(s)
Bpheo	Bacteriopheophytin
BSA	Bovine Serum Albumin
CD	Circular Dichroism
DTT	Dithiothreitol
EDTA	ethylenediamine tetraacetic acid
IR	Infra Red
kb	Kilo Base Pairs
kD	Kilo Daltons
LHC	Light Harvesting Complexes
NIR	Near Infra Red
PEG	Polyethylene Glycol
pfu	plaque forming unit
ps	picoseconds
RC	Reaction Centre
RF	Replicative Form
SDS	Sodium Dodecyl Sulphate
TEMED	NNN'N'- tetra methylethylene diamine
UV	Ultra Violet
S.D	Shine-Delgarno (sequence)
Tris	Tris(hydroxymethyl) aminoethane

ABSTRACT

Rhodopseudomonas acidophila is a species of photosynthetic bacterium which has more than one type of light harvesting complex (LHII). The type and amount of LHII complex synthesised depends upon a variety of environmental factors. This bacterial species therefore provides an excellent model system for examining the response of gene expression to the environment.

The following strategy was used to clone the light harvesting structural genes, and flanking regulatory elements. A genomic DNA library of *R. acidophila* strain 7050 was constructed using the replacement vector lambda EMBL3. The recombinant DNA was packaged *in vitro*, then the library amplified on *E. coli* strain NM621. The library was screened by *in situ* plaque hybridisation using a radiolabelled fragment from plasmid pLHIISB18 as a heterologous probe. This fragment contained the genes encoding the B800-850 antenna polypeptides of *Rb. sphaeroides* strain 2.4.1 [Kiley and Kaplan, 1987]. Positively hybridising plaques were picked then streaked out to give single plaques. These plaques were then picked into an array and rescreened.

DNA from each clone was purified, digested with restriction enzymes, and the restriction pattern of each clone checked against the others. This resulted in the identification of 23 unique clones which could be placed in four groups. Each of groups 1, 2 and 3 contained an overlapping series of clones. On the basis of restriction pattern these groups appeared unrelated to each other. Each of the clones within group 4 appeared unrelated to to other members within group 4. The clones also appeared unrelated to clones within groups 1-3. These group 4 clones were not examined further.

A restriction map was made of each clone within groups 1-3. From these maps a restriction map of the genomic segment spanned by each group was predicted.

A representative clone from each of groups 1-3 (clones 9, 6 and 16) was selected and their DNA cleaved with a variety of restriction enzymes. Small, positively hybridising fragments were then subcloned into into M13 mp19. The mp19 derivatives were screened by *in situ* plaque hybridisation as described previously for the library. Single stranded templates were prepared from positively hybridising plaques then sequencing carried out by the dideoxy chain termination method.

Using a DNA alignment computer package, Bestfit, the sequences were compared to the sequence of the pLHIISB18 fragment. Sequences showing considerable alignment were then translated in six reading frames. The deduced amino acid sequences were compared to those of the *R. acidophila* LHII polypeptides, which had been determined previously by amino acid sequencing.

The data derived using these methods suggests that genes encoding nine antenna polypeptides have been successfully cloned and sequenced. Clone 16 encodes an alpha and beta polypeptide of the B800-850 type. Clone 6 encodes a beta polypeptide of the B800-820 type. However, evidence from Southern blot analysis suggests that an additional polypeptide may also be encoded within this clone. Clone 9 encodes three alpha and three beta polypeptides clustered within a region of ≃7kb. These are thought to form two complexes of the B800-850 type and a complex of the B800-820 type. However, the latter may form a novel complex class.

The genes encoding the polypeptides are ordered beta-alpha and are probably transcribed on a polycistronic mRNA. They all have the initiation codon ATG and either TAA or TGA as the termination codon. An intergenic, non coding region of 12bp separates the genes encoding the beta and alpha polypeptides. Upstream (5') of each coding region lies a possible ribosome binding site (Shine-Delgarno sequence).

All of the nascent alpha polypeptides appear to contain carboxy terminal extensions of 7-8 amino acids which are probably translated, but are not found in the native polypeptides. It is envisaged that the polypeptides are inserted into the membrane as suggested in the "membrane trigger" hypothesis. The carboxy-terminal extensions are thought to be involved in the insertion mechanism and are probably removed posttranslationally.

Acknowledgements

I would like to take this opportunity to thank the large number of people who have been involved either directly, or indirectly in the work which forms the basis of this thesis.

In particular I would like to thank Prof. Richard Cogdell (Botany Dept.) and Dr. Kim Kaiser (Genetics Dept.) who shared the unenviable task of being my supervisors, a role in which they have excelled. They both have infinite patience and needed it. As friends I could not have asked for better and I look forward to collaborating with both of them in the future.

It has been my good fortune during the course of this work to have shared the company of my fellow students. With their technical tips, snippets of scientific information, and relentless off-beat banter they have made my work much more than an education. I would particularily like to thank Alistair Gardner for providing me with a variety of strains of photosynthetic bacteria; Joyce Moore, Arvind Hundal and Craig Gibbs for assisting me in getting the sequencing systems up and running; Dianne Harbison for helping me through the tedium of loading the data into the computer and Steven Russell for then guiding me around the data handling systems. I would also like to thank Steven Goodwin and Colin for keeping me informed on the latest developments of PCR; Simon Tomlinson for imparting some of his encyclopaedic knowledge of subtractive hybridisation and Fiona Russell, Jane Mackenzie and Andrew Farquarson for their many discussions on science, art and politics.

Both Linda Ferguson and Peggy Ennis have provided excellent technical assistance whenever it was required. I would especially like to thank Norman Tait for preparing the photographs submitted in this thesis and for his assistance in the compilation of my conference posters.

One group whose assistance cannot be overstated is the "media ladies" on the 3rd floor of the Genetics building. They work in terrible conditions of extreme heat and unpleasent smells. They are always friendly and helpful and provide media and labware of excellent quality in vast quantities.

I would also like to thank Prof. Richard Cogdell the head of the Department of Botany and Prof. Dave Sherrat and Dr. Rodger Sutcliffe the sucessive heads of the Department of Genetics for providing me with excellent facilities in which to carry out research.

Thanks also go to Samuel Kaplan and Patricia Kiley for providing me with plasmid pLHIISB18 and Rene Brunisholz for providing the R. acidophila strain 7050 polypeptide amino acid sequences before publication.

I am indebited to my mother, father and sister for their unflinching support during both my undergraduate and postgraduate periods of study, a debt which I could never begin to repay.

I would also like to thank my girlfriend Rosie, but words cannot express my gratitude, therefore to her I would like to dedicate this thesis.

Finally I would like to thank The University of Glasgow for providing me with a Resarch Studentiship which enabled me to carry out this work.

CONTENTS

CHAPTER 1 AN INTRODUCTION TO THE PHOTOSYNTHETIC BACTERIA

1.1 The Taxonomy of Purple Photosynthetic Bacteria:

It seems probable that "modern" photosynthetic bacteria represent surviving lines of ancient photosynthetic groups. These groups predated the emergence of oxygen evolving photosynthetic organisms and at one time were the only forms capable of utilising radiant energy on the primitive oxygen-free Earth [Stanier, 1970].

Most of the "modern" phototrophic bacteria are a physiological group of Gram-negative aquatic bacteria [Stanier, 1970]. They are characterised by their ability to grow in the absence of O_2 in the presence of light. Unlike the cyanobacteria, algae, and higher plants, the photosynthetic bacteria carry out anoxygenic photosyntheseis, that is, they do not generate molecular oxygen by photolysis [van Niel, 1944]. Instead these organisms utilise simple organic compounds, molecular hydrogen or sulphur compounds as electron donors. Despite these differences van Neil proposed a unifying equation for photosynthesis [van Niel, 1931, 1941].

$$CO_2 + 2H_2A \xrightarrow{\text{light + chlorophyll}} CH_2O + H_2O + 2A$$

Where H_2A is the electron donor, CH_2O is a fixed carbon storage product, and 2A the reduced product; ie. molecular oxygen in oxygenic systems, or sulphur, sulphate, protons, organic compounds and carbon dioxide in photosynthetic bacteria [Pfennig, 1978]. Such anoxygenic photosynthesis is fairly closely correlated with an anaerobic mode of life. The majority of these prokaryotes are strict anaerobes. These physiological properties thus effectively confine them to oxygen-free aquatic habitats that receive sufficient light to sustain photosynthesis [Pfennig, 1967]. Such environments include sewage ponds, grossly polluted pools and the hypolimnion of lakes [Pfennig, 1978].

All anoxygenic, bacteriochlorophyll containing, phototrophic bacteria belong to the order Rhodospirillales [Pfennig and Truper, 1971, 1974; Truper 1976]. This order has been divided into two suborders: Chlorobiineae (green bacteria) and Rhodospirillineae (purple bacteria). The division was made on the basis of differences in ultra-structure and pigment composition of the photosynthetic apparatus [Pfennig and Truper, 1971; Gibbons and Murray, 1978; Truper and Pfennig, 1978]. The original designation of "green" and "purple" bacteria was made on the basis of the colour of the growing cultures. Although examples exist in both groups which do not conform to this convention, these original group designations have retained their validity for most species with respect to the morphological organisation of the photosynthetic apparatus [Sprague and Varga, 1986].

In the green bacteria (Chlorobiineae) there are two cytologically distinct structures which form the photosynthetic apparatus, the chlorosome and the cytoplasmic membrane [Truper, 1976; Truper and Pfennig, 1978; Staehelin *et al.*, 1980]. The chlorosome is a lens shaped, non-unit membranous structure which contains the light harvesting pigments, bacteriochlorophylls (BChl) c, d, or e [Pfennig and Truper, 1973]. The reaction centres, where the primary photochemical event occurs, are housed in the cytoplasmic membrane. The chlorosomes sit above the reaction centres appressed onto the cytoplasmic membrane. However, although the chlorosomes are closely associated with the cytoplasmic membrane they are not derived from it [Sprague *et al.*, 1981; Sprague and Varga, 1986]. This suborder contains the families Chlorobiaceae and Chloroflexaceae [Truper and Pfennig, 1978].

In the purple bacteria (Rhodospirillineae) the photochemical reaction centres and light harvesting antennae are contained in a specialised system of membranes. These membranes, termed the intracytoplasmic or photosynthetic membranes, are derived from, and are continuous with, the cytoplasmic membrane [Pfennig and Truper, 1973; Pfennig, 1967]. The photopigments are BChl a or b [Cohen-

Bazire and Sistrom, 1966]. This suborder contains two families Chromatiaceae and Rhodospirillaceae. The division of the suborder was made on the ability of its members to utilise elemental sulphur as an electron donor for photosynthesis [[Pfennig and Truper, 1973; 1981].

Chromatiaceae or purple sulphur bacteria: Most members of this family are strictly anaerobic and therefore obligate phototrophs. All species are capable of photolithotrophic carbon dioxide fixation in the presence of sulphide. Under these conditions elemental sulphur is accumulated in the form of globules inside or outside the cells. In nature the Chromatiaceae occur in the anaerobic, sulphide containing parts of aquatic environments from moist and muddy soils to ditches, ponds, lakes, rivers, sulphur springs, salt lakes, estuaries (especially salt marshes) and other marine habitats, but not in the open ocean [Pfennig and Truper, 1973; 1981].

Rhodospirillaceae or purple non sulphur bacteria: This family is characterised by its inability to utilise elemental sulphur and sulphide as photosynthetic electron donors. This role is fulfilled by simple organic compounds such as acetate, pyruvate and succinate. In general all species are microaerophilic: [van Neil, 1944, 1957; Pfennig and Truper, 1974]. Many representatives may grow at full oxygen tension in the light or dark. In strains able to grow under microaerophilic to aerobic conditions the photopigment and intracytoplasmic membrane system decreases as the concentration of dissolved oxygen increases. Under strictly anaerobic conditions in the dark, growth and reproduction is very slow as fermentation is the energy providing process. Ecologically the ability to grow in the dark is believed to be of little consequence. These organisms are not found thriving under natural conditions unless light is periodically available to them [Keister, 1978]. In nature the Rhodospirillaceae occur in all aquatic habitats with the exception of the open ocean [Pfennig and Truper, 1973].

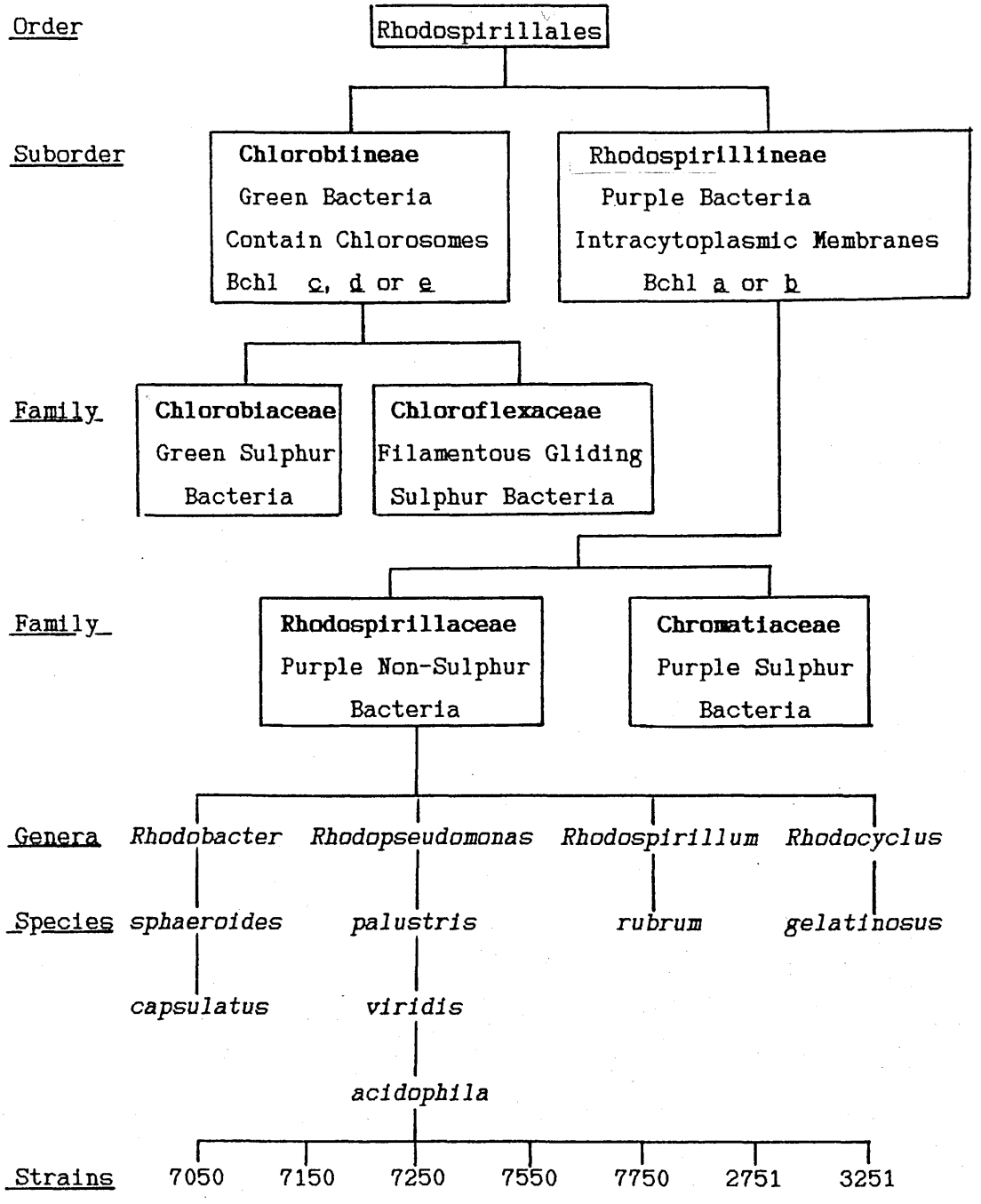


Figure 1

Summarised family tree of the photosynthetic bacteria as described in the text. Note strains of *R. acidophila* are listed as first described by Pfennig [Pfennig, 1969]. Since then additional strains such as 10050 have been isolated from the wild.

The family has been further divided into four genera; *Rhodobacter* [Imhoff *et al.*, 1984], *Rhodopseudomonas* [Molisch, 1907], *Rhodospirillum* [Molisch 1907] and *Rhodocyclus* [Pfennig, 1978]; differentiation has been made on the basis of differences in their morphology and physiology [Imhoff *et al.*, 1984] see figure 1.

Rhodopseudomonas acidophila was isolated by enrichment of bacteria in mud and water samples using a low pH, succinate-mineral salt medium. Seven strains were identified including strain 7050 isolated from water taken from Crystal lake, Urbana, Illinois and strain 7750 from a farm pond near Athens, Georgia [Pfennig, 1969].

The cells of this species are ovoid or slightly curved rods, 1-3μm wide and 2-4μm long which grow optimally at 30°C and reproduce by budding. Motility is by means of polar flagella, however the cells do not show phototactic responses.

Grown aerobically the cells are colourless to pink [Pfennig, 1969]. The colour of anaerobic cultures is dependent on the light intensity and the temperature at which the cells are grown. At 30°C and under a high light regime strains 7750 and 7050 are orange-brown in colour. At this temperature but under low light conditions strain 7750 is orange-brown whereas strain 7050 is purple-red [Pfennig, 1969]. However, if cells of strain 7750 are grown at 25°C under low light conditions they are also purple-red in colour and spectrally indistinguishable from low light grown cells of strain 7050 [R. J. Cogdell, personal communication]. These colour changes reflect the pigment protein-composition of the photosynthetic membrane and it is to this area that attention will now be turned.

PURPLE BACTERIAL PHOTOSYNTHESIS

In all known photsynthetic organisms the light reactions occur within membranes [Sauer, 1978; Trebst, 1974]. The membranes must fulfill two criteria, they must be closed, and they must be proton impermeable. These requirements are a prerequisite to the establishment of a transmembrane proton gradient and ultimately ATP synthesis [Mitchell, 1961; Capaldi, 1979; Hinkle and McCarthy, 1978; Racker and Stoeckenius, 1974; Racker, 1976; Miller, 1979; Heber and Walker, 1979]. In purple bacteria the creation of such a gradient involves the participation of photosynthetic units within the membrane.

Detergents have played a major role in investigating the components of the photosynthetic unit. Their use has permitted the solubilization of the intracytoplasmic membranes and the resolution of the components of the photosynthetic unit by classical protein purification techniques [Reed and Clayton, 1968; Thornber, 1970; Cogdell *et al.*, 1983; Cogdell and Thornber, 1979]. Such units consist of two structurally and functionally distinct pigment protein complexes; the photochemically active reaction centre (RC) and the light harvesting or antenna complexes (LHC) [Thornber *et al.*, 1978; Cogdell, 1983; Cogdell, 1986]. Light absorbed by the antenna complexes is passed by radiationless transfer to the reaction centre where the primary photochemical event and charge separation occurs [Clayton, 1978; Parson and Ke, 1982].

1.2 The Light Harvesting Antenna Complexes

The role of the antenna complexes is to absorb available light and transfer the energy to the reaction centre. Purple bacteria usually develop in aquatic environments. However, with depth and turbidity, the light available for photosynthesis decreases rapidly both in intensity and spectral quality [Kirk, 1983]. By absorbing light the antenna complexes effectively enlarge the cross sectional

capture area of the reaction centre thereby increasing the number of photons available for photochemistry [Sauer, 1978].

Spectral Taxonomy

Most bacterial antenna complexes can be placed in one of two distinct classes, see table 1.1. Historically these two classes have been called LHI and LHII. However, although still in use the lack of information in these titles has resulted in this nomenclature being superceded by (B)Bulk-890 or B800-850 protein classes respectively. The numbers in these titles refer to the appropriate bacteriochlorophyll near infra-red (NIR) absorption maxima [Cogdell and Thornber; 1979, Cogdell and Thornber, 1980; Thornber *et al.*, 1978, Thornber *et al.*, 1984].

So far all species of wild type purple bacteria studied contain an LHI class of antenna complex. This type of complex eg. the B890 from *Rhodospirillum rubrum* or the B875 complex from *Rhodobacter sphaeroides* surrounds the reaction centre and occurs in fixed stoichiometry to it. These complexes are characterised by having a single strong BChl absorption maximum in the NIR. In some strains of bacteria eg. *Rsp. rubrum* strain S1 and carotenoidless mutant strains of *Rb. sphaeroides* and *Rb. capsulatus* this is the only type of complex found [Sistrom, 1978; Thornber, 1986; Sauer and Austin, 1978; Cogdell, 1986; Aagard and Sistrom, 1972; Cogdell *et al.*, 1982; Picorel *et al.*, 1983; Broglie *et al.*, 1980].

Many species, in addition to the LHI complex, contain a second type of complex the LHII or B800-850 protein class. Within this class are three light harvesting types [Cogdell and Thornber, 1979; Cogdell and Thornber, 1980; Thornber *et al.*, 1978, 1984];

Type I B800-850: In most species this is the only type of LHII complex found. Spectrally it is characterised by having two strong Bchl absorption maxima at 800 and 850nm, the 850nm absorption band being ≈1.5 times as intense as that at 800nm. This type of complex

Antenna type	B-890 complex class		B-800–850 complex class		
	B-890-complex	B-875-complex	Type I	Type II	B-800–820-complex
Examples of bacteria containing antenna type	*R. rubrum* *Rps. viridis* *Chr. vinosum* *Rps. acidophila*	*Rps. sphaeroides* *Rps. capsulata* *Rps. palustris*	*Rps. sphaeroides* *Rps. capsulata* *Rps. gelatinosa* *Rps. acidophila* 7750	*Chr. vinosum* *Rps. acidophila* 7050	*Chr. vinosum* *Rps. acidophila* 7050
Bchla: carotenoid ratio	2:1	2:2	3:1	3:1	3:1
Nos. of polypeptide types in isolated complex	2 or 3	2	2 or 3	2	2
Intensity of CD spectrum of long wavelength band	Strong	Weak	Strong	Strong	Strong
Typical values of the millimolar extinction coefficient of the long wavelength Bchla absorption band[b]	*R. rubrum* B-890-complex 125 cm^{-1}	*Rps. sphaeroides* B-875-complex 126 cm^{-1}	*Rps. sphaeroides* B-800–850-complex 132 cm^{-1}	Not determined	Not determined

[a] The quantitative data are based on the analysis of only a few species and are only given as typical examples.
[b] The values are probably good to ± 10 cm^{-1}

Table 1.1

Table showing the classification of the light harvesting antenna complexes. Taken from Cogdell, 1986.

is found in *Rb. sphaeroides*, *Rb. capsulatus* and *Rhodocyclus gelatinosus*. It is also found in *Rhodopseudomonas acidophila* strain 7050 when grown at high light intensities and strain 7750 grown at high temperature (30°C) [Cogdell *et al.*, 1983; Broglie *et al.*, 1980; Clayton and Clayton, 1972; Cogdell and Crofts, 1978; Feick and Drews, 1978; Sauer and Austin, 1978; Thornber *et al.*, 1984].

Type II B800-850: This type of complex is found in *Chromatium vinosum*, *R. acidophila* strain 7050 grown at low light intensities and strain 7750 grown at low temperatures (25°C). It is characterised by the 800 and 850nm absorption maxima being of equal intensities. Unlike the type I complex the position of the "850" absorption band can vary between 835 and 855nm. The degree of variability depends on such factors as detergent concentration, ionic strength and pH. This suggests that the absorption spectrum may be an artefact generated by the methods used for isolation, or as a result of incomplete separation of B800-850 and B800-820 complexes [Thornber, 1970; Cogdell *et al.*, 1983; Thornber *et al.*, 1978; Cogdell and Thornber, 1979; Angerhofer *et al.*, 1986; Schmidt *et al.*, 1987].

B800-820: It has been observed that whenever a bacterium contains type II B800-850 complexes that it also contains B800-820 complexes. The status of this type of complex remains to be clarified. However, it does appear to be closely related to the type II B800-850 complexes. In *C. vinosum* the B800-820 complex and type II B800-850 complexes have similar sizes and pigment content, while in *R. acidophila* strains 7050 and 7750 they have very similar apoproteins [Thornber *et al.*, 1978; Thornber, 1970; Cogdell *et al.*, 1983; Angerhofer *et al.*, 1986; Schmidt *et al.*, 1987].

The Biochemistry of the Antenna Complexes

Antenna complexes consist of a supermolecular aggregate of "monomers" termed minimal units [Cogdell and Crofts, 1978; Feick

and Drews, 1978; Sauer and Austin, 1978; Shiozawa *et al.*, 1982]. The minimal units are composed of two components, light absorbing pigments and low molecular weight, amphipathic polypeptides [See reviews Amesz, 1985; Zuber, 1985; Cogdell 1986; Drews, 1985; Thornber *et al.*, 1983].

The Pigments: Two species of pigment molecules are involved in light harvesting, bacteriochlorophylls (BChls) and carotenoids (Crts) [Cohen-Bazire and Sistrom, 1966; Stanier *et al.*, 1976; Borisov, 1978]. *In vivo*, bacteriochlorophylls absorb light strongly in the near infrared region of the spectrum, well outwith the wavelengths utilised for oxygenic photosynthesis [Stanier *et al.*, 1981]. The majority of purple photosynthetic bacteria have Bchl a as their major light absorbing pigment. However, in a few species such as *Rhodopseudomonas viridis* and *Ectothiorhodospira halochloris* this role is fulfilled by bacteriochlorophyll b [Eimhejellen *et al.*, 1963; Thornber *et al.*, 1978]. In the case of *R. viridis* this generates a major absorption band at 1,020nm [Stanier *et al.*, 1981].

It should be noted that the spectral properties of purified Bchl *in vitro* differ considerably from those found *in vivo*. For example, purified Bchl a in ether has a long wavelength absorption maximum of 775nm, whereas in whole cells this is red shifted to between 800 and 910nm. This is largely, though not entirely a result of the non-covalent association of Bchl with the light harvesting polypeptides [Stanier *et al.*, 1976; Clayton, 1963, 1966; Thornber, 1978]. This spectral shift is used as the basis of the assay which determines the integrity of isolated light harvesting pigment-protein complexes [Cogdell, 1986; Stanier *et al.*, 1981].

The other group of pigments, the carotenoids, are bimodal in function. They act as accessory light harvesting pigments, absorbing light in the 450-570nm range [Cohen-Bazire and Sistrom, 1966; Stanier *et al.*, 1976], and they perform a photoprotective role by preventing the photodynamic effect [Cogdell, 1985]. That is, photosynthesis generates exited triplet states of

bacteriochlorophyll, such molecules may transfer their energy to oxygen forming a powerful oxidant, singlet oxygen. Carotenoids interact with both triplet bacteriochlorophyll and singlet oxygen dissipating the energy as heat.

The Polypeptides: The role of the antenna polypeptides is to anchor the light harvesting pigments in the correct orientation within the membrane [Breton and Vermeglio, 1982; Kell, 1984]. The protein composition of a number of antenna complexes has been investigated. In most cases they are composed of two polypeptide types, alpha and beta, in a 1:1 molar ratio [Brunisholz *et al.*, 1981; Fuglistaller *et al.*, 1984; Tadros *et al.*, 1984, 1985; Theiler *et al.*, 1984a and b, Cogdell *et al.*, 1983; Bissig *et al.*, 1988; Cogdell and Thornber, 1980]. The exceptions, the B1020 complex of *R. viridis*, the analogous complex in *Ectothiorhodospira halochloris*, and the type I B800-850 complex from *Rb. capsulatus*, have an additional gamma polypeptide which does not bind pigment [Bolt *et al.*, 1981; Feick and Drews, 1978; Brunisholz *et al.*, 1985]. In these cases the polypeptide ratio is 1:1:1 [Shiozawa *et al.*, 1982].

The role of non-pigmented polypeptides is uncertain. It has been suggested, that in the Bchl b containing species, they act as linker polypeptides between adjacent alpha-beta subunit pairs. This may account for the ordered hexameric arrays seen in electron micrographs of the complexes of these species [Englehardt *et al.*, 1983]. In *R. capsulatus* the gamma polypeptide has been implicated in the assembly and insertion of the B800-850 complex into the membrane [Kiley and Kaplan, 1988; Drews and Feick, 1978].

The ratio of alpha to beta polypeptides in LHII complexes is not always 1:1 [Bissig *et al.*, 1989a,b]. The B800-820 complex of low temperature grown *R. acidophila* strain 7750, for example, contains two beta type polypeptides. However, at higher temperatures a single beta polypeptide is found [Brunisholz *et al.*, 1987].

In *R. palustris* strain French high light and low light B800-850 complexes are found giving in total four alpha (α_1-α_4) and three beta polypeptides. It is still unclear whether all of these polypeptides are present in both high and low light B800-850 complexes. It may be that high and low light complexes contain different polypeptides, or they may contain the same polypeptides in a different stoichiometry. Preliminary investigations suggest that beta polypeptide exchange occurs between the two complexes and that the α_2-polypeptide is present predominantly in the low light grown cells. This gives approximate alpha polypeptide ratios (α_1:α_2:α_3:α_4) of 2:2:2:1 and 2:6:2:1 in high and low light grown cells respectively [Evans, 1989].

The basis for the distinction between alpha and beta polypeptides lies in their primary amino acid sequences. Within the polypeptides a number of amino acid are conserved, in particular the sequence AxxxHis. The histidine residue within this sequence is believed to non-covalently bind bacteriochlorophyll [Theiler and Zuber, 1984; Youvan and Ismail, 1985; Tadros *et al.*, 1987c]. In beta polypeptides two of these sequences occur whereas in alpha polypeptides there is only one [Zuber *et al.*, 1983; Zuber, 1986]. Also noticeable within the primary sequence is the conservation of aromatic residues positioned at a definite distance from the putative Bchl binding sites [Brunisholz *et al.*, 1984a].

Both alpha and beta polypeptides are small being composed of ≈50 amino acids, and extremely hydrophobic often being soluble in a mixture of chloroform and methanol [Zuber, 1986; Cogdell and Thornber, 1980; Thornber *et al.*, 1983]. A hydropathy plot shows a tripartite character which can be illustrated by the B890-α-apoprotein from *Rsp. rubrum* [Zuber *et al.*, 1983, Zuber, 1986; Theiler *et al.*, 1982; Brunisholz *et al.*, 1985], see figure 1.2. In this case the apoprotein has a central core of 23 hydrophobic amino acids while at either end the sequence is polar. This suggests that the apoprotein lies across the intracytoplasmic membrane, spanning it once, with the central hydrophobic core buried in the

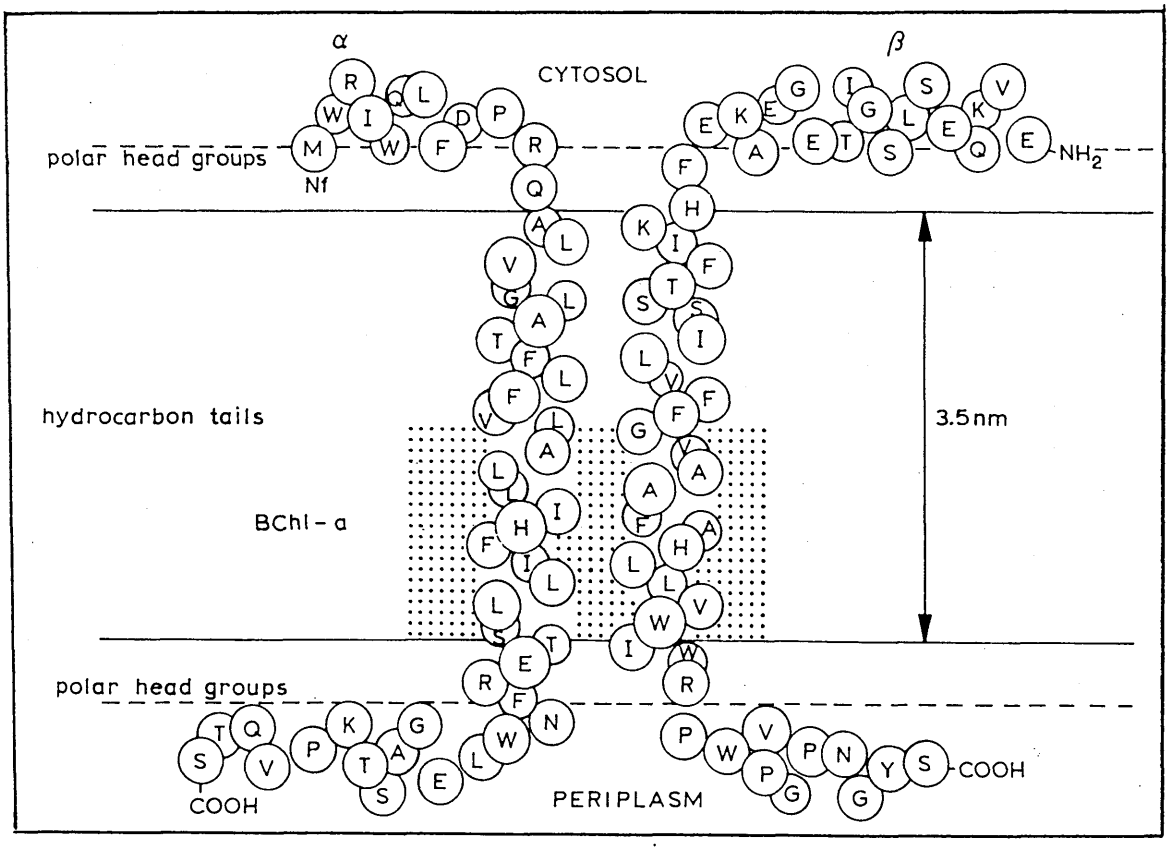


Figure 1.2

A model for the organisation of the alpha and beta antenna apoproteins in the photosynthetic membranes of *Rsp. rubrum*. The polypeptides have a tripartite character with polar N- and C-terminii which lie on the cytoplasmic and periplasmic sides of the membrane respectively. The central hydrophobic core of ≈20 amino acids is believed to form an α-helix, spanning the membrane once. The amino acids are represented by the single letter code. Taken from Brunisholz *et al.*, 1985.

non-polar interior of the membrane. The three-dimensional structure of the core is believed to be largely alpha-helical [Brunisholz *et al.*, 1986, Brunisholz, 1984b]. Evidence for this has come from examination of the native complexes by a combination of IR and Far UV Circular Dichroism spectroscopy [Theiler *et al.*, 1984b; Breton and Nabedryk, 1984]. These methods suggest an alpha helical content of 50-60% [Francis and Richards, 1980; Cogdell and Sheer, 1985].

It is possible by using *Rsp. rubrum* to obtain closed membrane vesicles that have both right side-out (sphaeroplasts) and inside-out (chromatophores) polarity. This has allowed both membrane faces to be interrogated for the presence of their antenna apoproteins. Using mild proteolysis and immunoprecipitation, the antenna apoproteins have been clearly detected at both membrane surfaces. It appears they lie across the photosynthetic membrane so that their N-terminii are on the cytoplasmic side and their C-terminii are on the periplasmic side [Brunisholz *et al.*, 1986; Brunisholz, 1984a and b].

In vivo Structure

All isolated light harvesting complexes have an apparent molecular weight between 80,000 and 180,000. Isolated complexes with sizes smaller than 80,000 tend to have poorly defined absorption spectra suggesting that *in vivo* the complexes consist of supermolecular aggregates of minimal units [Broglie *et al.*, 1980; Peters and Drews, 1983; Peters *et al.*, 1984].

It has not been possible to dissociate the macromolecular structure into basic subunits containing only one pair of the two different Bchl binding polypeptides [Broglie *et al.*, 1980, Peters and Drews 1983; Peters *et al.*, 1984; Picorel *et al.*, 1983; van Grondelle *et al.*, 1983]. However, a model for the minimal unit of the type I B800-850 complex from *Rb. sphaeroides* has been proposed on the basis

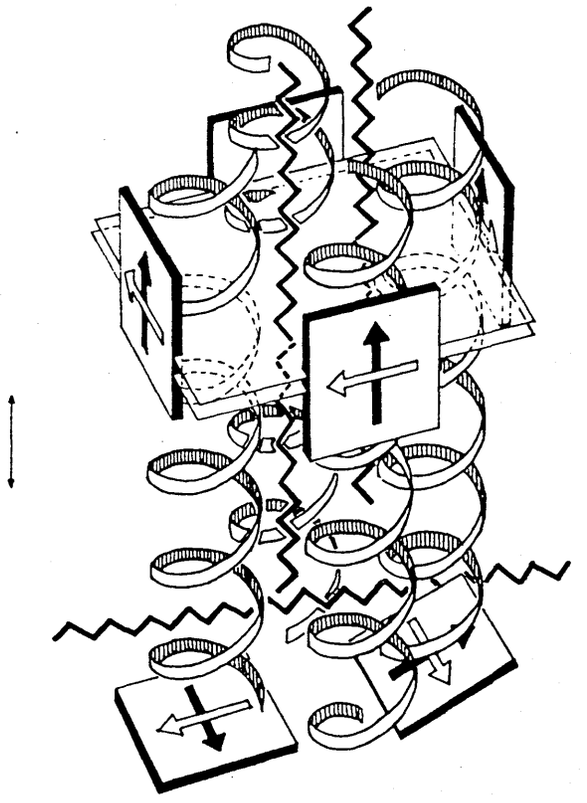


Figure 1.3

A schematic representation of the proposed model for the B800-850 antenna complex from *Rb. sphaeroides*. The basic unit consists of four Bchl 850 molecules (upper boxes), two Bchl 800 molecules (lower boxes), three carotenoids (zigzag lines) and two proteins, each consisting of two subunits. The helices are the α-helical transmembrane helices of the light harvesting polypeptides. The Q_y transitions (open arrows) of two of the Bchl 850 molecules (left front and right back) are in the same plane, while the Q_y transition of the remaining Bchl 850 molecules are in a parallel plane and vertically displaced by ≈1A. The Q_x transition moments (solid black arrows) of the Bchl 850 are perpendicular to these planes. The Q_y transitions of the Bchl 800 molecules are both in a plane parallel to Bchl 850 Q_y transitions, while the Q_x molecules are tilted out of this plane at an angle smaller than 24°. The bar represents 5A. Taken from Kramer *et al*, 1984a.

of fluorescence polarization and linear dichroism spectroscopy studies [Bolt *et al.*, 1981; Bolt and Sauer, 1979; Breton *et al.*, 1981; Breton and Vermeglio, 1982; Kramer *et al.*, 1984a], see figure 1.3. Each unit consists of two copies of each of the alpha and beta polypeptides, (alpha2beta2) giving four transmembrane helices, 4 molecules of Bchl 850, 2 molecules of Bchl 800, and 3 molecules of carotenoid [Evans, 1989; Cogdell and Crofts, 1978; Cogdell and Thornber, 1979; Sauer and Austin, 1978; van Grondelle *et al.*, 1982; Kramer *et al.*, 1984a]. The carotenoids associated with the Bchl B800 lie parallel to the membrane plane, whereas the carotenoids associated with Bchl 850 are roughly perpendicular to the plane [Kramer *et al.*, 1984a].

This model represents the minimum number of pigment molecules required to explain the spectroscopic data. However, the smallest B800-850 complex isolated was composed of 8-10 Bchl 850 molecules [Kramer *et al.*, 1984a]. Room temperature CD and 4th derivative analysis of the 800 and 850nm absorption bands suggest that the 800nm band represents monomeric bacteriochlorophyll and the 850nm band dimeric bacteriochlorophyll [Sauer and Austin, 1978; Cogdell *et al.*, 1980]. However, 77°K CD analysis suggests that within the multimeric structure the Bchl 800 molecules may be acting together rather than being strictly monomeric [Kramer *et al.*, 1984a].

The *in vivo* structure of LHI complexes is believed to be similar to that described above. In *Rsp. rubrum* strain S1 the minimal unit consists of an oligomer composed of 2 alpha and 2 beta polypeptides to which are bound 4 molecules of bacteriochlorophyll a and 2 molecules of spirilloxanthin [Brunisholz *et al.*, 1984a and b; Sauer and Austin, 1978; Cogdell and Crofts, 1978; Thornber *et al.*, 1983; Cogdell *et al.*, 1982; Picorel *et al.*, 1983]. The 890nm absorption band shows an intense CD spectrum typical of a bacteriochlorophyll dimer suggesting that two bacteriochlorophyll molecules form a closely interacting pair [Sauer and Austin, 1978]. The location of the AxxxHis sequence suggests that the tetrapyrrole rings are located in the periplasmic leaflet in the chromatophore membrane

[Brunisholz *et al.*, 1981, 1984b; Gogel *et al.*, 1983; Theiler *et al.*, 1984a and b; Tadros *et al.*, 1984; Kramer *et al.*, 1984b], and that the aromatic amino acids in conserved portions relative to histidine interact with the tetrapyrrole rings which lie parallel to the membrane plane [Brunisholz *et al.*, 1984; Breton and Vermeglio, 1982].

1.3 The Reaction Centre

Like the antenna complexes, pure reaction centre preparations have been obtained from detergent solubilised membranes by conventional protein purification techniques. This has permitted the characterisation of the reaction centre in terms of its chemical components [Reed and Clayton, 1968; Feher and Okamura, 1978; Okamura *et al.*, 1982]. Each reaction centre is composed of the following cofactors, 4 molecules of Bchl a, 2 molecules of bacteriopheophytin (Bpheo) a, 1 molecule of carotenoid. 1 atom of non-haem iron and 1-2 molecules of ubiquinone (or menaquinone as found in *C. vinosum* and *R. viridis*) [Cogdell *et al.*, 1976; Jolchine and Reiss-Husson, 1974 ,1975; Straley *et al.*, 1973]. These constituents are bound non covalently to a protein composed of three polypeptide subunits termed Light (L), Medium (M) and Heavy (H) present in a 1:1:1 ratio [Clayton and Haselkorn, 1972; Feher and Okamura 1978; Okamura *et al.*, 1982]. In *R. viridis* and (*Rb. sphaeroides*) these polypeptides are composed of 281 (274), 260 (320) and 307 (258) amino acid residues respectively [Williams *et al.*, 1983, Michel *et al*, 1986]. In *R. viridis* a cytochrome c molecule which "feeds" electrons to the Bchl "special pair" is closely associated with the reaction centre on its periplasmic side [for review see Hunter, 1988]

A simpler photochemically active unit can be obtained by treatment of reaction centres with chaotropic agents which effectively remove the H polypeptide. Though the resulting structure is less stable it shows no discernible difference in its primary photochemical activities and its cofactoral composition is identicle to that of

the intact reaction centre [Okamura *et al.*, 1974, Agalides and Reiss-Husson, 1983; Feher and Okamura, 1984]. Removal of the H-polypeptide however, does influence the transfer of energy between the primary and secondary quinones (Q_A and Q_B) [Debus *et al.*, 1985]. These results suggest that the photochemically active constituents reside with the L and M subunits. In addition it has been shown that substantial amounts of RC-H can be found in aerobically grown cells where there is no detèctable RC-M or L polypeptides [Chory *et al.*, 1984; Donohue *et al.*, 1986]. Perhaps because of this, there has been considerable speculation as to the role of RC-H. It might be inferred that the RC-H subunit is not necessary for RC function since some photosynthetic bacteria eg. *Rhodocyclus gelatinosa*, contain only two reaction centre subunits [Clayton and Clayton, 1978]. Studies of *Rb. sphaeroides* mutants suggest that the RC-H subunit may be involved in reaction centre assembly, perhaps providing a docking point within the membrane for insertion of the RC-L and M subunits. However, alternative roles such as a transmembrane environmental sensor has not been ruled out [Kiley and Kaplan, 1988].

Hydropathy plots of the L and M subunits has revealed in each five segments which are rich in hydrophobic amino acids and which are capable of forming membrane spanning helices perpendicular to the membrane plane [Williams *et al.*, 1983; Nabedryk *et al.*, 1982]. Sequence analysis suggests that each hydrophobic segment is long enough, being at least twenty residues long, to span the membrane in an alpha helix [Williams *et al.*, 1983, 1984]. These structural predictions have been confirmed by X-ray crystallographic analysis of crystals formed from reaction centres of *R. viridis*, [Michel, 1982, 1983; Michel and Deisenhofer, 1986;; Deisenhofer *et al.*, 1984, 1985a and b] and more recently from the *Rb. sphaeroides* R26 [Allen *et al.*, 1987a and b].

The electron density maps generated for the reaction centres of these species (resolving at 2.9 A and at 2.8 A respectively), have been used to construct a model describing the arrangement of

cofactors within the reaction centre. The model shows a roughly symmetrical arrangement around the vertical axis of Bchl and Bpheo pyrrole ring systems which form two branches, M and L. Two closely associated, noncovalently linked, Bchl molecules, the "special pair", are positioned near to the axis of symmetry [Deisenhofer *et al.*, 1984, 1985; Yeates *et al.*, 1987]. These Bchl molecules, known as P870, interact to give an absorption band near to 870nm and constitute the primary electron donor [Norris *et al.*, 1971; Feher *et al.*, 1975]. The positions of the centres of the tetrapyrrole rings are highly conserved, but there are some differences between the two species in the relative orientation of their macrocycles and sidechains. A striking feature of the structure is the presence of eleven transmembrane helices each consisting of approximately 22-31 residues; 5 membrane helices in both the L and M subunit, and 1 in the H subunit [Allen *et al.*, 1987b]. The bulk of the H subunit lies in the cytoplasmic side of the membrane and makes contacts with the cytoplasmic sides of the L and M subunits. Comparison of the structural data obtained by crystallography with predictions made from primary amino acid sequences and biophysical measurements shows consistent similarities between the data sets. This suggests that the reaction centre model proposed for these two bacterial species is representative of the reaction centre in the *in vivo* state.

Although a reaction centre can absorb light energy directly, energy is usually absorbed by the antennae then funnelled to the reaction centre as excitons. Reception of the energy by the reaction centre "special pair" induces their transition to the lowest excited singlet state [Norris *et al.*, 1973; McElroy *et al.*, 1969; Norris *et al.*, 1971; Netzel *et al.*, 1973] . Within ≈4 picoseconds (ps) the special pair donate an electron to the primary electron acceptor, Bpheo a, [Leigh, 1978; Prince and Dutton, 1978; Parson and Holten, 1986] see figure 1.5. Although a Bchl molecule lies between the special pair and Bpheo it does not appear to act as an electron acceptor, although it may facilitate the electron transfer. A transient charge separation occurs between the Bchl dimer and the

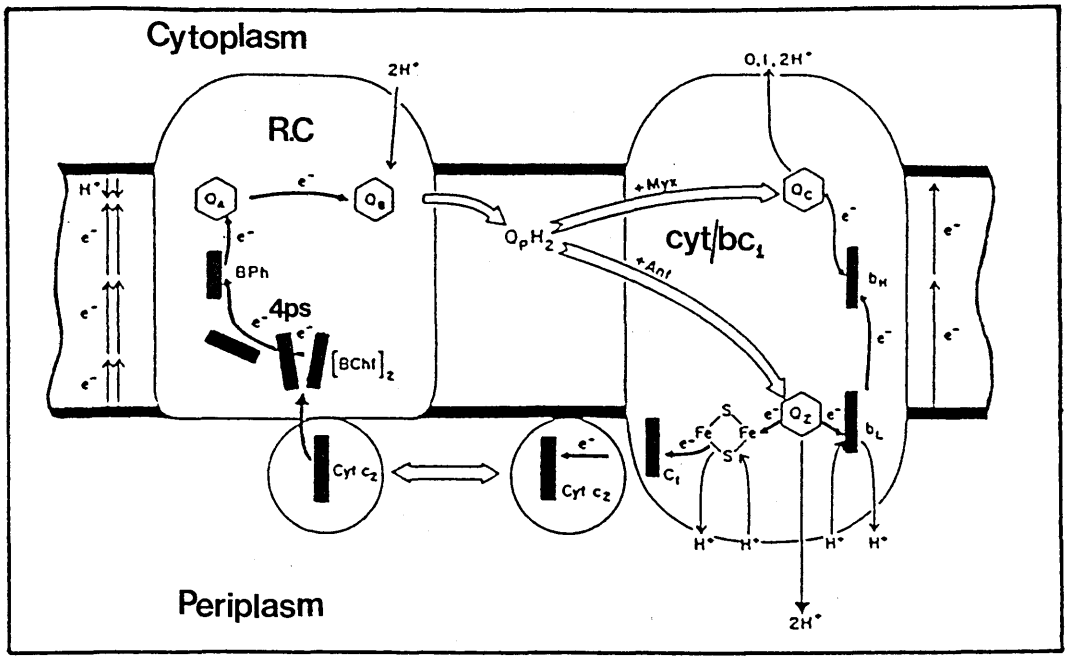
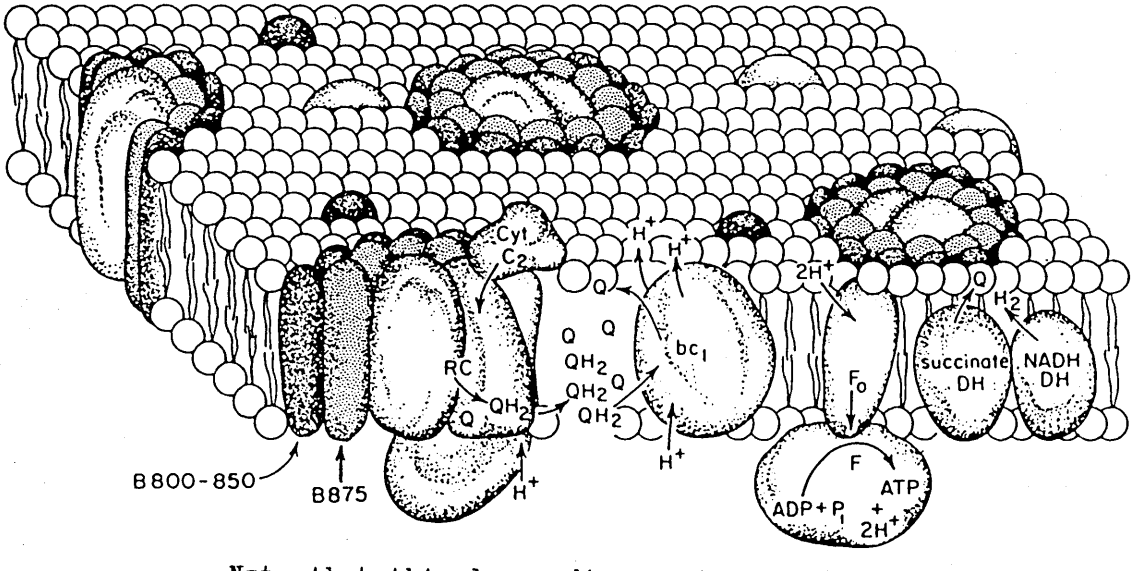


Figure 1.5

Cyclic electron flow between the reaction centre (RC) and the cytochrome bc_1 complex (cyt/bc_1). Membrane pool quinones reduce two quinones in the cytochrome bc_1 complex (Qc and Qz). Electrons then flow to cytochrome b_H and back to the membrane quinone pool, or to cytochrome c_2 in the periplasmic or lumenal space. Cytochrome c_2 then donated electrons to the reaction centre special-pair to complete the cycle. The cyclic electron flow results in a net transport of protons across the membrane. (From Robertson and Dutton 1988). This schematic representation is shown in an alternate form below (from Kiley and Kaplan, 1988). DH, Dehydrogenase; NADH, reduced nicotinamide adenine dinucleotide; ADP, adenosine 5'-diphosphate; ATP, adenosine 5'-triphosphate.



Note that this lower diagram is inverted

Bpheo molecule [Parson and Holten, 1986]. Within 200ps the electron is transferred further across the membrane to a quinone molecule, Q_A. Cytochrome c then donates an electron to the electron depleted "special pair" which stabilises the charge separation. The electron on Q_A is then passed via a second reaction centre quinone, Q_B, to the membrane quinone pool. [Robertson and Dutton, 1988; Hunter, 1988]. Light therefore drives a redox reaction resulting in charge separation and the formation of reduced quinone.

1.4 The Structural Relationships of the Pigment-Protein Complexes

Having examined the basic components of the photosynthetic unit as devolved biochemical entities, attention can now be turned on their organisation with respect to one another and the intracytoplasmic membrane.

In 1972 Aagaard and Sistrom sucessfully demonstrated the existence of two main type of antenna complex, B875 and B800-850 [Aagaard and Sistrom, 1972]. They also showed that the relative amounts of RC-Bchl a remained constant with respect to the longest wavelength component ie. B875. This suggested that a fixed stoichiometry exists between the reaction centre and B875 complexes. A number of developmental studies revealed that during photosynthetic membrane morphogenesis the reaction centre and B875 complexes were inserted into the membrane first. The B800-850 complexes were then added later [Aagaard and Sistrom, 1972; Takemoto, 1974; Niederman *et al.*, 1976; Drews *et al.*, 1977; Firsow and Drews, 1977; Pradel *et al.*, 1978; Schumacher and Drews, 1978; Hunter *et al.*, 1982]. In addition the synthesis of the reaction centre and B875 complexes were coordinated, whereas the synthesis of the B800-850 complexes occured independently [Aagaard and Sistrom, 1972; Lien and Gest, 1973; Niederman *et al.*, 1976]. These observations, and kinetic studies on fluorescence emission, led to a number of models being proposed to describe the *in vivo* relationships between reaction centres and

antenna complexes. However, the model which has recieved the greatest experimental support is that proposed by Monger and Parson [Monger and Parson, 1977]. They proposed that the reaction centres lie in a pool of interconnected B875 complex which in turn is surrounded by the B800-850 antenna complex.

From this initial model two variations arose, see figure 1.6. In the Puddle Model a single reaction centre is surrounded by B875 complex which in turn is surounded by B800-850 complex. This "photosynthetic unit" is separated from other similar units by the membrane lipid. The Lake Model proposes an association between several reaction centres within a lake of B875 antenna complex, which is surrounded by B800-850 antenna complex. This latter arrangement of complexes is supported by experimental evidence obtained by energy transfer studies [Clayton, 1980; Hayashi *et al.*, 1982]. The concentric arrangement of the antenna complexes as depicted by the model is permissible because a) the alpha helical structure of the polypeptides would permit side by side stacking of the apoproteins and b) interactions between the amino acid side chains of neighbouring helices would hold the structure together [Zuber, 1985]

The relationship between the reaction centre and the B875 complexes evolved into the theoretical concept that the reaction centre and B875 complexes were organised into a "core complex". The earliest, though indirect study of the complex, was that of Thornber (1970) in which a fraction, the so called fraction A, was sucessfully isolated from *C. vinosum* strain D. This contained the B875 antenna complex together with the reaction centre in a ratio of 45:1 respectively. Since then experimental support for the idea that the B875 complex is intimately associated with the reaction centre has come from a number of sources [Lien and Gest, 1973; Monger and Parson, 1977; Takemoto *et al.*, 1982; Drews *et al.*, 1983, Peter *et al.*, 1983 and 1984; Ueda *et al.*, 1985; Varga and Staehelin, 1985; Dawkins, 1988] in particular from mild treatment of the

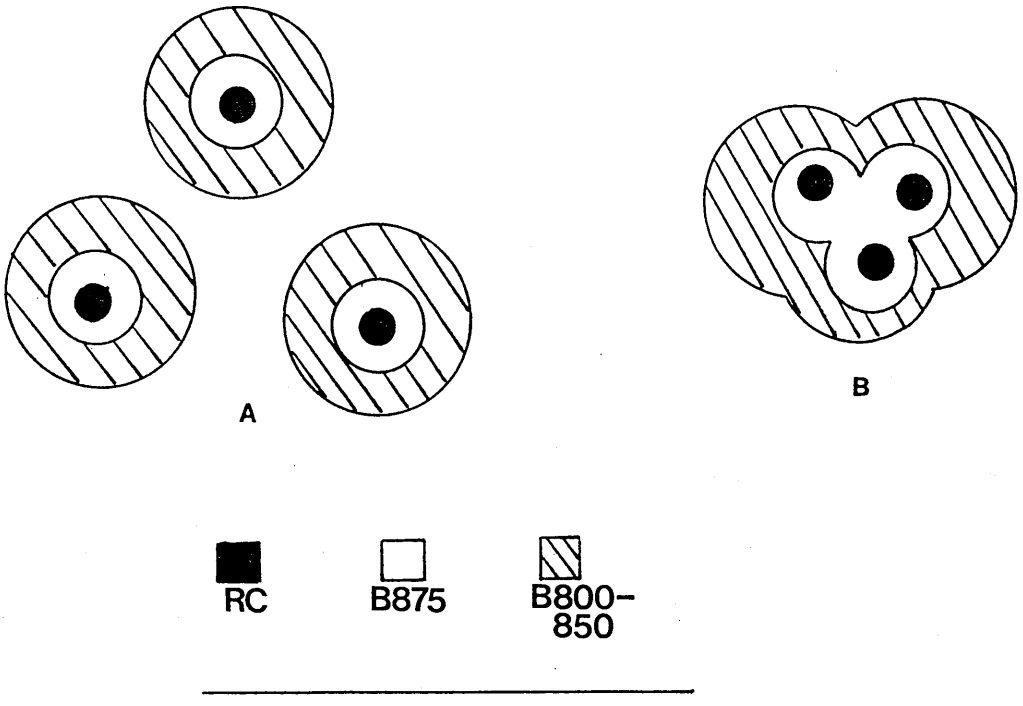


Figure 1.6

Shown above is a schematic diagram representing the two variants of the Monger and Parson model of the relationships between the reaction centre and the light harvesting complexes of the photosynthetic unit.

A=The "Puddle" model. In this model each reaction centre is surrounded by interconnected B875 complexes. These in turn are surrounded by the B800-850 complexes. There is no sharing of the B875 complexes between the individual reaction centres.

B=The "Lake" model. In this model each reaction centre is surrounded by interconnected B875 complexes. However, in this case the B875 complexes are shared by a number of reaction centres. Again the B875 complexes are surrounded by the B800-850 complexes.

photosynthetic membranes with detergent [Jay *et al.*, 1984; Dawkins 1988] and by immunofractionation techniques [Takemoto *et al.*, 1982].

High resolution electron microscopy combined with image processing of the photosynthetic membranes of *R. viridis* and *E. halochloris* have led to the low resolution of the core complex [Engelhardt *et al.*, 1983]. The micrograph images show a large central core structure which protrudes from both surfaces of the membrane [Miller and Jacob, 1983; Stark *et al.*, 1984] which is surrounded by six smaller electron dense areas [Miller, 1982].

Cross linking studies suggest that the H polypeptide of the reaction centre has close neighbour relationships with polypeptides of the B875 complex and that this plays an important structural role in the coordination of the reaction centre-antenna interactions [Takemoto *et al.*, 1982; Peter *et al.*, 1983].

The results of studies examining the assembly processes of the photosynthetic unit within the intracytoplasmic membrane have provided a quantitative description of the Monger and Parson model for *Rb. sphaeroides*. The results suggest that photosynthetic unit of chromatophores grown under low light conditions consists of 27 B800-850 complexes, existing as hexamers and higher oligomeric forms in discrete lakes, three dimeric B875 complexes, which may be associated directly with the reaction centre, and 14 B800-850 and 14 B875 complexes which appear to be functionally associated in various oligomeric states [Hunter *et al.*, 1981, 1982]. Additional evidence supporting the existence of two pools of B800-850 antenna complex, one of which is linked directly to B875 forming B800-850-B875 complexes is obtained from the results of fluorescence studies. It is suggested that these heterogeneous complexes represent fingers of B875 which extend into the B800-850 pool, and in so doing facilitate the energy transfer to the reaction centre from the outer array.

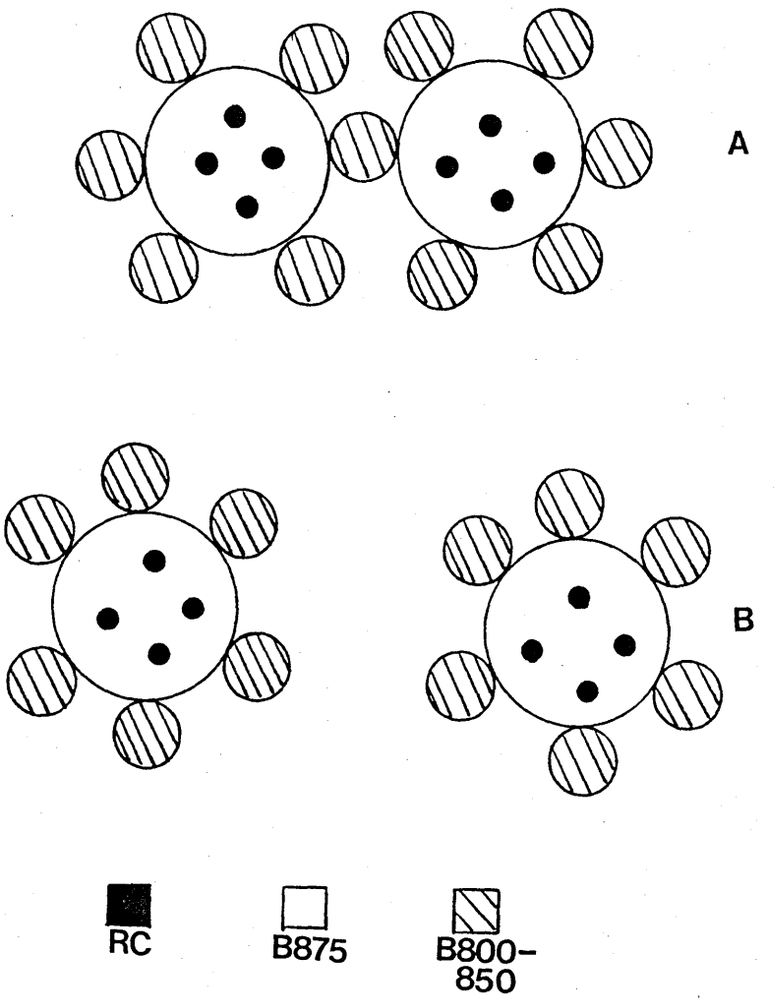


Figure 1.7

Based upon singlet-singlet annihilation measurements the following model for the organisation of photosynthetic units has been proposed.

The model suggests that at 4°K the photosynthetic unit consists of 4 reaction centres connected by ≈100 B875 Bchl a molecules and that the B800-850 complexes form about 5 domains which transfer energy, via a few contact sites, to the B875 complexes.

Two variations of this model are proposed; A=B800-850 complexes are shared between one or more core complexes; B=there is no sharing of the B800-850 complexes between the core complexes.

Fluorescence kinetic studies suggest that the reaction centre-B875 core complexes are aggregated in such a way that excitation energy can pass between different reaction centres within a lake of B875 antenna complex. A number of studies have been carried out to determine the B875 lake size [Vredenberg and Duysen, 1963; Monger and Parson, 1977; Campillo *et al.*, 1977; Bakker *et al.*, 1983]. The results suggest that the lake size is large, but decreases considerably when the experiments are repeated at low temperatures. In the case of *Rsp. rubrum* the domain consists of approximately 1000 connected antenna molecules, whereas at 4°K this number is reduced to about 150. Similarily, at 4°K in *Rb. sphaeroides* there are about 100 molecules which connect 4 reaction centres [Vos *et al.*, 1986; Westerhuis *et al.*, 1987]. The existance of a minimal domain size receives independant experimental support from developmental studies which conclude that at the earliest stage of intracytoplasmic membrane development at least four reaction centres are connected [Hunter *et al.*, 1985]. As development proceeds it is suggested thst the minimal units merge to form larger domains [Vos *et al.*, 1986]. The barriers between the units still exist but are believed to be overcome easily by excitations at room temperature and only present difficulties at lower temperatures ie. 4°K. It has also been suggested that at lower temperatures the B800-850 complex of *Rb. sphaeroides* forms five domains which transfer energy to B875 via a few contact sites. Thus at lower temperatures the Monger and Parson model no longer applies, and a new structure has been proposed [Vos et al 1986], see figure 1.7 .

1.4 The Antenna Complex and Reaction Centre Genes: Structure and Regulation.

In purple bacteria oxygen regulates the formation of the photosynthetic apparatus. When cultivated in the dark, in an oxygen saturated medium the cells lack photosynthetic membranes. Lowering the partial pressure of oxygen to ≈70Pa induces photosynthetic membrane formation. [Drews, 1986]. The synthesis of the photosynthetic membrane is not light dependent, but light does regulate the size, composition and number of photosynthetic units per cell. Therefore, at the molecular level there is a two-tier system of gene control; low oxygen concentrations "switch genes on," light intensity then modulates their expression. How this type of control is achieved, can in principle, be determined by examining the structure of the genes and their promoters.

Genetic studies have shown that the genes encoding the Bchl and carotenoid biosynthetic enzymes, reaction centre subunits and the polypeptides of the B875 complex are clustered [Marrs, 1981; Pemberton and Harding, 1986; Youvan *et al.*, 1985; Scolnik *et al.*, 1980; Sistrom *et al.*, 1984; Taylor *et al.*, 1983; Kiley *et al.*, 1987; Youvan and Marrs, 1984; Yen and Marrs, 1976] see figure 1.8. This was first demonstrated by the isolation of an R-prime from *Rb. capsulatus* carrying an insert of approximately 50kb of genomic DNA [Marrs, 1981]. The plasmid, pRS404, was shown to complement a variety of deficiencies in various aspects of photosynthetic function. Sistrom *et al.* subsequently isolated an R-prime, pWS2, from *Rb. sphaeroides* which appears structurally and functionally equivalent to pRS404 [Sistrom *et al.*, 1984].

In *Rb. capsulatus* and *Rb. sphaeroides*, the genes encoding the structural polypeptides of the photosynthetic unit have been placed within three transcriptional units, *puf* (formerly the rxcA locus), *puc* and *puh*A [Kaplan and Marrs, 1986; Belasco *et al.*, 1985; Donohue *et al.*, 1986; Kiley and Kaplan, 1987; Ashby *et al.*, 1987; Zhu and

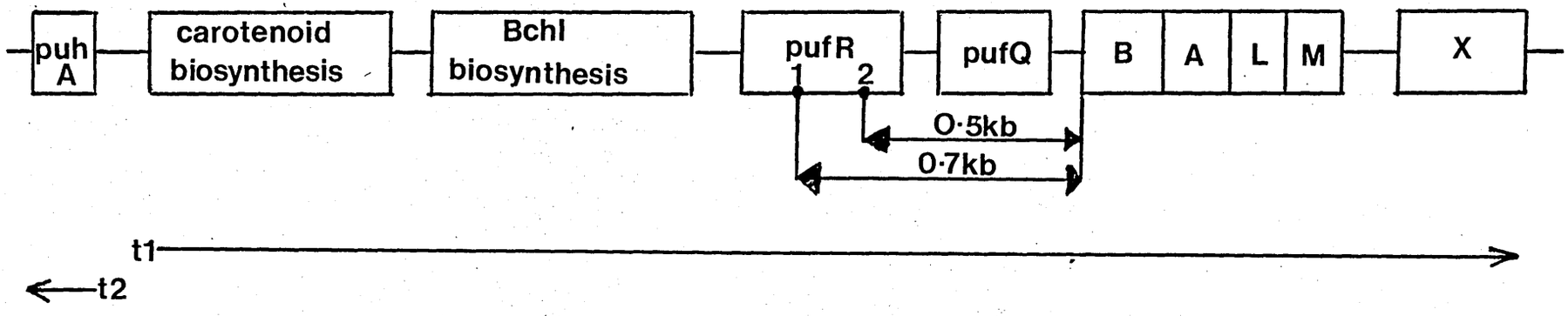


Figure 1.8
Shown above is a schematic representation of the organisation of the genes involved in bacterial photosynthesis, see text. 1 and 2 within *pufR* are the putative promoters P_{puf1} and P_{puf2} respectively. Arrows t1 and t2 indicate the relative directions of transcription of the genes to which they are subjacent.

Kaplan, 1985; Zhu *et al.*, 1986a; Zhu *et al.*, 1986b; Williams *et al.*, 1984; Youvan *et al.*, 1984]. The genes coding for the L and M subunits of the reaction centre (*pufL* and *M*), and the alpha and beta polypeptides of the B875 complex (*pufA* and *B*), have been placed in the transcriptional unit *puf*. Within this unit they are ordered *B, A, L* and *M* [Williams *et al.*, 1984; Youvan *et al.*, 1984; Youvan *et al.*, 1985].

In addition to the structural genes, *puf* consists of three other genes *pufR* or ORFR, *pufQ* and *pufX* (formerly ORF C2937 in *Rb. capsulatus*). [Farchaus and Oesterhelt, 1989, Williams *et al.*, 1983; Youvan *et al.*, 1984; Alberti and Hearst, 1989; Bauer *et al.*, 1987; Bauer *et al* 1988; Kiley and Kaplan, 1988; Zhu and Hearst, 1986]. The genes encoding the structural polypeptides of the variable B800-850 antenna complex and the H subunit of the reaction centre have been placed in the transcriptional units *puc* and *puhA* respectively [Youvan *et al.*, 1984; Donohue *et al.*, 1986; Kiley and Kaplan, 1987; Ashby *et al.*, 1987; Youvan and Ismail, 1985].

In *R. viridis* a similar genomic organisation of the *pufBALM* genes occurs [Michel *et al.*, 1986]. However, in this case an additional structural gene, encoding the reaction centre cytochrome subunit, is found downstream of *pufM*. In this species the existence of the genes *pufR, Q* and *X* have not been demonstrated [Kiley and Kaplan, 1988].

In *Rb. capsulatus* the gene encoding the H subunit of the reaction centre, *puhA*, is approximately 40kb upstream of *pufB* and is transcribed in the opposite direction to *the puf* operon[Youvan *et al.*, 1984]. In *Rb. sphaeroides puf* and *puhA* lie within a 100kb region on the R-prime pWS2 [Sistrom *et al.*, 1984] but are separated by at least 40kb [Youvan *et al.*, 1984; Kiley and Kaplan, 1988; Scolnik and Marrs, 1987; Williams *et al.*, 1986]. In both species, between *puf* and *puhA* lie the genes encoding the enzymes for carotenoid (*crt*) and Bchl biosynthesis [Pemberton and Harding, 1986; Youvan *et al.*, 1984, Yen and Marrs, 1976; Taylor *et al.*, 1983;

Zsebo *et al.*, 1984; Zsebo and Hearst, 1984]. In *R. viridis* the *puhA* gene is at least 100Kb upstream of *pufB* [Lang and Oesterhelt, 1989].

The genes encoding the alpha and beta polypeptides of the variable B800-850 antenna complex are not found within either pRS404 or pWS2 [Scolnik and Marrs, 1987; Youvan *et al.*, 1984; Sistrom *et al.*, 1984]. In *Rb. sphaeroides* and *Rb. capsulatus* these genes are found over 100kb away from *puf* and are clustered within a single operon, *puc* [Kaplan, 1989 ; Williams, 1989]. In both species the operon encodes one alpha and one beta polypeptide and the genes are ordered beta-alpha [Ashby *et al.*, 1987; Kiley and Kaplan, 1987; Youvan and Ismail, 1985]. In *R. palustris* there are four *puc* operons [Tadros and Waterkamp, 1989]. Two of these are clustered within a 1.5kb genomic segment and separated from each other by about 500bp. The remaining operons are separated from the clustered pair by at least 1kb and at least 6kb from each other. As yet the distances between the *puf* and *puc* operons is unknown. [Habib Tadros personal communication]. These operons possess the same structural organisation as those found in the other two species.

To digress briefly, the structural similarity between *pufBA* and *pucBA* operons has led to the hypothesis that gene duplication may have played an important role in the evolution of photosynthetic structural genes [Youvan and Ismail, 1985]. It has been shown that in *Rb. capsulatus* the L and M subunits of the reaction centre are extensively homologous [Youvan *et al.*, 1984]. Similarily the percentage homology between the beta (34%) and alpha subunits (27%) from the B875 and B800-850 complex suggest that both beta and both alpha genes may have had a common ancestor. Further speculation suggests that the LHII genes arose from ancestral LHI genes rather than vice versa because the LHI is more central to the reaction centre in the sense of the structure of the core complex, energy flow and gene regulation. Secondly, the LHII genes are outwith *puf*, the core complex gene cluster, a possible consequence of gene duplication involving a transposition to a second chromosomal locus [Youvan and Ismail, 1985]. This type of argument can be extended

further in *R. palustris* where further transpositional events may have given rise to the multiple *puc* operons [Tadros and Waterkamp 1989].

The *puf* operon is the most thoroughly studied transcriptional unit encoding bacteriochlorophyll binding proteins [Kiley and Kaplan, 1988]. It encodes two polycistronic transcripts, a high abundance, small (0.5kb) transcript specific to the *pufBA*, and a low abundance, large (2.5kb) transcript encoding *pufBALMX* [Belasco *et al.*, 1985; Zhu *et al.*, 1986]. The ratio of small to large transcripts in photosynthetically growing cells varies from 8-15:1. The maintainance of this mRNA ratio results directly in the molar excess of light harvesting (LHI) polypeptides over reaction centre subunits (12 LHI:1 RC) found in the core complex [Schumacher and Drews, 1978; Kaufmann *et al.*, 1982; Dierstein, 1984].

In *Rb. capsulatus* this stoichiometry is maintained, in part, as a result of differences in the stability of the two mRNA populations [Klug *et al.*, 1987; Chen *et al.*, 1988; Belasco *et al.*, 1985]. The differential stability arises from structural differences within the transcripts. The large transcript is initiated 5' to *pufB* and terminates in two hairpin loops, 3' to *pufX* [Youvan *et al.*, 1984; Belasco *et al.*, 1985]. These loops in addition to halting transcription, protect the 3' end of the transcript from 3'-5' exonucleases. Within the transcript there are three notable features; an intercistronic stem-loop between *pufA* and *pufL*, a relatively large number of endonucleolytic cleavage sites in the *pufLMX* region and relatively few of such sites in the *pufBA* region [Youvan *et al.*, 1984; Belasco *et al.*, 1985; Chen *et al.*, 1988]. Once generated the large transcript undergoes endonucleolytic cleavage, which because of the larger number of sites is likely to be in the *pufLMX* region. The loss of the two 3' hairpin loops permits attack of the remnants of the 3' end of the transcript by 3'-5' exonuclease. The attack is halted when the exonuclease reaches the intercistronic hairpin loop which provides 3' protection

to the remaining, smaller, more stable, 0.5kb transcript [Chen *et al.*, 1988].

In *Rb. sphaeroides* the differential mRNA stoichiometry is maintained by a different mechanism [Kiley and Kaplan, 1988; Zhu *et al.*, 1986]. As in *Rb. capsulatus* the *pufA-L* junction is partitioned, but in this case by two stem-loop structures [Kiley *et al.*, 1987; Zhu *et al.*, 1986b]. These have the characteristics of rho independent transcription terminators, and are believed to cause differential transcription termination. Unlike *Rb. capsulatus*, the large and small transcripts do not share the same 5' terminus. The 5' end of the 2.5kb transcript maps 75bp upstream of *pufB*, whereas that of the 0.5kb transcript is a further 29bp upstream, suggesting differences in abundance may be due to initiation at independent transcription promoters. [Zhu *et al.*, 1986b; Kiley *et al.*, 1987; Scolnik and Marrs, 1987]. Although, the possibilty that these different 5' terminii are generated as a result of postranscriptional modification of the transcripts has not been ruled out [Kiley and Kaplan, 1988; Farchaus and Oesterhelt, 1989]. Therefore, in this species a combination of differential transcription initiation together with differential transcription termination has been implicated in the formation of the intracellular amount of each transcript. Deletion of the purported intercistronic terminator region has yielded a mutant which grows photosynthetically but is phenotypically B875- [Davis *et al.*, 1988]. Such a phenotype is consistent with the intercistronic region acting as a differential terminator though more work is required to establish the validity of this hypothesis [Kiley and Kaplan, 1988].

The *puf* operon is transcriptionally repressed by molecular oxygen [Clark *et al.*, 1984; Zhu and Hearst, 1986; Klug *et al.*, 1985; Zhu *et al.*, 1986]. However, although the concensus sequences for an oxygen-regulated promoter have not been unequivocally identified. In *Rb. capsulatus*, a region ≈700bp upstream of *pufB* is critical for oxygen-dependent gene regulation [Bauer *et al.*, 1988; Adams *et al.*, 1989]. Sequence analysis of this "promoter" area, named Ppuf1 has

demonstrated the presence of a 26bp region of dyad symmetry. Immediately downstream is a sequence that closely resembles the consensus sequence recognised by the sigma sixty (*ntrA*) subunit from RNA polymerase which transcribes a variety of bacterial genes involved in nitrogen metabolism [Bauer *et al.*, 1988; Gussin *et al.*, 1986]. A feature common to promoters using this subunit is that they all appear to be highly induced by an upstream binding element [Ames and Nakaido, 1985; Reitzer and Magasanaik, 1986]. 188bp downstream of Ppuf1 is a second promoter Ppuf2. This is believed to be responsible for the low level (3-4%) constituive expression of *puf* observed during aerobic growth. This second promoter unlike the first does not contain any obvious secondary structure nor sequence homology to the ntrA-dependent promoters [Bauer *et al.*, 1988].

Located between the promoters and the *pufB* structural genes is an open reading frame designated *pufQ* which codes for a 74-amino acid polypeptide. This shows significant alignment with the *pufL* and *M* gene products but lacks the two histidines that have been implicated in the binding of cofactors to the polypeptides [Allen *et al.*, 1986; Adams *et al.*, 1989; Deisenhofer *et al.*, 1985a and b; Williams *et al.*, 1986]. Insertions into *pufQ* result in a massive decrease of Bchl or any precursors in the Mg^{2+} branch of the Bchl biosynthetic pathway. This has led to the suggestion that this gene product functions as a carrier for Bchl intermediates [Bauer and Marrs, 1987; Burgess *et al.*, 1988; Bauer *et al.*, 1988; Marrs *et al.*, 1989; Davis *et al.*, 1988].

Although the *pufBA* and *pufBALM* transcripts share the same 5' termini , recent S1 analysis with end labelled probes that extend approximately 900bp upstream of *pufB* have revealed additional 5' mRNA termini to those previously known to exist [Adams *et al.*, 1989]. The most 5' of these highly labile transcripts extend upstream to the hairpin loop region that contains the oxygen regulated promoter Ppuf1. It seems likely that an area of bidirectional symmetry within this region functions as an operator

to bind proteins that regulate puf transcription initiation in response to changes in O_2 levels [Adams *et al.*, 1989].

This work has led to a revised transcriptional model for the puf operon. It is now believed that the operon *pufQBALMX* is transcribed as a single polycistronic message. Transcription begins at a promoter located approximately 600bp upstream of *pufB*. Initiation of transcription is regulated by a trans acting factor that either directly or indirectly senses the amount of O_2 available to the cell and acts at site located near the promoter. Transcription occurs from the promoter through to the region 3' of *pufX*. The first 300bp of the nascent transcript are not translated, translation begining at the start codon of *pufQ* and continuing through to *pufX*. Before transcription is terminated by the hairpin-loops 3' to *pufX*, cleavage and degradation of the highly labile segment 5' to *pufB* begins. The half life of this region which codes for *pufQ* is tens of seconds. This rapid rate of degradation and hence low level of expression reinforces the belief that the role of Q is catalytic rather than structural. The endonucleolytic cleavage of the slower degrading region 3' to *pufB* then proceeds as described earlier [Adams *et al.*, 1989]. Sequence comparison of *pufQ* with the region upstream of *pufB* in *Rb. sphaeroides* suggests the existance of *pufQ* in this organism [Belasco *et al.*, 1988].

Although oxygen affects the expression of *pufQ* which in turn influences the expression of the other photosynthetic genes (see later) there still exists the fundamental question; how does oxygen regulate gene expression? Several hypotheses such as redox potential control [Marrs and Gest, 1973] or aporepressor-corepressor interactions [Kaplan, 1978] have been proposed as mediators of the environmental signal, but the exact mechanism has not been clarified. Yamamoto and Droffner have suggested that the expression of genes required for anaerobic growth in *Salmonella typhimurium* requires the supercoiling of chromosomal DNA, whereas the expression of genes required for aerobic growth depends on DNA

relaxation [Kranz and Hazelkorn, 1986;Yamamoto and Droffner, 1985]. Superhelical tension has also been implicated in the regulation of expression of numerous prokaryotic and eukaryotic genes [Gellert, 1981; Lam and Chua, 1987; Novak and Maier, 1987; Stirdivant et al 1985]. In bacteria DNA topology is mainly determined by two enzymes, topoisomerase I and gyrase [Otter and Cozzarelli, 1983; Wang, 1985]. In a study of mRNA accumulation from essential photosynthetic genes, Zhu and Hearst demonstrated the inhibition of gyrase resulted in the loss of detectable *puf* and *puhA* RNA [Zhu and Hearst, 1988]. Loss of detectable *puhA* RNA occurred within 15mins of addition of gyrase inhibitors, a time comparable to the half life of this RNA species. In contrast, mRNA levels for the *fbc* operon which codes for the cytochrome bc1 complex and is constitutively expressed under aerobic and anaerobic conditions is unaffected by gyrase inhibitor treatment [Gabellini and Sebald, 1986; Gabellini *et al.*, 1985; Zhu and Hearst, 1988].

The rate of decrease of *puf* and *puhA* RNA levels have been interpreted to imply that the observed effects result directly from DNA gyrase inhibition and are not a secondary response to drug treatment. This has led to the hypothesis that DNA superhelicity in the region of the genes for photosynthesis is altered by DNA gyrase leading to repression or derepression of transcription [Zhu and Hearst, 1988]. A simple model would then be that different promoters have different sensitivities to DNA topology. An alternative would be a more complex interaction between the promoter, an upstream regulatory region of DNA [Thompson and Mosig, 1987], or regulatory protein and gyrase.

The different sensitivities of promoters to DNA topology has been demonstrated in other organisms [Gellert, 1981; Thompson and Mosig, 1987; Menzel and Gellert, 1983; Menzel and Gellert, 1987] This model may explain why most of the essential genes are clustered on a 50kb region of the chromosome [Marrs, 1981; Sistrom *et al.*, 1984]. Since the bacterial chromosome has been shown to be composed of topologically independent domains of torsionally strained DNA, the

photosynthetic gene cluster could be contained within such a domain and be subject to facile regulation by gyrase [Sinden and Pettijohn, 1981; Pettijohn and Pfenninger, 1980; Worcel and Burgi 1972].

However, Cook *et al.* have developed an assay that uses psoralen cross-linking to detect local changes in chromosomal supercoiling [Cook *et al.*, 1989]. From their results they conclude that the requirement for DNA gyrase in the transcription of photosynthetic genes is not related to gyrase induced change in local chromosomal superhelicity. The loss of mRNAs for photosynthesis after treatment with gyrase inhibitors results indirectly from inhibitor treatment. That is, gyrase is an essential enzyme for DNA metabolism and loss of gyrase activity may have numerous effects on the cell none of which directly reflect its immediate role in bacterial physiology. This led to an alternative hypothesis to that of Yamamoto and Droffner for the role of gyrase in transcription. Whenever a cell switches from one form of metabolism to another a substantial involvement in the production of new enzyme systems is required. This in turn requires a high level of transcription of the genes coding for the structural proteins. Recent evidence suggests that on highly expressed regions of topologically anchored DNA, RNA polymerases divide the template into domains and introduce positive and negative supertwists by rotating the DNA through the transcription complex [Brill and Sternglanz, 1988; Gamper and Hearst, 1982; Giaever and Wang, 1988; Tsao *et al.*, 1989; Wu *et al.*, 1988]. The requirement for gyrase under these conditions could simply reflect the need to maintain the chromosome at its steady state superhelicity by acting as a swivel for restoring negative supercoils. Thus instead of activating transcription by altering DNA topology, gyrase could be required as a consequence of transcription [Cook *et al.*, 1989].

Of the genes encoding bacteriochlorophyll binding proteins, the expression of the *puc* operon (encoding the LHII antenna complexes) is the most stringently regulated by oxygen concentration [Zhu *et*

al., 1986; Zhu and Hearst, 1986; Zhu and Kaplan, 1986; Kiley and Kaplan 1987]. In *Rb. sphaeroides*, for example, grown photosynthetically there is three-six times more *puf* and *puhA* specific mRNA than in cells grown chemoheterotrophically in 30% oxygen. In contrast, the amount of *puc* specific mRNA is at least 100 fold greater in photosynthetically grown cells [Zhu and Kaplan, 1985; Donohue *et al.*, 1986; Kiley and Kaplan, 1988].

In *Rb. sphaeroides*, Northern blotting has shown that the two genes, *pucB* and *pucA*, are encoded by a single 640bp polycistronic transcript [Kiley and Kaplan, 1987]. The use of a coupled transcription-translation system suggests that promoter sequences are within 211bp upstream of the structural genes. This system has also been used to show the presence of transcriptional units downstream of the *puc* genes [Kiley and Kaplan, 1987]. How these genes interact with the *puc* operon is unclear, but it is known that sequences extending 2.0kb downstream contribute to B800-850 assembly [Kaplan, 1989].

In *Rb. capsulatus* the use of polyacrylamide rather than agarose gels has resolved two *puc* specific messages of 491 and 505bp [Zucconi and Beatty, 1988]. This differs from previous reports which have only mentioned one transcript [Zhu et al., 1986; Zhu and Hearst, 1986]. Perhaps because of this, the number of transcripts found in *Rb. sphaeroides* is under review [Kaplan S. personal communication].

In *Rb. capsulatus* transcription initiates at two different sites. It is not known if there are two promoters (ie. one for each transcript) or a single promoter that directs transcription initiation at two sites. Examination of the nucleotide sequence upstream of the start sites showed the lack of *E. coli* -10 or -35 consensus sequences. This is not surprising in view of the fact that of all *Rb. capsulatus* promoters tested in *E. coli*, not one has proved functional [Johnson *et al.*, 1986]. However, there exists a direct repeat which maps upstream of the 5' end of the longer transcript. The centres of these repeated sequences are separated

by 14bp, the approximate distance between the 5' ends of the two transcripts. In addition, the centre of the most upstream repeated sequence is approximately 25bp from the start site of the longer transcript, whereas the centre of the downstream sequence is about 26bp upstream from the initiation site of the shorter transcript. The spacing of the repeated sequences in relation to each other and to the 5' ends of the mRNA suggests the sequences may be involved in transcription initiation. The transcripts share a common 3' terminus which is in the region of a potential 10bp stem-loop structure which may serve as a transcription terminator. Both transcripts have the same rate of decay, are present in equal amounts, and show no apparent change in this pattern between high and low light grown cells.

The function of the third 14kd colourless polypeptide found in the *Rb. capsulatus* B800-850 complex remains obscure. A similar mist shrouds the genes and transcripts which encode it [Zucconi *et al.*, 1988; Shiozawa *et al.*, 1980; Drews and Oelze, 1981]. In addition to the promoters descibed by Zucconi *et al.*, regions 300bp upstream and 600bp downstream of *pucBA* are also required for B800-850 expression [Tichy *et al.*, 1988; Oberle' *et al.*, 1988]. Tn5 insertions into the region 600bp downstream of pucBA significantly reduces the level of puc mRNA [Oberle' *et al.*, 1988]

The levels of LHII RNA are strongly influenced by the concentration of environmental oxygen [Zucconi *et al.*, 1988; Zhu and Hearst, 1988; Zhu and Kaplan, 1988; Klug *et al.*, 1985; Zhu *et al.*, 1986; Zhu and Hearst, 1986]. Comparison of the relative quantities of LHI and LHII mRNAs in photosynthetically grown high light cells indicate that the LHII mRNA is the more abundant [Zucconi *et al.*, 1988; Zhu *et al.*, 1986]. The results of transcription inhibition experiments suggest that *pucBA* mRNA is more stable under anaerobic conditions than that of *pufBA* which in turn is more stable than *pufLM* and *puhA* [Zhu *et al.*, 1986].

When the cells are transfered from anaerobic to aerobic conditions there is an immediate decrease in the levels of all the mRNA species. However, the level of LHII mRNA is more sensitive to oxygen as its level decreases more rapidly than the level of LHI mRNA. These results suggest that oxygen not only represses the expression of genes coding for the core complex but accelerates the degradation of LHII mRNA. The mRNAs encoding several enzymes in the bacteriochlorophyll biosynthetic pathway are regulated in a similar manner [Biel, 1986; Biel and Marrs, 1983; Zhu and Hearst, 1986]. How oxygen both inhibits transcription and accelerates the degradation of LHII mRNA is unknown but possiblities include; LHII mRNA becomes more susceptible than LHI mRNA to RNase in the presence of oxygen, or that the two mRNA species are degraded by two RNases which differ in activities in the presence of oxygen, or oxygen blocks the translation of LHII mRNA resulting in its releases from the ribosome making it susceptible to cleavage [Zucconi *et al.*, 1988].

Light appears to modulate the quantity of LHII polypeptides at a different level. *Rb. capsulatus* grown photosynthetically under low light conditions ($2W/m^2$) contain about four times the amount of B800-850 antenna complex as cells grown under high light intensity ($140W/m^2$). However, the low light cells contain about a quarter of the LHII mRNA of the high light grown cells. The steady state levels of mRNAs depend on both rates of transcription and degradation. High light grown LHII mRNA has a half life of 10mins whereas the same transcripts in low light grown cells have a half life of 19mins. As the mRNA isolated from high light grown cells degrades more rapidly than the low light counterpart, but is present in a greater quantity, then the level of transcription must be greater in the high light grown cells. If it is assumed that the amount of mRNA parallels the amount of polypeptide then there would be a pool of light harvesting polypeptides that had not bound bacteriochlorophyll in high light grown cells [Zucconi *et al.*, 1988].

It has been shown by Dierstein that the half life of alpha polypeptides which have not bound bacteriochlorophyll is much shorter those which have. Therefore, bacteriochlorophyll in some way protects the polypeptide from proteases. Perhaps the apoproteins exist in a protease sensitive conformation that changes to a resistant conformation with the binding of bacteriochlorophyll [Dierstein, 1983; Dierstein, 1984; Klug *et al.*, 1986].

From mRNA levels it has been shown that high light grown cells have the potential to generate four times as much antenna complex as low light grown cells. However, since less Bchl is generated in high light grown cells the newly translated polypeptides would be turned over at a much greater rate as there would be a limiting number of Bchl molecules for binding. Free bacteriochlorophyll is not detected in wild type cells so the ratio of polypeptides to pigments must be greater than 1. It is well established that the total Bchl content of *Rb. capsulatus* increases as the light intensity decreases.[Aiking and Sojka, 1979; Biel, 1986; Drews, 1986; Drews and Oelze, 1981; Lien *et al.*, 1973; Schumacher and Drews, 1979]. When cells are shifted from high to low light intensities more Bchl is produced which could bind to the available pool of light harvesting polypeptides. These chromoproteins would then be resistant to degradation and could become involved in light harvesting. An advantage of this type of control is that it would allow a rapid increase in light harvesting ability. Expansion of the photosynthetic unit would require the synthesis of key enzyme(s) involved in Bchl synthesis rather than de novo initiation of transcription of light harvesting polypeptide genes [Zucconi *et al.*, 1988].

A pivotal role in this pathway appears to be played by the *pufQ* gene product [Bauer and Marrs, 1987; Adams *et al.*, 1988; Bauer *et al.*, 1988; Ashby *et al.*, 1987]. In *Rb. sphaeroides* a mutant, PUFB1, generated by deletion of part of pufQ, lacks detectable RC and B875 complexes. However , this also resulted in a pleiotropic reduction in the level of the B800-850 complexes, bacteriochlorophyll and

carotenoid content. A similar effect has been observed in analogous *Rb. capsulatus* mutants [Davis *et al.*, 1988]. Though there is a decrease in the level of B800-850 complexes in the membrane there was a substantial increase in the quantity of *puc* specific mRNA. A similar uncoupling between the levels of *puc* specific transcripts and cellular levels of B800-850 complexes have been observed whenever mutants defective in RC or B875 assembly are observed [Davis *et al.*, 1988; Farchaus and Oesterhelt, 1989].

Unlike the genes encoding the reaction centre and antenna polypeptides and enzymes involved in the bacteriochlorophyll biosynthetic pathway, the genes encoding the enzymes involved in carotenoid biosynthesis are regulated in an opposite fashion [Zhu and Hearst, 1986; Zhu *et al.*, 1986]. That is when the cells are growing under anaerobic conditions and the light intensity or oxygen concentration increases, there is an increased expression of the *crt* genes. Under steady state high oxygen concentrations and high light intensities the levels of mRNA encoding a number of carotenoid biosynthetic enzymes are high, though the cells remain unpigmented. Likewise when shifted from low to high light or anaerobic to aerobic conditions levels were also seen to increase. Such shifts are likely to be a response to minimise photoxidative damage which occurs in high light in the presence of oxygen [Foote and Denny, 1968; Krinsky, 1979; Cogdell, 1981; Cogdell, 1985]. Whether the increases in concentration of *crt* mRNA occur due to increases in transcription or decreases in degradation is as yet unclear.

In addition to genes encoding structural polypeptides or biosynthetic enzymes there are a number of genes which have been implicated in the assembly of the photosynthetic unit [Farchaus and Oesterhelt, 1989; Farchaus *et al.*, 1989; Kiley and Kaplan, 1988; Youvan *et al.*, 1984; Zhu and Hearst, 1986] These genes appear to have a parallel genomic organisation in the species examined.

In *Rb. sphaeroides* an open reading frame, *pufR*, is found upstream of *pufQ* [Kiley and Kaplan, 1988]. Alberti and Hearst have found an

organisationally analogous open reading frame, ORFR, in *Rb. capsulatus* [Alberti and Hearst, 1988]. ORFR is 1473bp long and lies 1bp upstream of *pufQ*. The transcription of both genes are regulated by light and oxygen concentations. The role of the ORFR gene product is unknown and a function has not been proposed. The function of the *pufR* gene product is unclear but it believed to be involved in LHI assembly [Kiley and Kaplan, 1988]. Evidence for this comes from complementation analysis of a mutant, RS103, which is deficient in B875 complexes [Jackson *et al.*, and Meinhardt *et al.*, 1985]. This mutant can be restored to wild type with the pufR gene region in trans. It is known that this mutant is not defective in *pufBALMX* transcription, thereby suggesting that the *pufR* gene, or closely linked region, controls synthesis or assembly of the B875 complexes. This same region can also restore a *puhA* deletion /insertion mutant which is phenotypically B875- RC- to B875+ (when grown anaerobically in the dark on DMSO) without the concomitant restoration of RC complexes. It has recently been suggested that *pufR* and sequences 3kb upstream of *puhA* act in concert, to promote B875 assembly [Kaplan 1989].

The control of reaction centre insertion into the membrane as well reaction centre assembly appear to be controlled by the H-subunit of the reaction centre which is encoded by *puhA* [Kaplan, 1989; Donohue *et al.*, 1986; Kiley *et al.*, 1988; Youvan *et al.*, 1984]. *In vitro*, synthesis of an immunoreactive *Rb. sphaeroides* RC-H subunit using *puhA* plasmids indicate that the promoter sequences required for expression are within a 450bp upstream region of the *puhA* gene [Donohue *et al.*, 1986]. An analogous situation exists in *Rb. capsulatus*, initiation occuring at a site within the upstream region of ORF F1696 [Youvan *et al.*, 1984]. In both species Northern blot analysis with a *puhA* specific probe indicates that the RC-H polypeptide is encoded by two different transcripts of ≃1400 and ≃1100bp, the larger having a 5' terminus upstream of the smaller. Whether these different 5' terminii arise through differential RNA processing or because they are initiated at different sites is unknown. The sizes of the transcripts suggest that the puhA gene is

transcribed as a monocistronic operon and that the role of the 5' non-coding (0.4-0.6kb) may be regulatory [Donohue *et al.*, 1986].

Northern/dot blot analysis has also revealed that the overall transcription of this gene is controlled by light intensity and oxygen concentration. Light intensity determines the relative amounts of the two RNA species which varies from 2:1 to 3:1 for *Rb. sphaeroides* and 0.5:1 to 2.5:1 for *Rb. capsulatus* for the ratio of large:small transcript when grown under high and low light conditions respectively [Donohue *et al.*, 1986; Zhu and Hearst, 1986]. Although the levels of both transcripts decrease in dark or aerobic conditions they can still be detected within total RNA isolated from aerobic cells lacking detectable amounts of Bchl or ICM [Clark *et al.*, 1984; Zhu and Hearst, 1986]. This is perhaps not surprising in view of the fact that immunochemically detectable levels of the RC-H subunit are present within membranes of aerobically growing cells [Chory *et al.*, 1984; Clark *et al.*, 1984].

The gene *pufX* which lies downstream of *pufM* has recently been implicated in photosynthetic growth [Farchaus and Oesterhelt, 1989; Farchaus *et al.*, 1989; Kiley and Kaplan, 1988]. Farchaus and Oesterhelt using interposon mutagenesis created a mutant lacking a region extending from midway through *pufL* to a point 300bp downstream of *pufX* ie. *pufLMX* minus [Farchaus and Oesterhelt, 1989]. The mutant lacked L and M reaction centre subunits with reduced levels of the H subunit. Both the B875 and B880-850 antenna complexes were present in the photosynthetic membrane, though the lack of reaction centre made the mutant phenotypically photosynthetic minus. In trans complementation with a plasmid containing a *pufBALMX* insert complemented the deletion. Preliminary experiments with a plasmid containing *pufBALM* insert ie. X- but containing the 3' transcription terminators/exonuclease inhibitors suggested that *pufX* was not required to restore the photosynthetic+ phenotype. However, later experiments suggest that the *pufBALM* fragment is insufficient for the restoration of the phenotype [Farchaus *et al.*, 1989]. Therefore, either *pufX* and/or the region

3' of *pufX* is critical for restoring the photosynthetic plus phenotype. Spectroscopic and protein analysis of cells complemented with the *pufBALM* construct indicated that the pigment-binding complexes including those of the reaction centre were expressed and showed normal absorption characteristics. In addition there was a 20% reduction in the amount of B800-850 complexes, but a corresponding increase in the amount of B875 complex. The level of total photobleachable RC Bchl remained unchanged [Farchaus *et al.*, 1989]. These observations suggest that *pufX* may be involved in maintaining core complex stoichiometry in addition to assisting the insertion of the reaction centre into the membrane [H. Gruenberg personal communication].

Although the core and LHII complexes are encoded within different operons, there is clearly considerable interaction of gene expression. This work has been undertaken to begin the unravelling of the interactions which may take place in *R. acidophila* strain 7050. As this organism can produce at least three different types of LHII complexes the interactions which occur between the *puf* and *puc* operons, and the *puc* operons and the environment is likely to be even more complex.

CHAPTER 2 METHODOLOGY

Strategy for the Isolation, and Sequencing of the *R. acidophila* Strain 7050 Light Harvesting Genes.

In order to clone and sequence the genes encoding the *R. acidophila* strain 7050 light harvesting (LHII) polypeptides a genomic DNA library was constructed. Summarised this involved the following. Random genomic DNA fragments were generated by digestion of *R. acidophila* strain 7050 genomic DNA with Sau3AI. The digestion products were then resolved on the basis of size by passage through a sucrose velocity gradient. Fragments of ≈20kb in size were ligated to the arms of the replacement vector, lambda EMBL3. The recombinant molecules were then encapsidated using an *in vitro* packaging system. The library was amplified, then screened using a radiolabelled probe corresponding to the genes encoding the B800-850 structural polypeptides from *Rhodobacter sphaeroides*. Positively hybridising plaques were picked, rescreened then grouped according to restriction pattern. Four different groups of clones were found. Clones from three of these groups were mapped, then small, positively hybridizing fragments from a member of each group subcloned into mp19. These fragments were then sequenced by the dideoxy chain termination method. The sequences were then compared to those of the genes encoding the B800-850 polypeptides from *Rb. sphaeroides* using "Bestfit" a sequence comparison computer package. Sequences which showed similarity to these genes were translated and the deduced amino acid sequences compared to those of the LHII polypeptides of *R. acidophila*.

Techniques which were used routinely are presented in this chapter. The composition of solutions, media, etc. are presented in Appendix 1. They can be found in Appendix 1 by refering to the appropriate number in parenthesis in this chapter eg. 10x TBE (1.28). Some of the sections within the chapter refer to other sections within the

chapter. In these cases the first number in parenthesis begins with 2 eg. (2.13).

General notes on reagents, equipment, and their suppliers are described in Appendix 2.

All centrifuge rotors used, unless otherwise stated, are Beckman types. All DNA restriction and modifying enzymes were used in accordance with the manufacturer's recomendations, see (1.33).

Contents:

2.1 Growth of *R. acidophila* Strains 7050 and 7750.

Rhodopseudomonas acidophila strains 7050 and 7750 were grown in Pfennig's medium (1.1) which uses succinate as the main carbon source [Pfennig, 1969]. All cultures were grown in a thermostatically controlled growth room at approximately 30°C. Incandescent bulbs were used to create two light regimes:

High light (42 $W.m^{-2}$): Two rows of 3 x 150W bulbs were placed 75cm apart, the bulbs within each row being separated by 25cm from their neighbour. The bacterial cultures were placed midway between the two rows of bulbs.

Low light (9 $W.m^{-2}$): Cultures were placed 70cm from a single 40W bulb. Light intensities were measured with a UDT Model 40X OptoMeter in the wavelength range 450-910nm.

Stock bacterial cultures were kept as agar stabs (1.2) which were stored on the laboratory shelf. When liquid cultures were required, the stabs were filled with Pfennig's medium, to exclude air, and placed in the growth room for 7-14 days. During this time the cells grew out onto the surface of the agar and into the liquid. The liquid was transferred to a clean McCartney bottle, filled with Pfennig's medium, then returned to the growth room for 3-4 days. The treatment which the cells now recieved depended upon the strain:

Strain 7050: The culture was divided equally between two flat sided 500ml bottles which were then filled with Pfennig's medium. One bottle was placed under a high, the other under a low light regime. After 3-4 days the strain of the culture was checked by its colour. In strain 7050 the high light cultures were red-brown and the low light cultures red-purple. About 50ml of the low light culture was then added to ≈450ml of Pfennig's medium in a 500ml flat sided bottle. The cells were returned to low light conditions for no

longer than 18 hours. If left beyond this time DNA extraction proved immpossible as the cells would not lyse.

Strain 7750: The culture was added to a 500ml flat sided bottle, filled with Pfennig's medium then placed in high light for 18 hours.

The cells were harvested by centrifugation in a JA10 rotor at 12,000 x g at 4°C for 10 minutes. The supernatants were decanted and a total of 100ml of lysis solution (50mM glucose, 25mM Tris.HCl, 10mM EDTA pH8.0) added to the pellets. The pellets were resuspended to wash the cells then pelleted as before. The supernatants were discarded and the pellets resuspended in a total of 50ml of lysis solution. The cells were used immediately for DNA isolation.

2.2 Preparation of High Molecular Weight Genomic DNA.

This procedure is mainly a hybrid of two methods, that of Birnboim and Doly (1979) for the isolation of plasmid DNA, and that of Klug and Drews (1984) for the isolation of genomic DNA from *Rb. capsulatus*. However, parts of the protocol are my own and other parts, though not mine, are difficult to attribute to particular individuals or groups. The quantities of some of the reagents may appear very "rough and ready" but the method works and is presented in full, as used.

About 100mgs of lyzosyme (Grade I from chicken egg white) were placed in a 1.5ml Eppendorf tube, 1ml of lysis buffer (50mM Glucose, 25mM Tris.HCl pH 8.0, 10mM EDTA pH 8.0) added then the tube vortexed until the lysozyme dissolved. The lysozyme solution was added to ≈15mls of freshly harvested cells in a 50ml Falcon tube. The mixture was vortexed, then placed in a 37°C water bath for 1 hour to allow degradation of the cell walls. During this period the cells usually changed to a salmon pink colour and the suspension became granular in appearance.

A test aliquot (≈1ml) of cell suspension was removed to an Eppendorf vial containing a few drops of 10% w/v SDS and inverted a few times. If the cells in the test aliquot lysed, the solution cleared and became viscous. If they did not lyse more lysozyme was added to the main lysis solution and incubation continued for at least another hour. An aliquot was then tested for lysis as before. If the cells did not lyse at this stage the main lysis suspension was discarded and lysis of a fresh batch of cells attempted.

When the cells in the test solution lysed, 10% w/v SDS was added to the main lysis solution to a final concentration of 1%. The mixture was swirled gently then left to stand for 30 minutes at room temperature. During this time ≈100mgs of RNase A were placed in a test tube and 1ml of lysis buffer added. The mixture was vortexed until the solids dissolved, boiled for 15 minutes, to destroy DNase activity, then left to cool slowly to allow renaturation of the RNase [Maniatis *et al.*, 1982]. About 100mg of Proteinase K was added to a 1.5ml Eppendorf tube, 1ml of lysis buffer added and the mixture vortexed until the solids dissolved. The DNase free RNase and Proteinase K solutions were added to the lysate and gently mixed in. The lysate was left to stand at room temperature overnight.

An equal volume of phenol/chloroform (1.14) was added to the lyste and the mixture placed on a "Rock and Roller" for 30 minutes. To separate the phases the mixture was centifuged in a JS-7.5 rotor at 10,000 x g for 15 minutes at 4°C. Avoiding the protein interface, the upper, DNA containing phase was removed with a wide bore pipette to a clean Falcon tube. This extraction-centrifugation procedure was repeated until centrifugation generated little protein at the interface and most of the colour had been removed from the DNA solution.

If neccessary the DNA was concentrated by extracting the DNA solution with an equal volume of butan-1-ol to remove water. However, where possible this was avoided as it appeared to lead to shearing of the DNA.

The DNA was then placed in a tight dialysis sac (1.16) and dialysed against five, 5 litre changes of TE (1.20). Each change was made after ≈12 hours. The DNA concentration was then measured with a spectrophotometer and an ethidium bromide plate (2.6), then tested for its abilty to be cut with a range of restriction enzymes. The digests were run with an uncut DNA sample on a 0.4% agarose gel (2.5). If the size of the uncut DNA was >50kb, and was capable of being digested with restriction enzymes, it was used as the starting material for future experiments; eg. the construction of the genomic library or for genomic Southern blotting. Meanwhile it was stored at 4°C until required.

2.3 Preparation of Lambda EMBL3 Vector DNA.

The method is based on the method of Hans Lehrach and Annemarie Frischauf presented in the EMBL Laboratory Manual. The manual is undated and gives only an address: EMBL, Heidelberg, Meyerhofstr 1, Germany. Isolation of the vector DNA is a seven stage process involving; i) isolation of a plaque and checking the phenotype of the phage, ii) using the plaque to make a high titre lysate, iii) using the lysate to inoculate a batch culture, iv) removing cell debris, v) precipitation of the phage, vi) banding the phage in CsCl gradients, and vii) purifying the phage DNA. The phage were propagated on *E. coli* strains Q358 and Q359.

Isolation of a plaque 0.4 mls of Q358 plating cells (2.4) were added to 3mls of L-broth top layer (1.5) and poured onto an L-broth agar (1.4) plate. A loopfull of EMBL3 phage stock was streaked onto the lawn of Q358 and the plate incubated at 37°C overnight. An isolated plaque was then picked into 1ml of phage buffer (1.19) then vortexed and left for 1 hour at room temperature.

Checking the phage and host phenotypes: To test that the phage had the Spi^+ phenotype (see Chapter 3.1), 40µls of phage suspension were removed, and a series of tenfold dilutions made. 0.1mls of each

dilution were dispensed into each of four tubes. 0.4mls of Q358 plating cells were added to two of these. 0.4mls of Q359 plating cells were added to the other two. The suspensions were vortexed briefly then left on the bench for 20mins to allow phage attachment. 3mls of BBL-top layer (1.8) was added to the cells and the mixture poured onto a BBL-broth agar (1.7) plate. The plates were incubated overnight at 37°C.

In the morning the titre of the stock was determined by counting the plaques on the Q358 plates. The genetic properties of the phage and the Q359 cells were determined by counting and examining the Q359 plates for plaques. The plaques on the Q359 plates, at a given dilution, were more than ten thousand-fold fewer in number and much smaller in size (pin pricks) than those propagated on the Q358 host. This suggested that both the phage, and the Q359 host, had the required phenotypic characteristics for the construction and plating of a genomic library.

The High Titre Lysate: To make a high titre phage lysate for large scale lambda culture, 0.1mls ($\simeq 10^7$ pfu) of the isolated plaque phage stock were added to each of two tubes containing 0.4mls of Q358 plating cells. The contents were vortexed and left on the bench for 20 minutes to allow the phage to adhere to the cells. 3mls of L-broth top layer agarose (1.5) were added, and the lawns poured onto thick, wet, L-broth agar plates (1.4). The plates were tapped on the shelf of the incubator (lid side up) so that the condensation on the lids would land on the surface of the agar. This helped to spread the phage. The plates were incubated overnight at 37°C. 5mls of phage buffer were added to each plate and the plates rocked from side to side for about 30 minutes. The liquid (the high titre phage lysate) was removed to a tube and the contents stored at 4°C until, required.

Batch Culture: 200mls of L-broth (1.3) and 2mls of an overnight culture of Q358, were added to each of two, 1 litre conical flasks. The flasks were incubated at 37°C on an orbital shaker until the

A_{600} of the culture was $\simeq 0.4cm^{-1}$. 5mls of phage lysate were then added to each flask and incubation continued for 4 hours.

To release phage from unlysed cells, 0.4mls of chloroform were added to each flask and shaking continued for 10 minutes. The lysates were spun in a JA-14 rotor at 12,000 x g for 10 minutes at 4°C to pellet the cell debris. The supernatants, containing the phage, were transferred to bottles containing a small quantity of DNase and RNase to degrade host DNA and RNA. The phage DNA was not degraded as it was protected from digestion by the capsid proteins.

Precipitation of the Phage: PEG 6000 and sodium chloride were added to a final concentration of 10% w/v and 2% w/v respectively and dissolved by mixing on a "Rock and Roller". The lysates were placed in buckets of ice in a cold room (4°C) overnight to precipitate the phage. The phage were then pelleted by centrifugation in a JA-10 rotor at 12,000 x g for 10 minutes at 4°C. The supernatants were decanted, the bottles returned to the rotor, rotated through 180° and spun for a few minutes. While still in the rotor any remaining liquid was removed from the bottles with a Pasteur pipette. The inside of the bottles were dried with tissue, then the pellets dissolved in a total volume of 10mls of P buffer (50mM Tris.HCl pH 7.5; 100mM NaCl; 10mM $MgSO_4$).

Banding the Phage: The phage suspension was extracted once with chloroform, to remove any remaining host cell debris, then the phases separated by centrifugation in a JA-20 rotor at 48,000 x g for 10 minutes at 4°C. The upper phase containing the phage was passed through a 0.45μm filter. The volume of the filtrate was made up to 10mls with P buffer, 7.5g of caesium chloride added, and the solids allowed to dissolve. The solution was transferred to a 13.5ml centrifuge tube, topped up with 0.75g/ml caesium chloride-P buffer solution, then spun in a Ti-70 rotor at 230,000 x g for 18 hours at 20°C. The bluish phage band, clearly visible against a black background, was removed with a needle and syringe in a volume of about 1ml. The band was transferred to a fresh 13.5ml tube,

topped up with 0,75g/ml caesium chloride-P buffer solution then centrifuged as before. The band was removed in a volume of ≈1ml and transferred to a tight dialysis (1.16) sac then dialysed against two, 1 litre changes of P buffer.

Purification of the DNA: The volume of the phage solution was measured. 0.5M EDTA pH 8.0 was added to a concentration of 20mM, 20mg/ml proteinase K added to 200µg/ml, and 10% w/v SDS added to 0.5%. The mixture was inverted a few times then incubated at 65°C for 1 hour to degrade the capsid proteins [Maniatis *et al.*, 1982]. The solution was extracted with phenol (1.13), phenol-chloroform (1.14), and chloroform. During each transfer a wide bore pipette was used to reduce DNA shearing. The DNA solution was then dialysed against five, 5 litre changes of TE (1.20) each change lasting about 12 hours. The concentration of the DNA was determined, (2.6) then the size and quality of the vector DNA checked on a 0.4% agarose gel (2.5). To ensure the correct phage DNA had been isolated, the DNA was digested with a number of restriction enzymes. The restriction pattern was then checked against that known for the vector.

2.4 Plating Cells

Plating cells were made to optimise the infection of *E. coli* by lambda EMBL3 and its derivatives. In order for a single lambda particle to form a plaque it must attach to, and infect, a live bacterium. Preparation of plating bacteria increases the number of live bacteria available for infection by phage. This method is a modification of that by Maniatis *et al.*, 1982.

A single bacterial colony was picked into 20mls of the appropriate broth and grown overnight with shaking at 37°C. Within this culture there would be a number of dead cells which can attach phage but not lead to phage propagation. Therefore, 200µls of the overnight culture were removed and used to inoculate 20mls of broth. The culture was shaken at 37°C for 4 hours by which time the cells were

entering stationary phase. Within this culture the number of dead cells from the overnight culture would be proportionately reduced. The culture was centrifuged in a JA20 rotor at 2,000 x g for 10 minutes at 4°C to pellet the cells, then the supernatant decanted. The pellet was resuspended in 8ml of 10mM $MgSO_4$, the cells placed on ice, and used for plating within the next 8 hours.

2.5 Agarose Gel Electrophoresis

Agarose gel electrophoresis is a technique which was used to resolve mixtures of different sizes of DNA fragments into bands containing DNA fragments of identical size. The resolved fragments or bands, could then be: a) sized, by comparison with bands of known size, b) identified, by Southern blot analysis or c) excised from the gel and used for cloning or making radiolabelled probes.

The mechanism behind the resolution process is as follows. DNA, under conditions of neutral pH is negatively charged, therefore DNA fragments loaded at the negative pole (the cathode) of a gel will migrate through the gel towards the positive pole (the anode). The gel consists of a microscopic network of polysaccharide molecules which slow the migration of the DNA fragments. The extent to which the fragments are retarded depends on their molecular size ie. the distance the DNA fragments migrate is inversely proportional to the $\log_{10}$ of their molecular weight [Aaij and Borst, 1972; Sharp *et al.*, 1973; Maniatis *et al.*, 1982]. However, parameters other than size are also involved in the rate of migration. These parameters, and how they were utilised in particular applications to give the greatest fragment resolution are described below.

The agarose concentration:

A DNA fragment of a given size migrates at different rates through gels containing different concentrations of agarose [Maniatis *et al.*, 1982]. This is due to the higher agarose concentration

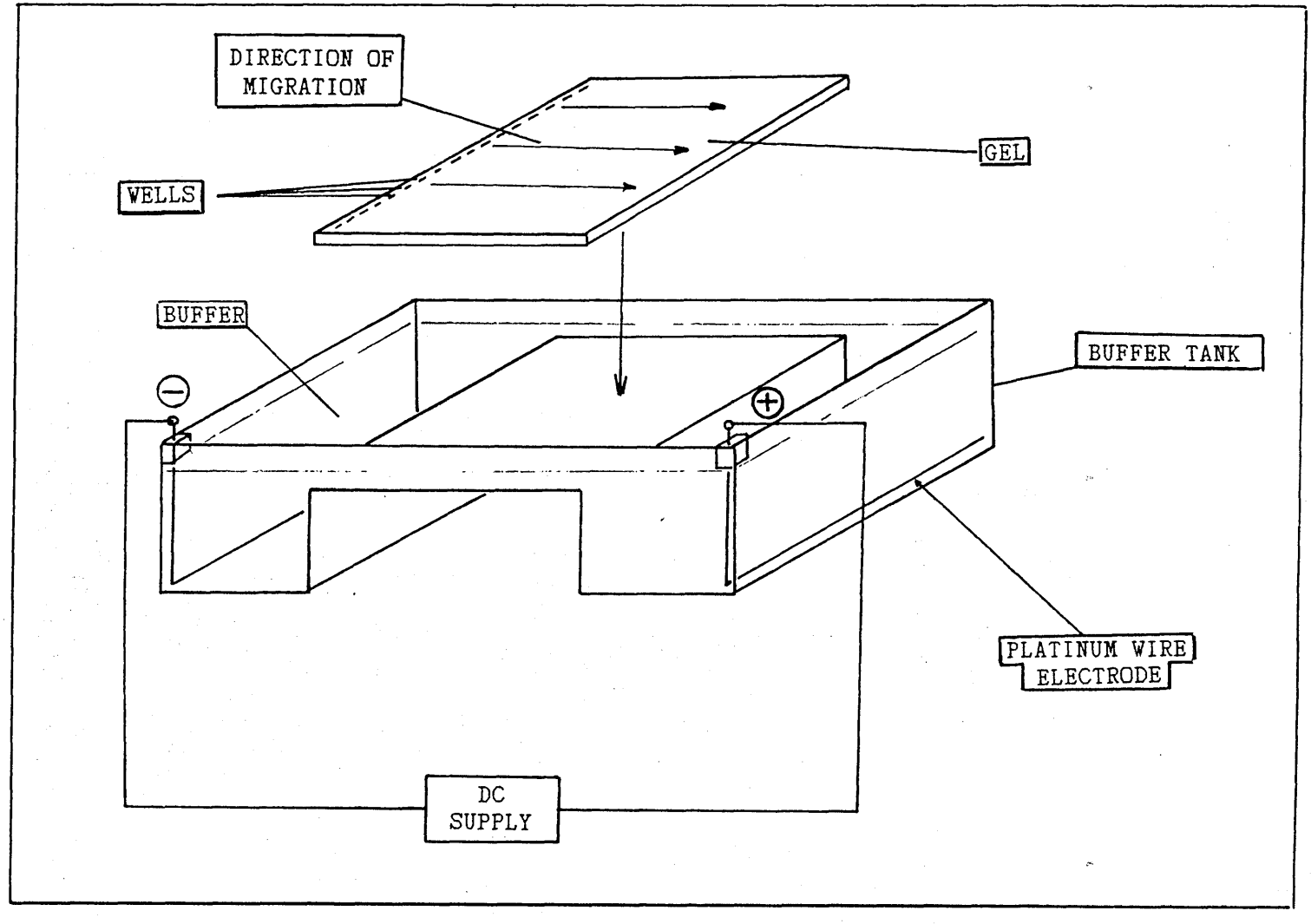


Figure 2.3 Agarose Gel Electrophoresis Apparatus.

generating a denser polysaccharide network which increases the resistance of DNA to migration within the gel. By using different concentrations of agarose DNA fragments of different sizes were resolved, see table below.

Size of DNA	Examples of Application	%age Agarose
›20kb	Sizing genomic inserts for cloning into EMBL3, checking the integrity of lambda DNA.	0.4%
≃0.6-12kb	Resolving restriction digests of EMBL3 genomic clones	0.8%
≃0.4-7kb	Comparing linear and supercoiled M13 RF/and plasmid DNA (see below). Isolation of B800-850 polypeptide encoding fragment from pLHIISB18 (2.7)	1.0%

The Conformation of the DNA:

Closed circular and linear DNAs migrate at different rates through agarose gels. This difference was used to determine if closed circular M13 RF DNA had been succesfully cut at a unique restriction site, linearising the DNA. Although methods exist for detecting differences between two DNA populations of unknown superhelicity, [Maniatis *et al.*, 1982] and determining the form of each population [Thorne, 1966]. In this case, only the difference in DNA superhelicity before and after cleavage had to be detected. The following method proved reliable. 1µl of 10mg/ml ethidium bromide solution (1.30) was added to samples of cut and uncut M13 RF DNA which were then loaded onto a 1.0% agarose minigel. The ethidium bromide induces conformational changes in the DNA which accentuates the differences in the two DNA forms by further altering their migration rates. The samples were electrophoresed for 20 minutes

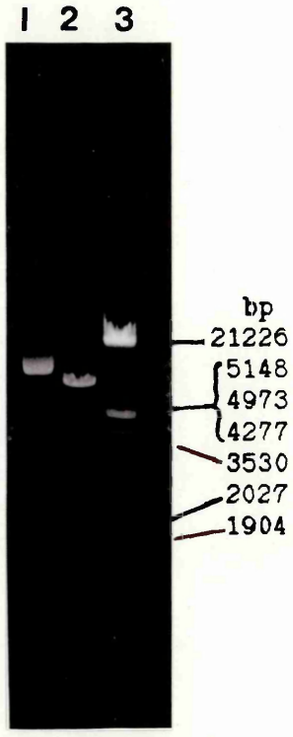


Figure 2.2

The difference in migration rate between linear and closed circular M13 mp19 is shown above.

Lane 1: 0.5µgs of mp19 DNA after digestion for 1 hour with 5 units of SstI. The reaction volume was 10µls which included 1µl of 10x SstI reaction buffer.

Lane 2: 0.5µgs of mp19 DNA after incubation for 1 hour in a 10µl "reaction" which contained 1µl of 10x SstI buffer. The SstI enzyme was absent the volume it would have occupied being made up with water.

Lane 3: λ cI857 DNA which has been digested with EcoRI+HindIII. Note that this marker DNA appears to have run anomalously. It is thought that the upper band is the 21226bp fragment, the next band a mixture of 5148, 4973 and 4277bp fragments and the band just below that the 3530bp fragment.

1µl of ethidium bromide was added to the samples in lanes 1 and 2 before loading. The gel was 1% agarose and was run at ≈15V cm^{-1}.

then the bands visuallised by transillumination. Differences in migration rate between the cut and uncut samples could be clearly seen, see figure 2.2.

The Voltage Gradient Across the Gel:

At low voltages, the rate of migration of linear DNA fragments is proportional to the voltage gradient. However, as the field strength is increased the mobility of high molecular weight fragments increases differentially. Thus the effective range of separation decreases with increase in voltage [Maniatis *et al.*, 1982]. This led to the use of two voltage gradient regimes. To resolve fragments in the 0.5-10kb size range, the gels were run at ≈2.5 Vcm^{-1} for 4 hours. For fragments larger than 10kb the gels were run at ≈1 Vcm^{-1} for 18 hours.

Minigels (described below) were always run with a ≈15 Vcm^{-1} voltage gradient.

Gel Types

Gel size: Two sizes of gel were used. Large gels had dimensions of 20 x 20cm and were used for analytical purposes eg. determining the sizes of DNA bands for restriction mapping, or for Southern blotting. All photographs displayed in this work, unless otherwise stated, are of large gels. Minigels had dimensions of 7.5 x 5cm and were run as a "quick check" to see if a piece of DNA had been cut by a particular enzyme or to detect if DNA was present in a particular sample.

All agarose gels were of the horizontal slab type, see figure 2.3.

Gel agarose: The gels were formed from one of two types of agarose, general purpose agarose and low melting point agarose.

General purpose agarose was used for all gel applications except when DNA bands were to be excised and purified from a gel. In these cases low melting point agarose was used. This type of agarose has a lower component of sulphonated agarose, an inhibitor of enzymes such as ligase, Klenow fragment, and restriction enzymes.

Gel buffers: Two types of electrophoresis buffer were also used, a borate based buffer, TBE, and an acetate based buffer, TAE (1.27 and 1.28 respectively) [Maniatis *et al.*, 1982]. TBE was found to be suitable for most applications and was used with general purpose agarose.

TAE was used whenever DNA was to be purified from a gel slice as it permits greater DNA recovery than borate based buffers. During electrophoresis the acetate in TAE is oxidised to carbonate which raises the pH of the buffer solution. Therefore, when TAE buffer was used it was either circulated between buffer tanks by a peristaltic pump, or changed regularly during electrophoresis

Casting, Pouring and Running Gels

Large gels: These were cast onto a glass plate the sides of which had been taped to form a shallow tray. The glass plate was placed on a level bench then a comb placed ≈1cm from one end of the tray and ≈0.5mm from the glass.

A suitable volume (usually ≈200mls) of agarose was prepared by pouring the appropriate amount of agarose powder into a conical flask and adding buffer. The flask and contents were weighed, then placed in a microwave oven until all the solids had dissolved. A conical flask proved the most suitable vessel as it prevented excessive evaporation of water during heating, and a skin forming on the surface during cooling. The solution was allowed to cool to ≈60°C. The flask was then reweighed, and if necessary brought back to the original weight by the addition of water. The gel was then poured. When the gel had set, the comb and tape were removed, the gel and glass plate placed in the gel tank and covered to a depth of ≈1cm with buffer.

Minigels: These were prepared in essentially the same way as described above, however, the gels were cast in a casting tray. Approximately 30mls of agarose were used for each gel.

Occasionally large volumes of relatively dilute sample needed to be loaded. In these cases, the gel was placed in the tank, sufficient buffer added to just bring the level up to, but not over the top of the gel. The samples were loaded into the dry wells, then the current switched on for ≈15 minutes. This allowed time for the migration of the DNA into the gel before it was covered with buffer. The samples were electrophoresed using the conditions required for the particular sample, then the gel stained in ≈0.5μg/ml ethidium bromide (1.30) solution for 20 minutes. The gel was then transferred to a medium wavelength transilluminator (312nm) and the bands visualised by UV light [Sharp *et al.*, 1973].

Photographing gels:

Gels were photographed onto Polaroid type 667 film (ISO 3000) through a Kodak 22A Wratten filter using a Polaroid camera system [Maniatis *et al.*, 1982].

2.6 Measuring DNA Concentrations

DNA concentrations were measured using a spectrophotometer and/or ethidium bromide plates.

Specrophotometric method: This method was used mainly with preparations of genomic DNA when there was sufficient material available to permit accurate measurement. 10μls of DNA solution were added to 990μls of TE (1.20) in a 1.5ml Eppendorf tube. The mixture was vortexed briefly then added to a quartz $1cm^{-1}$ pathlength UV cuvette (1.5ml volume). The absorbance was determined at wavelengths of 260 and 280nm. The DNA concentration was determined using the following equation:

$$\frac{(\text{Absorbance at 260nm} - \text{Absorbance at 280nm}) \times 2}{1000\mu\text{ls}/10\mu\text{ls ie. vol. of DNA soln.}} \times 50 = \text{DNA conc. } \mu g/ml$$

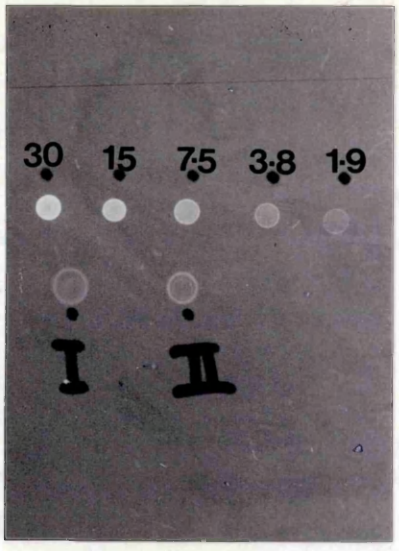


Figure 2.4

Ethidium bromide plate showing DNA concentration standards of 1.9-30ngs/µl. Samples I and II have have concentrations of between 3.8 and 7.5ngs/µl as judged by the intensity of their spots.

Ethidium bromide plate method: This method was used to determine the DNA concentration of samples where there was very little DNA available for analysis. Ethidium bromide plates were prepared by adding 3µls of 10mg/ml ethidium bromide solution (1.30) to 100mls of melted 1% agarose solution. The solution was poured into Petri dishes to a depth of ≈1mm. If the agarose film was poured thicker, excessive background light obscured low concentrations of sample. Once set the plates were either used immediately or wrapped in foil or stored in a light tight container at room temperature.

Lambda cI857 DNA, of known concentration, was diluted in TE to give a range of concentration standards ranging from ≈1 - 500ng/µl. A line of dots, was made with a fine marker pen on the bottom of the ethidium plate. 1µl of a concentration standard was spotted (without stabbing the agarose) under each dot. Under a separate dot was spotted 0.5 or 1µl of the DNA sample(s) to be measured. The plate, without lid, was placed, agarose surface up, under a 60 watt bench lamp until the liquid in all the spots had evaporated. The plate was then placed agarose surface down on a medium wavelength (312nm) transilluminator. The intensity of the transmitted light from the sample spot could then be compared to those of the concentration standards, and from this an estimate of the DNA concentration of the sample made. An example of an ethidium bromide plate is shown in figure 2.4.

2.7 PLASMID pLHIISB18

The library was screened with a radiolabelled heterologous probe encoding the B800-850 alpha and beta polypeptide genes from *Rhodobacter sphaeroides* strain 2.4.1. These genes were isolated by Patricia J. Kiley and Samuel Kaplan and donated by Samuel Kaplan for this work in the form of plasmid pLHIISB18 [Kiley and Kaplan, 1987]. The construction of this plasmid is summarised below.

The B800-850 beta polypeptide amino acid sequence was used to construct a synthetic deoxyoligonucleotide probe. *Rhodobacter sphaeroides* genomic DNA was isolated then digested with PstI. The fragments were run on an agarose gel and fractions of 2-4kb electroeluted onto DEAE paper. The DNA was purified and fractions hybridising to the deoxynucleotide probe identified by Southern blot analysis. The positivley hybridising fractions were ligated into the PstI site of pUC19 and transformed into *E. coli* strain JM83. Recombinant plasmids containing the B800-850 structural genes were identified by Southern blot analysis of PstI digested plasmid DNA derived from lysates of small scale cultures.

A 730bp StuI-BamHI fragment containing the B800-850 structural genes was excised from one of these plasmids. The fragment was subcloned into the HincII and BamHI sites of pUC18 giving rise to the recombinant pLHIISB18. The plasmid was transformed into *E. coli* strain JM83 and maintained using 50μg/ml ampicillin. The restriction map of the plasmid and nucleotide sequence of the insert are presented in figures 2.5 and 2.6

Use of the pLHIISB18 Insert as a Radiolabelled Probe

The plasmid insert was used as a probe for the genes encoding the *R. acidophila* light harvesting polypeptides. The probe was used against *R. acidophila* genomic, EMBL3 recombinant, and M13 recombinant DNA.

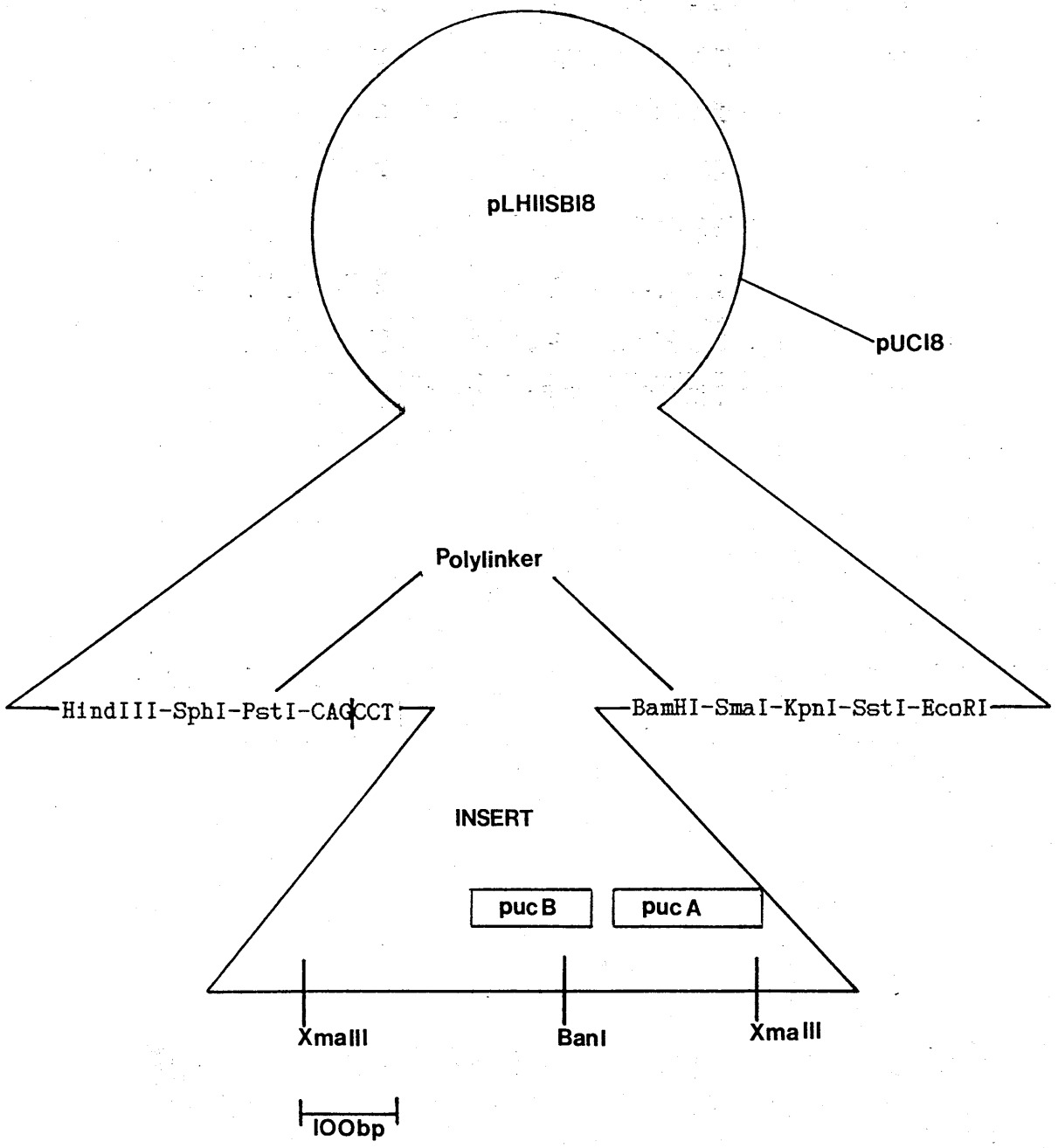


Figure 2.5

Schematic diagram of plasmid pLHIISB18. The insert containing the genes, *puc*BA, encoding the light harvesting B800-850 polypeptides of *Rb. sphaeroides* was subcloned into the HincII (CAG↓CCT) and BamHI sites of the pUC18 polylinker. The sequence of this insert is presented in Figure 2.6.

Note the scale refers only to the insert.

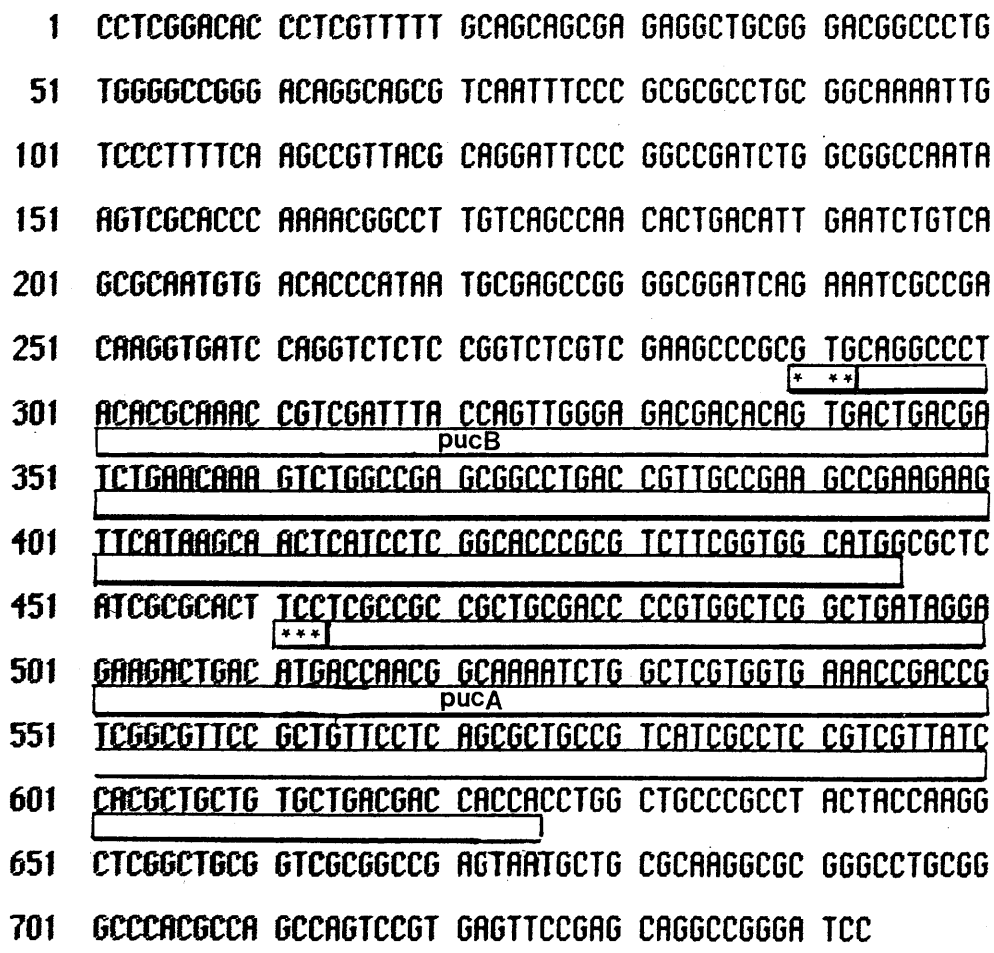


Figure 2.6

Shown above is the nucleotide sequence of the insert from plasmid pLHIISB18. The genes encoding the beta and alpha B800-850 light harvesting polypeptides are shown blocked as *pucB* and *pucA* respectively. The nucleotides encoding the initiating methionine are shown as *** .

When used as a probe to genomic or EMBL3 recombinant DNA the plasmid was digested with EcoRI and HindIII. However, these enzymes excised the insert with small flanking pieces of pUC18 polylinker. It was shown (data not presented) that under the conditions of filter hybridisation and washing described (2.13), that radiolabelled pUC18 did not hybridise to *R. acidophila* genomic DNA. It was also shown that the HindIII-EcoRI fragment from pLHIISB18 did not hybridise to non recombinant EMBL3 DNA. This suggested that the pLHIISB18 HindIII-EcoRI fragment could be used for probing both genomic and genomic-recombinant DNA without "false positive" signal occuring due to cross hybridisation.

Sequencing required the construction of M13 mp19 recombinants. Although not tested experimentally, it seemed probable that the pUC18 polylinker sequences present in the HindIII-EcoRI fragment of pLHIISB18 may hybridise to the mp19 polylinker sequence. By cutting the plasmid with BamHI and PstI a 600bp fragment was released free of most pUC18 sequences. This fragment was radiolabelled and used as a probe to M13 recombinants.

To avoid continual use of "the HindIII-EcoRI fragment of pLHIISB18" or "the BamHI-PstI fragment of pLHIISB18" when refering to radiolabelled probes, the name pLHIISB18 alone has been used in the text to describe the fragments.

2.8 Isolation of Plasmid pLHIISB18 DNA

Culture of the the cells: A loopful of *E. coli* strain JM83, containing plasmid pLHIISB18, was streaked onto an L-agar (1.4) plate containing 50µg/ml ampicillin. The plate was incubated overnight at 37°C.

400mls of L-broth (1.3) containing 50µg/ml of ampicillin, were added to each of two, 2 litre conical flasks. A single, well isolated

colony from the overnight plate was then picked into the broth. The flasks were incubated at 37°C on an orbital shaker for 18 hours. The cells were then pelleted by centrifugation in a JA-10 rotor at 12,000 x g for 10 minutes at 4°C.

Isolation of plasmid and M13 RF DNA.

This protocol is a modified version of the alkaline lysis method [Birnboim and Doly, 1979] as described in Maniatis *et al.*, 1982. The same basic method was also used for the isolation of M13 RF DNA (2.16). Differences between the protocols are outlined in the text.

The pelleted cells were resuspended in 20mls of lysis buffer (50mM glucose; 25mM Tris.Cl pH 8.0; 10mM EDTA pH 8.0). 100mgs of lysozyme (Grade I from chicken egg white) were dissolved in 2mls of lysis buffer which were then added to the bacterial suspension and mixed in by vortexing. The suspension was left to stand at room temperature for 10 minutes, during which time the cell walls were degraded by the lysozyme. To lyse the cells and denature the DNA, 40mls of 0.2M NaOH; 1% w/v SDS were then mixed in by gentle inversion. At this stage the suspension changed from turbid to clear. The lysate was left on ice for 10 minutes, then to renature the DNA, 20mls of 5M potassium acetate pH 4.8 were added and mixed in by several sharp inversions of the tube. The lysate was returned to ice for 10 minutes then the cell debris and tangled mass of chromosomal DNA pelleted by centrifugation in a JA-10 rotor at 16,000 x g for 30 minutes at 4°C. The supernatant, containing the plasmid, was poured through a Kleenex tissue to remove pieces of loose pellet and SDS solids. To precipitate the DNA, 50mls of isopropanol were added to the supernatant, the solution vortexed, then left to stand at room temperature for 15 minutes. The DNA was pelleted by centrifugation in a JA-20 rotor at 48,000 x g for 30 minutes at 20°C. The supernatant was poured off, the pellet washed with 80% ethanol, then dried in a vacuum desiccator. The pellets were dissolved in a total volume of 10ml of TE (1.20) then 10g of caesium chloride added and dissolved by mixing on a "Rock and

Roller". 1ml of 10mg/ml ethidium bromide solution (1.30) was added then tube covered in foil and left at room temperature for 1hr. Using a needle and syringe, and avoiding the protein aggregates which had accumulated, the solution was transferred to the appropraiate centrifuge tube, see below.

Plasmid and M13 RF DNA were prepared differently at this stage:
Plasmid: The DNA solution was transferred to a 39ml centrifuge tube, the volume made up with ethidium bromide/caesium chloride solution* then spun in a VTi-50 rotor at 167,000 x g for 18 hours at 20°C. The band, usually visible by daylight, was removed in a volume of ≈2ml. The DNA was transferred to a 5.1ml centrifuge tube, topped up with ethidium bromide/caesium chloride solution* then re-banded by centrifugation in a VTi-65 rotor at 194,000 x g for 18 hours at 20oC. The band was removed in a volume of ≈1ml and transferred to a 5ml test-tube.

M13 RF: The DNA solution was transferred to a 13.5ml centrifuge tube, topped up with ethidium bromide/caesium chloride solution* and spun in a Ti-70 rotor at 230,000 x g for 18 hours at 20°C. The band was visualised with a long wavelength (360nm) transilluminator and removed in ≈2ml. The DNA was transferred to a VTi-65 tube, topped up with ethidium bromide/caesiumchloride solution* then re-banded by centrifugation in a VTi-65 rotor at 194,000 x g for 18 hours at 20°C. The band was removed in a volume of ≈1ml to a 5ml glass tube.
*Ethidium bromide/caesium chloride solution: 1ml of 10mgs/ml ethidium bromide solution (1.30) was added to 10mls of TE (1.20) and 10g of ceasium chloride.

Plasmid and M13 RF DNA were then purified by the same method. To extract the ethidium bromide, 1ml of water saturated butanol (1.15) was added to the DNA solution which was then vortexed and allowed to settle for 5 minutes. The butanol phase (the upper or most pink) was removed and another 1ml of water saturated butanol added to the DNA solution. This process was repeated, usually four times, until all the ethidium bromide had been extracted.

To remove the caesium chloride, the DNA solution was transferred to a tight dialysis sac (1.16) and dialysed against five, 1 litre changes of TE. Each change lasted about 2 hours. The DNA concentration was measured on an ethidium plate (2.6), then the DNA digested and the restriction pattern checked on a 1% agarose gel (2.5).

2.9 Genecleaning

Genecleaning was used to purify DNA from gel slices. The Genecleaning reagents (based on those of Vogelstein and Gillespie, 1979) were purchased as a kit and used as follows.

The DNA sample containing the restriction fragment of interest was run on a low melting point agarose gel made in TAE buffer (1.28), (2.5). The appropriate band was excised, placed in a weighed Eppendorf vial, then the weight of the gel slice determined. 2.5 volumes of 6M sodium iodide solution were added to the gel slice for each unit of gel slice weight. The gel slice was melted at 55°C for 5 minutes to give a diluted DNA solution, to which was then added, 5µls of Glassmilk (a silica matrix suspension) and the mixture vortexed briefly. The suspension was placed on ice for 5 minutes to allow the DNA to bind to the silica. The silica, with bound DNA. was then pelleted by brief microcentrifugation (≈5 seconds) and the supernatant decanted. The pellet was resuspended in 500µls of a high salt solution (New Wash), to wash the DNA/Glassmilk. These were then pelleted as before. The washing procedure was repeated twice more. The molar amounts of the New Wash ingredients are not given in the kit instructions. However, the instructions state "The New Wash contains a mixture of ethanol/NaCl/Tris/EDTA....with a mixture of Tris acid-Tris base such that the pH is within the range 7-8.5 depending on temperature". The last drops of New Wash were removed from the pellet with a glass capillary then the pellet resuspended in 15µls of water. The suspension was heated to 55°C

for 3 minutes, to release the DNA, then spun briefly to pellet the silica. The supernatant, containing the DNA, was transferred to a clean tube and the concentration of the solution determined using an ethidium plate (2.6).

2.10 Preparation of ^{32}P Radiolabelled Probes

Radiolabelled probes were prepared according to the "oligo-labelling" or random priming method of Feinberg and Vogelstein, [Feinberg and Vogelstein, 1983] which involves the following. DNA to be labelled is denatured by boiling to form a single stranded template. The template solution is added to a "reaction mix" containing random hexamer oligo-deoxyribonucleotides, dATP, dGTP and dTTP and to which is then added [^{32}P]dCTP and the large fragment of DNA polymerase I (Klenow fragment).

The reaction begins by the random hexamers binding to complementary regions of the template. These act as primers for the synthesis of a complementary radiolabelled second strand by the Klenow fragment. The labelled DNA is separated from unincorporated radionucleotides by molecular sieve chromatography then denatured and used as a probe. Probes prepared in this way regularly had specific activities greater than 10^8cpm/μg of DNA.

Three different types of DNA were radiolabelled by this method: *R. acidophila* genomic DNA; the stuffer fragment from lambda EMBL3; and the insert from plasmid pLHIISB18. Before labelling, the DNA from each of these was prepared as follows:

R. acidophila genomic DNA: The purified DNA (2.2) was diluted in TE (1.20) to a concentration of 2ngs/μl. 5μls (10ng) of the diluted DNA solution were used in each labelling reaction.

Lambda EMBL3 Stuffer Fragment: 5µgs of EMBL3 DNA were cut with SalI and the digestion products run on a 1% low melting point agarose gel in TAE (2.5), (1.28) The gel was stained with ethidium bromide solution (1.30) and the stuffer fragment excised from the gel. The DNA was Genecleaned (2.9) out of the agarose and the concentration determined on an ethidium plate (2.6). 10ngs of DNA were used in each labelling reaction.

Plasmid pLHIISB18: 20-30µgs of DNA were cut with either: EcoRI and HindIII when probing lambda clones and genomic DNA, or PstI and BamHI when probing M13 clones, see (2.7) for explanation. The digested DNA was run on a 1% low melting point agarose gel in TAE. The gel was stained with ethidium bromide solution and the band excised with a scalpel blade. The DNA was Genecleaned out of the agarose and 0.5µls spotted onto an ethidium plate to determine the DNA concentration. 10ngs of DNA were used in each labelling reaction.

The labelling reactions were carried out as follows: To denature the DNA, the solution was boiled for 10 minutes then immediately placed on ice. The following were added to a 0.5µl Eppendorf vial in this order:

32µls of water
10µls of Reaction Mix, see below.
2µls of 10mg/ml nuclease free BSA.
5µls of denatured DNA solution (2ngs/µl).
3µls of 800µCi/mmol ^{32}P-α-dCTP.
1µl of Klenow fragment (1unit/µl).

The Reaction Mix was composed of the following four solutions:

Solution A:
1ml of solution D.
18µls of β-mercaptoethanol (14.4M).
5µls of 100mM dATP.
5µls of 100mM dTTP.
5µls of 100mM dGTP.

Solution B:
2M HEPES pH 6.6 (pH with 4M NaOH).

Solution C: Hexadeoxyribonucleotides; 90 A_{260} units/ml in TE.

Solution D: 1.25M Tris.HCl pH 8.0; 0.125M $MgCl_2$.

These were mixed together in the following proportions: A=200µls, B=500µls and C=300µls to give 1ml of Reaction Mix. When not in use the Reaction Mix was stored at -20°C.
The reaction was incubated at room temperature overnight then a small amount of Orange G dye added. This acted as a marker of unincorporated nucleotides during the next step.

The labelled DNA was resolved from unincorporated radionucleotides by passage down a G-50 Sephadex column. The column was prepared by packing a small amount of siliconised glass wool into the bottom of a 1ml syringe and adding Sephadex G-50 (1.18) up to the 1ml mark. The column was washed through with several volumes of TE (1.20) then the surface allowed to dry briefly. The orange reaction mixture was loaded onto the column and the surface again allowed to dry briefly before being topped up with TE. 3 drop aliquots were collected from the column. The labelled DNA was usually collected in aliquots 4-6, the unincorporated radionucleotides, which ran with the Orange-G, in aliquots 9-12.

Immediately before use the probe was denatured by boiling for 5 minutes then immediately placed on ice for 5 minutes. The probe was then added to ≈4ml of pre-hybridization solution (1.25). The mixture ie. hybridization solution (1.26) was then added to the filters to be probed (2.13).

2.11 Screening Bacteriophage plaques by *In Situ* Hybridisation (Preparation of Filters)

This method was used to screen both EMBL3 and M13 recombinant plaques. The method is essentially that of Mason and Williams, 1985 and Maniatis et al., 1982

The phage were plated, grown overnight at 37°C, then placed at 4°C for 1 hour to allow the agar surface to harden. A nitrocellulose filter, cut slightly smaller than the plate, was laid onto the surface and left in position for 1 minute. During this time non-symmetrical orientation marks were punched through the filter into the agar with an India ink coated needle. The filter was removed from the plate, and placed DNA, or lawn side, upwards on a piece of 3MM paper moistened with denaturing solution; 0.5M NaOH; 1.5M NaCl. This treatment removes the capsid proteins and, in EMBL3, denatures the DNA. After 1 minute the filter was removed to a piece of 3MM paper moistened with neutralising solution; 1M Tris.HCl; 1.5MNaCl, where it remained for 5 minutes. The filter was then moved to a piece of 3MM paper soaked in 4x SET (1.21) for 5 minutes, before being left to air dry on a clean piece of 3MM paper. Another filter was laid onto the same plate and orientation marks punched corresponding to those made previously. After 3 minutes the filter was removed from the plate and denatured, neutralised and washed like the first. Duplicate filters were taken from each plate as they allowed positively hybridising plaques to be discerned from random background spots. The dried filters were placed DNA side down on a clean, medium wavelength (312nm) transilluminator and irradiated with UV light (312nm) for 4 minutes. The filters were then probed immediately or stored in a sealed polythene bag at -20°C until required.

2.12 Southern Blotting

DNA was transferred from gels to nylon or nitrocellulose membranes using slight modifications of the Southern blotting technique described by Southern, 1975.

After the gel containing the DNA of interest had been run, stained and photographed, the wells and excess gel material were trimmed away with a scalpel blade. The gel was transferred to a glass baking dish and the DNA partially hydrolysed by soaking in several volumes

of 0.25M HCl for 20 minutes. This treatment facilitated an even transfer of different sized DNA bands, as high molecular weight DNAs were cleaved into smaller, more readily transferred fragments [Wahl *et al.*, 1979]. The treatment the gel received next depended on the nature of the DNA:

EMBL3 and M13 Recombinant DNA: The DNA was denatured by soaking the gel in several volumes of 0.5M NaOH; 1.5M NaCl for 30 minutes. The DNA was transferred to a nylon filter using 0.25M NaOH; 1.5M NaCl transfer buffer.

R. acidophila Genomic DNA: The DNA was denatured by soaking the gel in several volumes of 0.5M NaOH; 1.5M NaCl for 30 minutes. The denaturing solution was then decanted and the gel neutralised by soaking in several volumes of 1M Tris.HCl pH 8.0; 1.5M NaCl for 1 hour. The DNA was transferred to a pre-wetted nitrocellulose filter using 10x SSC (1.22) transfer buffer. The nitrocellulose was pre-wetted by soaking for 5 minutes in distilled water then for 5 minutes in 2x SSC.

The two types of DNA were treated differently because empirically the conditions described were found to favour the highest signal to noise ratio. In both cases the DNA transfer was allowed to proceed for ≈18 hours. The gel-filter "bilayer" was then placed filter side down on a piece of 3MM paper and the positions of the wells marked with a ball point pen. The filter was peeleed away from the dehydrated gel and placed in a shallow bath of 4x SET (1.21) for 5 minutes. Meanwhile the dehydrated gel was restained then examined on a transilluminator to determine if the DNA transfer was complete.

Nitrocellulose filters were blotted on 3MM paper then placed DNA side down on a medium wavelength transilluminator (312nm). Irradiation for 3 minutes was used to fix the DNA to the filter. DNA transferred to nylon filters with alkali did not require this treatment.

The filters were then either probed with ^{32}P labelled DNA probes (2.10),(2.13), or stored in a sealed polythene bag at -20°C until required.

2.13 Screening Filters with Radiolabelled Probes

All filters shown or described in this work were screened using the following protocol.

The filters with fixed DNA were washed in each of the following for a few minutes; water, 0.1% v/v Triton X-100, then 4x SET (1.21). The damp filters were stacked between two polythene sheets and three of the four edges of the polythene heat sealed to form an open bag. Sufficient prehybridization solution (1.25) was poured into the bag to allow the filters to float about freely. The prehybridisation solution, contained Denhardt's solution (1.23) and denatured herring sperm DNA (1.24), to decrease background by reducing the non specific binding of the probe to the filters. Air bubbles were removed from the bag which was then heat sealed at its fourth side. The bag was placed in a sandwich box half filled with water which in turn was placed in a 65°C shaking water bath.

After ≃4 hours the bag was removed, one corner cut off, and the prehybridization solution decanted until 1-2ml remained. The denatured probe, (10ngs of random primed DNA (2.10)) was added to 4mls of fresh prehybridization solution and the mixture transferred to the bag. Bubbles were removed, the corner sealed, then the bag returned to the water bath.

After 18 hours the bag was removed from the bath and the hybridisation solution poured off. The filters were immediately removed from the bag and placed in a room temperature solution of 1% w/v SDS where they remained for 5 minutes. The filters were then washed for 20 minutes in 500mls of: 3xSET; 0.1% w/v SDS; 0.05% w/v

tetra-sodium pyrophosphate in a shaking water bath at 50°C. The wash was then repeated twice more. The filters were transferred to a bath of 4x SET at room temperature for 10 minutes, blotted on a piece of 3MM paper then covered in Saran Wrap. The filters and a piece of X-Omat S film were sandwiched between two intensifying screens, placed in a cassette, and left overnight at -70°C. The film was developed automatically using a Kodak X-OMAT processor.

2.14 M13 Vector-Host Biology

The M13 vector, mp19, was used to generate the template DNAs required for the sequencing reactions. Wild type M13 is a male-specific filamentous bacteriophage of *E. coli* [Dehardt *et al.*, 1978]. The virus particle, which contains a single stranded circular DNA genome, attaches to the F pilus of the host for infection. As with other single stranded phage only the (+) strand DNA is encapsidated into viral particles. Following infection, the M13 (+) strand DNA serves as a template for the synthesis of complementary (-) strand DNA. The double stranded form of the viral DNA is referred to as the replicative form (RF). The RF is amplified to approximately 200 copies per cell. Following amplification, the (-) strand DNA is no longer replicated; however, the (+) strand DNA continues to be synthesised. As the new (+) strand DNA is synthesised it displaces the previous (+) strand DNA. The disassociated (+) strand is packaged in the viral coat proteins to generate the phage particle. The (-) strand contiues to serve as a template for new (+) strand DNA synthesis. After the phage is assembled it is extruded out of the host cell through a non-lytic mechanism.

Several variants of the M13 phage have been created specifically for the purpose of generating templates for dideoxy sequencing. These are the M13 mp vectors [Yanisch-Perron *et al.*, 1985; Messing, 1983;

Messing and Viera, 1982; Norrander *et al.*, 1983]. The vector used during the course of this work was mp19.

To perform dideoxy sequencing analysis the DNA of interest was subcloned into the RF form of the vector. The recombinant DNA was then used to transform competent *E. coli* cells. The recombinants containing the cloned DNA fragment of interest were selected by *in situ* hybridisation using a ^{32}P labelled PstI-BamHI fragment from pLHIISB18. The purified, positively hybridising plaques were used to propagate recombinant phage, the DNA from which was used as the dideoxy sequencing template.

Modification of M13 and its host has permitted a simple colourimetric analysis to be used to identify recombinant phage. This analysis was a useful marker in determining the efficiency of the RF digestion, ligation and transformation reactions. The colourimetric system employs β-galactosidase, an enzyme in the lac operon. The tetrameric enzyme will convert X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) to a blue chromophore, thus allowing visual detection of β-galactosidase activity. Mutations in the 5'-region of the gene coding for B-galactosidase (lac Z gene) produce subunits that are not capable of assembly and do not demonstrate β-galactosidase activity. However, if these subunits associate with a short polypeptide corresponding to the N-terminal portion of the enzyme, β-galactosidase activity is restored. This association is referred to as α-complementation [Messing *et al.*, 1977, Bethesda Research Laboratories, 1980].

M13 DNA has been modified by insertion of a portion of the *E. coli* lac operon into an intergenic region of M13. This vector DNA sequence (M13 mp19) could produce a peptide capable of α-complementation. The lac operon has been deleted from the host genome and the mutated form of the gene inserted into the F episome. Upon infection of the cells with M13 mp19 these cells produce active β-galactosidase (via α-complementation) and consequently blue plaques are obtained in the presence of the indicator X-gal.

Introduction of unique "in frame" restriction sites within the lac Z gene has permitted α-complementation to continue.. However, insertion of most DNA fragments into these sites results in a peptide incapable of α-complementation and therefore colourless or "white" plaques are produced.

In summary, the modified M13 host does not demonstrate β-galactosidase activity. Upon infection of the host by an M13mp vector containing a subcloned DNA fragment a white plaque is obtained in the presence of the indicator. However, if the M13mp virus does not contain a subcloned fragment, α-complementation will occur and the resulting plaque will be blue.

Both the *E. coli* host (in this work strain JM101) and its episome have been modified. It is essential for M13 infection that the host strain carries the F episome as it encodes genes involved in F pilus formation. The genes responsible for proline biosynthesis have been deleted from the *E. coli* genome and inserted into the F episome. Maintainance of the host on minimal media therefore forces the retention of the F episome.

As mentioned above the lac Z gene has been deleted from the genome of the host and a mutant form of the gene (lac Z delta M15) capable of α-complementation, inserted into the F episome. In addition, the F episome also carries the lac IP mutation, which results in overproduction of the lac repressor. This allows the expression of the lac operon to be regulated by the introduction of the isopropylthio-β-D-galactoside (IPTG), a lac operon inducer.

2.15 Growth of *E. coli* Strain JM101

E. coli strain JM101 was the host for the M13 vectors used in this work. When single colonies were required a loopful of the glycerol stock was streaked onto a Minimal H-agar (1.11) plate and incubated

overnight at 37°C. The plate was stored at 4°C and colonies removed when needed. The plate was kept for a maximum of 2 weeks then replaced [Bethesda Research Laboratories, 1980].

The cells were used in two types of culture:
Liquid cultures: The cells were grown in either 2xYT (1.9) or L-broth (1.3) at 37°C with shaking.
Plates for the Isolation of M13 plaques: The cells were added to 0.6% Top Agar (1.12) and plated onto a bottom layer of H-Agar (1.10).

2.16 Isolation of M13 Replicative Form (RF) DNA.

Restriction fragments to be sequenced were subcloned into the double stranded RF DNA of M13 mp19. The RF DNA was prepared as follows. 50ng of M13 mp19 RF DNA was used to transform 0.3ml of JM101 competent cells (2.17).

The following morning a well isolated blue M13 plaque was picked into 20mls of L-broth (1.3) and the culture shaken at 37°C for 6 hours. The cells were pelleted by centrifugation in a JA-20 rotor at 12,000 x g for 10 minutes at 4°C and the supernatant (cleared phage lysate) stored at 4°C until required.

A single colony of JM101 was added to 5mls of L-broth then the culture shaken overnight at 37°C. 400mls of L-broth, and 4mls of the overnight culture were poured into each of two, 2 litre conical flasks. The flasks were shaken at 37°C until the A_{600} of the culture $\simeq 0.4cm^{-1}$. This took about 1-1.5 hours, then 1/2 of the cleared phage lysate was added to each flask. After 4 hours the cells were pelleted by centrifugation in a JA-10 rotor at 12,000 x g for 10 minutes at 4°C.

The DNA was isolated using the method described in (2.8).

2.17 Transformation of *E. coli* with M13 RF DNA

"Wild type" and recombinant M13 DNAs were introduced into JM101 by transformation of calcium chloride treated competent cells as described by Cohen *et al.*, 1974 and Dagert and Ehrlich, 1979.

Preparation of JM101 Competent Cells:

A well isolated colony of JM101 was picked into 20mls of 2xYT (1.9) medium and grown overnight with shaking at 37°C. For each transformation reaction (to be performed) 3mls of 2xYT were placed in a single conical flask and inoculated with 30µls of the JM101 overnight culture. The culture was shaken at 37°C until the A_{550}=0.4-0.5cm^{-1}, usually about 1.5-2 hours.

The flask containing the culture, and equipment such as centrifuge tubes, JA-20 rotor and 50mM $CaCl_2$ were cooled on ice to 0°C. After 20 minutes the cells were pelleted by centrifugation in a JA-20 rotor at 2,000 x g for 10 minutes at 4°C. The supernatants were then decanted. The tubes containing the cell pellets were kept on ice and the cells gently resuspended in 1/2 the original volume of 50mM $CaCl_2$ using a wide bore pipette. After 30 minutes the cells were pelleted as before, the supernatant decanted and the cells gently resuspended in 1/10th of the original volume of 50mM $CaCl_2$. The competent cells were kept on ice and used within the next 4 hours.

Preparation of lawn cells:

200µls of the overnight culture (used for the preparation of competent cells) was used to inoculate 20mls of 2xYT. The inoculum was shaken at 37°C until the competent cells were used for plating ie. ≃3-7 hours.

Transformation

5ml glass tubes were placed on ice and into each was added 0.3mls of JM101 competent cells. An appropriate volume of RF solution, (containing 10-50ngs of DNA) was added to each tube. The contents were mixed and the tube returned to ice for 20 minutes. The cells were heat shocked, to stimulate the uptake of DNA, by transferring the tubes to a 42°C water bath for 2 minutes. 0.4 mls of lawn cells were added to the tube followed by 3mls of warm 0.6% Top Agar (1.12). The contents were then poured onto a warmed H-Agar plate (1.10) then incubated overnight at 37°C. Recombinant and non-recombinant M13 plaques, visualised by virtue of the "blue-white selection system", were then picked and/or counted depending on the experimental requirements.

2.18 Preparation of Single Stranded Templates.

3mls of 0.6% Top Agar (1.12) were added to 0.4mls of JM101 lawn cells (2.17) and the suspension poured onto an H-Agar (1.10) plate. A loopful of the M13 recombinant lysate was streaked onto the 0.6% Top Agar and the plate incubated overnight at 37°C.

An overnight culture, (required for the following morning) was prepared by inoculating 5mls of 2xYT (1.9) with a colony of JM101. The culture was shaken overnight at 37°C.

200µls of the overnight culture and a well isolated white recombinant plaque were added to 20mls of 2xYT. The tube was shaken for 4 hours at 37°C to allow growth of the phage. The culture was transferred to a 50ml Falcon tube and the cells pelleted by centrifugation in a Jouan swing out rotor at 4,000 x g for 20 minutes at room temperature. The supernatant was poured into a clean Falcon tube and the centrifugation repeated as before. Without touching the sides or bottom of the tube, 15mls of cleared

lysate were removed with a pipette and transfered to a clean Falcon tube containing 3mls of 20% w/v PEG 8000; 2.5M NaCl; 5mM EDTA pH8.0. The contents were mixed by vortexing, then left to stand at room temperature for 10 minutes to allow the phage to precipitate. The phage were then pelleted by centrifugation in a Jouan swing out rotor at 4,000 x g for 10 minutes at room temperature. The supernatant was decanted and the tubes respun for 5 minutes to bring droplets of the precipitating solution into the conical base of the tube. The phage pellet was resuspended in the remnants of the precipitating solution then transferred to a clean 1.5ml Eppendorf vial. The tube was microfuged for 5 minutes to re-pellet the phage and the supernatant removed. The vial was respun for a few seconds, more supernatant removed, then respun again and the last vestage of supernatant removed with a drawn out Pasteur pipette.

750µls of TE (1.20) were added to the pellet and the vial left to stand for 30 minutes before being vortexed vigorously for ≈15 seconds. 750µls of phenol-chloroform (1.14) were added, the mixture vortexed well, then microfuged for 5 minutes. 700µls of the aqueous phase were removed to a clean vial, 700µls of phenol-chloroform added then the mixture vortexed, then microfuged for 5 minutes. 600µls of the aqueous phase were removed, 600µls of chloroform added then the mixture vortexed and microfuged for 5 minutes. 450µls of the aqueous phase were removed and to it added 45µls of 3M sodium acetate pH 5.2 (1.17) and 1ml of 90% ethanol. The mixture was vortexed briefly, placed on ice for 1 hour, then microfuged for 15 minutes. The supernatant was discarded and the invisible DNA "pellet" washed with 80% ethanol. All liquid was removed from the vial with a drawn out Pasteur pipette then 20µls of TE (1.20) added to the pellet. The pellet was allowed to dissolve for 30 minutes. A 1µl aliquot of the template solution was removed and added to 3µls of TE and 1µl of loading dye (1.29). The sample was run on a 1% agarose minigel (2.5), then checked for integrity, approximate quantity and contamination with JM101 genomic DNA, see figure 2.7.

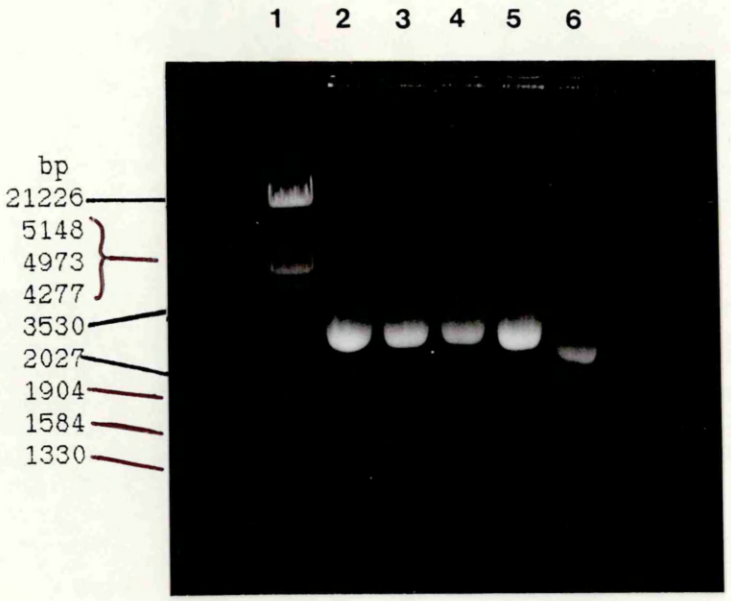


Figure 2.7

Lane 1: Lambda cI857 DNA marker cut with EcoRI+HindIII.

Lanes 2-5: Single-stranded sequencing template DNAs.
The templates are derivatives of mp19 each containing a HpaII insert (≈ 250 bp) from lambda clone 9 (See Ch.8).
Note there is very little smearing of the template DNAs and minimal evidence of *E. coli* host genomic DNA.

Lane 6: M13 mp19 (+) strand.

This is a minigel to which 0.5μgs/ml of ethidium bromide had been added before pouring.

2.19 The Sequencing Reactions

The templates were sequenced by the dideoxy chain termination method of Sanger *et al.*, 1977.

This method involves the *in vitro* synthesis, by a DNA polymerase, of a DNA strand which is complementary to the DNA template. Synthesis of the new strand is initiated at only one site, where an oligonucleotide primer has been annealed to the template. The synthesis reaction is terminated by the incorporation of one of four nucleotide analogues, the 2'3'-dideoxynucleoside 5' triphosphates (ddNTPs). These lack the 3' hydroxyl group required for DNA chain elongation.

When mixtures of dNTPs and and one of the four ddNTPs are used to synthesise a complementary strand, enzyme catalysed polymerisation will be terminated in a fraction of the population of chains at each site where the ddNTP can be incorporated. Four separate reactions, each with a different ddNTP therefore gives complete sequence information.

A radioactively labelled nucleotide is also included in the synthesis. This permits the labelled chains of various lengths to be visualised by autoradiography after separation by high resolution, polyacrylamide gel electrophoresis.

The sequencing reactions were carried out using reagents from a Sequenase Version 2.0 kit (1.32). The polymerase provided with the kit was a modified version of T7 DNA polymerase. This lacked the 3'⇒5' exonuclease activity of the wild type enzyme.

The kit instructions suggested ranges of incubation times and template concentrations which give satisfactory results, because of this the exact conditions used (determined empirically) are described overleaf.

Each sequencing reaction is composed of three parts, the annealing, labelling and termination reactions.

The Annealing Reaction:

The following reagents were dispensed into a 0.5µl Eppendorf vial:

Primer	1µl
Reaction Buffer	2µl
Template solution	7µl

The contents were mixed by pipetting, and the vial placed in a heating block at 65°C. After 2 minutes the heating block was switched off and with vial still in place, allowed to cool to <30°C (a time of ≈45 minutes). The vial was microfuged for a few seconds then placed on ice.

The Labelling Reaction:

Before use the Labelling Mix had to be diluted to the correct concentration for the length of template to be sequenced.

Length of template to be sequenced	Vol. of Labelling Mix	Vol. of Water	Working dilution
1-30 bases	4µl	56µl	15 fold
30-300 bases	4µl	16µl	5 fold
>300 bases	-	-	Undiluted

The Sequenase enzyme was diluted 1:8 with ice cold Enzyme Dilution buffer before use.

The following were added, in order, to the annealed template-primer mixture:

DTT 0.1M	1.0µl
Diluted Labelling Mix	2.0µl
800µCi/mmol [α-^{35}S] dATP	0.5µl
Diluted Sequenase	2.0µl

The reaction mixture was incubated at room temperature for 5 minutes.

The Termination Reactions:

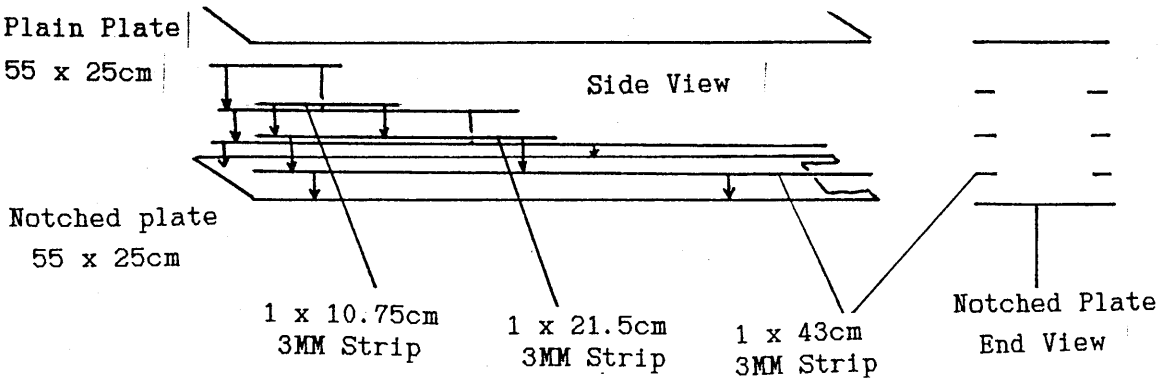
Four 0.5ml Eppendorf vials were required for each template to be sequenced. The vials were labelled G, A, T and C and placed on ice. Into each vial 2.5µls of the appropriate ddNTP was added ie. ddGTP into vial G, ddATP into vial A and so on. 2 minutes before the Labelling reaction was due to finish, the Termination vials were placed at 37°C to pre-warm the contents. When the Labelling reaction was complete, 3.5µls of it were dispensed into each of the four Termination vials. The vials were incubated at 37°C for a further 5 minutes, then 4µls of stop solution added to each vial. The completed reactions were usually run on a sequencing gel within the next 2 hours.

2.20 Sequencing Gels

Pouring and Pre-running:

Glass plates of the dimensions shown in fig. 2.8 were washed in detergent, rinsed thoroughly with tap water then dried. The gel sides of both plates were wiped over with a tissue soaked in chloroform or acetone, followed by a tissue soaked in methanol. The surfaces were dried, wiped over with a Silane (5% dichloromethylsilane in carbon tetrachloride) soaked tissue then polished to remove smears and tissue particles. Spacers were cut from 3MM paper to the dimensions shown in fig. 2.8 and the plates assembled. The sides and base of the plates were sealed with waterproof tape and the plates clamped together with bulldog clips. 70mls of 8% acrylamide (1.31) were poured between the plates, with care being taken to avoid trapping air bubbles. The flat edge of a shark's tooth comb was inserted between the plates then clamped tightly in place with bulldog clips.

After 1.5-2 hours, the tape at the bottom of the plates, all the bulldog clips, and the comb were removed. The slot formed by the comb was rinsed with 1X TBE (1.27) then the comb replaced so that



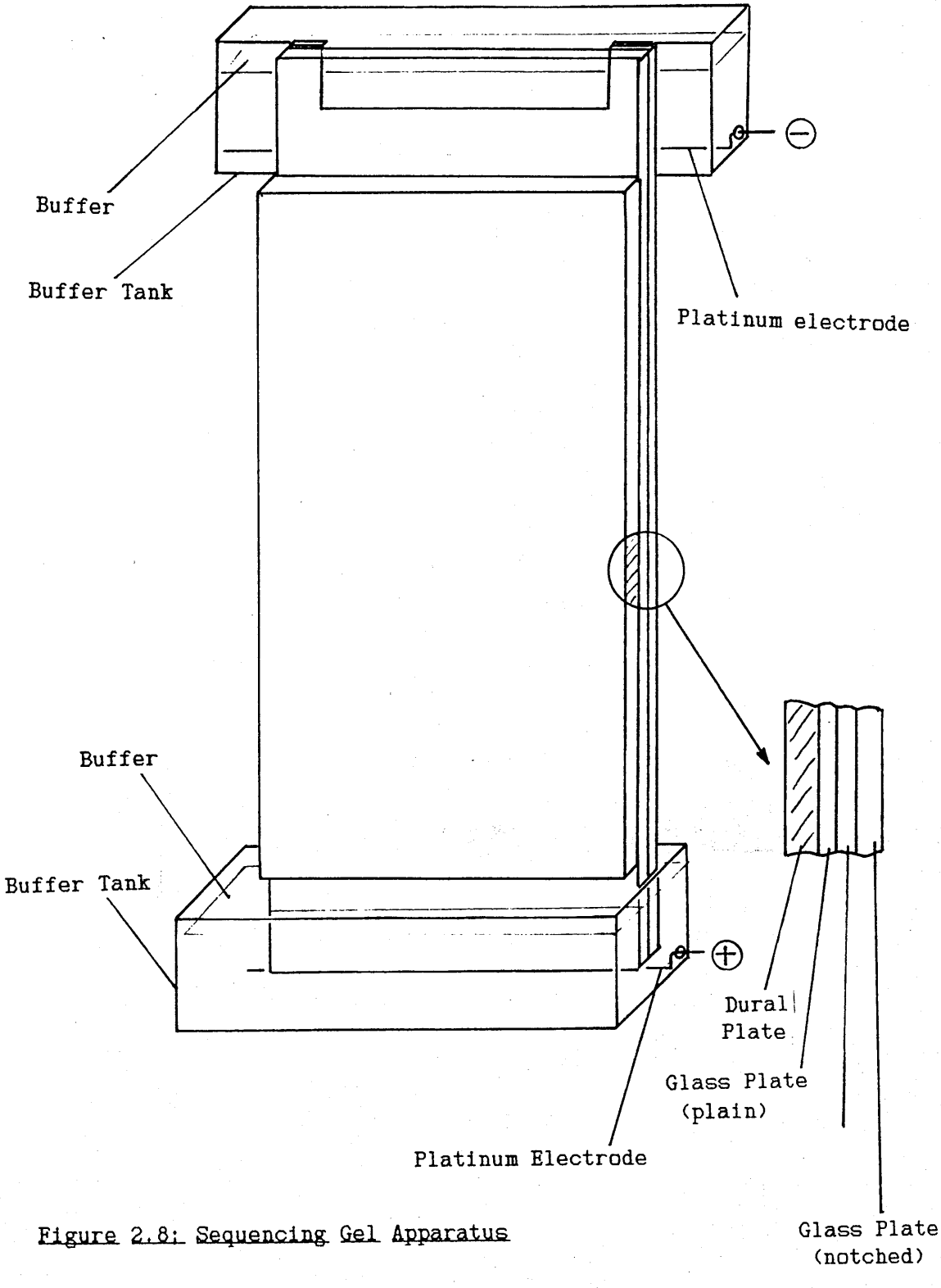


Figure 2.8: Sequencing Gel Apparatus

the teeth just caused indentations in the gel. The plates were clamped to the gel stand and a metal plate (see fig. 2.8) clamped to the front glass plate. The buffer tanks were filled with 1X TBE then the gel run for ≈1 hour at a constant power of 65 watts, 1500 volts and 35 milliamps. This pre-running raises the temperature of the gel which assists in keeping the DNA denatured during electrophoresis. The gel was considered ready for loading when the temperature in the centre of the metal plate was just too high to be comfortable on the back of the hand. Before samples were loaded, the buffer in the tanks was replaced with fresh 1X TBE.

Loading and Running:

Immediately before loading, the wells were washed out with 1x TBE to remove urea (which if not removed led to diffuse bands). The samples were heated at 80°C for 3 minutes to separate the templates from the newly synthesised DNA strands. The samples were kept at this temperature until they were loaded to prevent secondary structure formation, however they were not kept at this temperature for longer than 5 minutes in total. 3µls of each sample were loaded per well and no more than 8 samples were loaded at one time. If more than 8 samples were to be loaded, the gel was run for 5 minutes between loadings.

The gels were run at 1500V, 35mA and 65W for the required period of time. How long this time was depended on the region of the template to be resolved. Most templates were initially given a medium length run (≈4 hours). If the sequence information from this proved useful, shorter and longer runs (2 and 7 hours) were then carried out. The movement of the markers in the loading dye to particular positions within the gel coincided quite well with the following times. After 2 hours the bromophenol blue was about half way down the gel, after 4 hours it had just migrated out of the gel. After 7 hours the xylene cyanol had just migrated out of the gel. When the gels were to be run for 7 hours the buffer in the tanks was changed after 3½ hours.

Once electrophoresis was complete the buffer was discarded and the gel plates removed from the sequencing apparatus. The gel plates were placed, notched plate uppermost, on a flat surface and the plates gently separated with a spatula. Care was taken when separating the plates to prevent the gel tearing, and to ensure the gel remained attached to the lower plate. The 3MM spacers were trimmed away with a scalpel blade, then the lower plate and gel submerged in a tray containg 2.5 litres of 10% v/v acetic acid, 10% v/v methanol solution for 30 minutes. This process fixed the gel and removed the urea. If the urea was not removed the consequences would vary from the gel sticking to the photographic film during autoradiography, to the gel fracturing into innumerable pieces during drying.

The lower gel plate, with loosely attached gel, was removed from the tray and allowed to drain. A piece of 3MM paper, with dimensions slightly larger than that of the gel, was immersed in the tray of fixative for a few seconds, then laid on top of the gel without trapping air bubbles or forming creases. Copious amounts of tissue were then applied to the 3MM paper to absorb most of the liquid. With the tissue-3MM-gel-lower plate in place the notched plate was placed on top of the tissue then the ensemble inverted so that the notched plate was lower most. The upper plate was gently removed, care being taken not to remove any of the gel with it. Saran wrap was laid over the gel, and the Saran wrap-gel-3MM paper sandwich dried in a vacuum gel drier for 1.5 hours at 80°C. The dried gel was placed in an autoradiography cassette (without an intensifying screen) with a piece of X-OMat S film for ≈18 hours at room temperature. The film was then developed automatically in an X-OMAT processor.

After about 18 hours exposure a DNA sequence of 200-250bp could usually be read from the autoradiograph. Exposures of 72-144 hours often revealed a further 50bp. Generally, longer exposures did not reveal bands which could be read with certainty and were therefore not used.

2.21 Nucleotide Sequence Analysis

Nucleotide sequences, generated by dideoxy sequencing, were compared with the nucleotide sequence of the insert of plasmid pLHIISB18 (2.7). Comparisons were made by using Bestfit, a programme which is part of a sequence analysis software package (The GENETICS COMPUTER GROUP, version 5* from the University of Wisconsin).

The Bestfit programme makes an optimal alignment of the best segment of similarity between two sequences by using the "local homology" algorithm of Smith and Waterman [Smith and Waterman, 1981]. The magnitude of the computer's job is proportional to the area of the surface of comparison. That area is determined by the product of the lengths of the two sequences being compared.

The programme gives scores for or against features within the alignment, for example the scores given to base matches and mismatches were +1 and -0.9 respectively. In this way the programme constructs a path matrix that represents the entire surface of the comparison with a score at every position for the best possible alignment path to that point. Random alignments should have a path value that average approximately zero. If the best path to any point has a negative value, a zero is entered at that position.

After the path matrix is complete, the highest value on the surface represents the best region of similarity between the two sequences. The best path from this value back to the point where the values revert to zero is the alignment shown by Bestfit. The alignment is the best segment of similarity between the two sequences. The percent similarity between the sequences was determined automatically by the programme and where applicable have been quoted in the text.

Bestfit always finds a solution to any two sequences that are compared, even if there is no "genuine" similarity between them.

Therefore the alignments found had to be analysed critically and a decision taken as to whether the segment shown was a real or random region of relative similarity. In reality this was not difficult, random segments of similarity tended to be short and/or with a large number of base mismatches/gaps in the aligned sequences. "Genuine" regions of similarity would extend for many tens of bases with only a few base mismatches or gaps in the alignment.

CHAPTER: 3 MATERIALS AND METHODS

3.1 Genomic Library Construction (Theoretical Considerations)

The library

The perfect genomic DNA library contains DNA sequences representative of the entire genome, in a stable form, as a manageable number of overlapping clones. The cloned fragments should be small enough to be mapped easily by restriction enzyme analysis, but large enough to contain entire genes and their flanking sequences. The library should be easy to construct from small amounts of starting material and easily screened. Ideally the library should amplify without loss or misrepresentation and be capable of storage for years without loss of titre. Phage lambda vectors provide an easy and efficient means for the routine construction of genomic DNA libraries [Kaiser and Murray, 1985].

Bacteriophage lambda is a double stranded DNA virus with a genome size of approximately 50kb. Since the first demonstration of the feasibility of using lambda as a vector [Murray and Murray, 1974; Rambach and Tiollais, 1974; Thomas *et al.*, 1974] a large variety of vectors have been produced. One of these, the lambda replacement vector, EMBL3 [Frischauf *et al.*, 1983], was used to construct the *R. acidophila* genomic DNA library.

The vector

For cloning purposes the EMBL3 genome can be considered as consisting of three main regions (see fig. 3.1); a 20kb left arm encoding the virion proteins; a central, non-essential stuffer fragment (14kb); and a 9kb right arm containing the origin of replication, major promoters and other essential genes. At the extremeties of both arms are 12 base complementary single strand projections, the cohesive ends (cos).

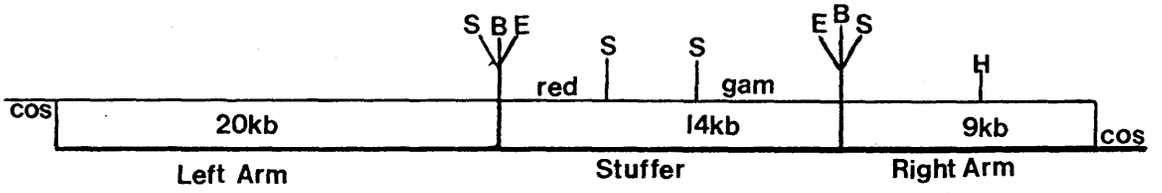


Figure 3.1

Schematic representation of the lambda EMBL3 genome which consists of three main regions: 20kb left arm, 14kb stuffer fragment and a 9kb right arm. At the extremities of the arms are 12bp complementary single stranded projections *cos*.

B=BamHI, S=SalI, E=EcoRI and H=HindIII.

The stuffer fragment is flanked by polylinkers in inverse orientation:

Left Arm--SalI-BamHI-EcoRI--**Stuffer**--EcoRI-BamHI-SalI--**Right Arm**

These sites are unique within the vector for these enzymes, with the exception of two internal SalI sites in the stuffer.

Although any of these enzymes can be used for cloning, the vector was designed specifically for use with BamHI. The rationale behind this is as follows. Cleavage with BamHI generates a staggered break at the hexanucleotide sequence 5'-G↑GATCC-3'. In a piece of random sequence DNA any hexanucleotide sequence occurs, on average, once every 4096 base pairs. Too infrequently for the random cleavage of donor DNA required for the construction of the ideal library. However, the core of the BamHI sequence, is also the recognition site of the enzyme Sau3AI. This cleaves at 5'-GATC-3', which occurs on average every 256bp. Thus EMBL3 DNA cleaved with BamHI can accept an exogenous DNA fragment (the insert) prepaired by an essentially random partial digestion with Sau3AI.

If BamHI cleaved EMBL3 DNA, was mixed with Sau3AI cleaved genomic DNA and ligated, the result would be a mixed population of EMBL3 recombinant and EMBL3 'wild-type' vector DNAs. On plating, the latter DNA species would lead to a background of non-recombinant phage. This problem can be reduced by physical removal of the stuffer fragment so that only vector arms and insert DNAs can participate in the ligation reaction. This is usually achieved by velocity gradient fractionation of the cleaved vector DNA [Maniatis *et al.*, 1982].

In the case of EMBL3 such a complex approach is not required. The BamHI sites of the vector are flanked internally by EcoRI sites. Cleavage with both of these enzymes releases a 10bp EcoRI-BamHI fragment from each end of the stuffer fragment. These fragments can be removed by differential isopropanol precipitation. The stuffer

fragments, which now have EcoRI ends are thus rendered incapable of competing with insert DNA fragments for ligation to the vector arms.

An additional advantage of using EMBL3 as a vector is the genetic selection system against non-recombinant phage. The stuffer fragment of EMBL3 encodes the genes *red* and *gam*. Possesion of these genes prevents non-recombinant phage from being propagated on an *E. coli* host containing a P2 lysogen, that is, these phage posses the Spi+ phenoype, Sensitive to P2 Interference. By plating the genomic library on an *E. coli* host containing a P2 lysogen (for example, strain Q359), only recombinant phage lacking the stuffer fragment (Spi-) will form plaques. Recombinant and non-recombinant phage can be propagated on the isogenic strain Q358 which lacks P2.

The interaction between the vector and the host

The viabilty of a particular genomic clone depends on viral replication and packaging. These are the outcome of a complex interaction between the Chi elements, *red* and *gam* genes of the vector, and the *rec* genes of the host.

During the course of parental vector growth, two distinct forms of replicating phage chromosomes co-exist within the cell [Furth and Wickner, 1983]. The first, called the theta (θ) form, is a typical bidirectional mode of replication that forms replication bubbles in circular molecules. The second type is the sigma (σ) or rolling circle form, in which one of the DNA strands is nicked and displaced by the procession of continuous DNA synthesis on the circular template. The displaced strand is copied in a normal fashion by discontinous DNA synthesis. These types of replication generate different products: θ replication generates monomeric circles, while σ produces concatenated linear molecules.

The substrate for the lambda packaging process must contain at least two cohesive end sequences in the same orientation [Feiss and Becker, 1983]. Products of σ replication are packaged directly. In contrast, since the monomeric circles generated by θ replication

contain only a single *cos* site, they can be packaged only if recombination between monomers takes place. Thus most of the packaged chromosomes are products of σ replication.

Early in the lambda life cycle, replication is primarily by the θ mode. At later times, σ type intermediates are seen in addition to the θ molecules [Better and Freifelder, 1983]. This depression of rolling circle replication is probably due to the accumulation of the lambda gamma protein (the *gam* gene product) which is an inhibitor the nucleolytic activity of the host RecBCD enzyme (Exonuclease V or Exo V for short) [Enquist and Skalka, 1973]. These observations led to the suggestion that rolling circle structures are sensitive to activity of the RecBCD enzyme.

However, *gam*$^-$ phage (such as EMBL3 recombinants) fail to inhibit the RecBCD enzyme and are thus unable to produce concatemers via σ replication. The *gam*$^-$ recombinants can be propagated in *rec*BCD$^+$ hosts, provided that they can undergo molecular recombination to produce di- or multimeric (and therefore concatenated) circular DNA molecules. Unfortunately, this cannot be achieved by the phage encoded recombination enzymes (Red) as the *red* genes, lie adjacent to the *gam* genes on the EMBL3 stuffer fragment which has been replaced by insert DNA. Hence, *red*$^-$*gam*$^-$ recombinants rely entirely on host recombination (Rec) functions. In addition, genetic selection for the *red*$^-$*gam*$^-$ (Spi$^-$) recombinants on hosts lysogenic for P2, requires the use of Rec$^+$ cells. Maintainance of the *recBC* product is essential for survival of the P2 lysogen.

However, only phage molecules which posses one or more copies of a Chi (Cross over Hotspot Instigator) sequence will act as a suitable substrate for recombination via the *RecBC* pathway. As a proportion of the inserts in a genomic library would carry, by chance, a Chi sequence, then these recombinants would have a selective advantage over Chi$^-$ phage. To overcome this problem EMBL3 has been endowed with Chi sequences in the vector arms.

The selection of inserts

There are lower (≃40kb) and upper (≃52kb) limits to the size of phage genomes which can be packaged into phage heads. Cutting genomic DNA with Sau3AI and ligating it to vector arms will result in a proportion of DNA molecules which will fail to fulfill these parameters. More disturbing however, are inserts which place the genome on the limits of packagable size (ie ≃9 and 22kb) and which tend to be susceptible to Rec-dependent deletion and duplication, or Rec-independent deletion and rearrangement. Equally worrying, are packageable anagram sequences" which arise from the insertion of small, concatenated inserts into the vector. Dephosphorylation of the inserts with calf intestinal phosphatase before ligation to the vector arms would prevent this. However, the general consensus appears to be that this only exchanges one problem for another ie. that of removing the phophatase! All of these problems can be overcome by selecting inserts of ≃20kb in size for ligation to the vector arms. This can be achieved either by fractionation of the partially digested genomic DNA on a sucrose-sodium chloride velocity gradient or by cutting the appropriate sized DNA fragments out of a low melting point agarose gel. Although technically easier, the latter has two main disadvantages; a) the gel agarose may contaminate the DNA with unremovable ligase inhibitors and b) the size of the DNA may be reduced in size by photonicking when the band is being excised from the gel or by cleavage during subsequent purification steps. For these reasons the insert DNA used to construct the library was size selected after passage over a sucrose-sodium chloride velocity gradient.

The size of the library

The number of recombinant phage required in the library so that any piece of genome is represented can be calculated from the equation:

$$N=In(1-P)/In(1-x/y)$$

Where x is the insert size, y is the size of the haploid genome, p is the probability of the library containing any sequence and N is the number of clones in the library. Unfortunately the size of the *R. acidophila* genome is not known. However, it seemed reasonable to

assume that it would lie in the range between that of *E. coli* (4.2 x 10^6bp) and *Saccharomyces cerevisiae* (1.4 x 10^7). As 20kb inserts were to be cloned, and aiming for a 99% probability (p=0.99) of the library containing a particular DNA sequence, then the library must contain, at least, 9.2 x 10^2 - 3.2 x 10^3 independent clones in order to represent the entire genome.

3.2 Genomic Library Construction (Materials and Methods)

Partial digestion of genomic DNA

DNA inserts, of ≃20kb in size, were required for the construction of the genomic library. These inserts were generated by partial digestion of high molecular weight genomic DNA with Sau3AI. However, before committing large amounts of genomic DNA to digestion, the optimal digestion conditions were determined first by calculation, then empirically in a small scale reaction.

The Calaculation

The digestion conditions were calculated as follows:
Sau3AI, on average, will cut a piece of random sequence DNA approximately once every 250bp.
The size of the fragments required for cloning are 20,000bp.
Therefore, to generate 20kb fragments one eightieth of the sites need to be cut. ie. 250/20,000=1/80.
Under optimal conditions 1 unit of restriction enzyme will digest 1µg of lambda cI857 DNA to completion in 1 hour. For this purpose this can be approximated to:

1 unit of enzyme digests 80/80 sites in 1µg of genomic DNA in 1hr
≑1/80 " " " " 1/80 " " 1µg " " " " 1hr
≑1/10 " " " " 1/80 " " 8µg " " " " 1hr
≑1/5 " " " " 1/80 " " 8µg " " " 0.5hr

That is, 8µg of DNA digested with 0.2 units of enzyme for 30 minutes should yield the maximum number of 20kb fragments.

These parameters were used as a rough guide, to determine the conditions required to carry out a small scale test reaction.

The Test Digest

The literature [Maniatis *et al.*, 1982; Kaiser and Murray, 1985] suggested that small test reactions should be carried out by adding serially diluted enzyme to tubes containing fixed amounts of DNA. The reactions are then carried out for a fixed period of time and the digestion products analysed on a gel.

The disadvantage of this method is that pipetting viscous DNA and serially diluting very small volumes, tends to be inaccurate irrespective of the care taken. Therefore, in this case, enzyme was added to the DNA in a single reaction and aliquots removed at various times around that calculated.

The digestion conditions were tested experimentally as follows. Genomic DNA was prepared (2.2) and the concentration determined (0.4µg/µl)(2.6). Sau3AI at a concentration of 4units/µl was added to 1x ReAct 4 buffer, mixed in by vortexing, microfuged briefly, then placed on ice. The diluted enzyme had a concentration of 0.05units/µl. The following were added to a 0.75ml Eppendorf vial:

30µls of water.
6µls of 10x React 4.
20µls of 0.4µg/µl genomic DNA.

The contents were mixed by gentle inversion of the vial, microfuged briefly, then placed in a 37°C water bath for 30 minutes, to allow the contents to equilibrate. Meanwhile, 2µls of EDTA pH 8.0 were added to each of six Eppendorf vials, which were then placed on ice.

4µls of diluted Sau3AI (0.05 units/µl) were added to the reaction vial, mixed in with a glass capillary and the vial returned to the water bath. After 3, 7, 15, 30, 60, and 120 minutes, a 10µl aliquot of the digest was removed and dispensed into a vial containing chilled 0.5M EDTA. The vial was incubated immediately at 65°C, to denature the enzyme, then after 10 minutes returned to ice. 2µls

of loading dye (1.29) were added to each sample, which were then run on a 0.4% agarose gel (2.5). The digestion time which gave the largest number of 20kb fragments was used as the basis for a larger digest.

Large Scale Digest

As before, aliquots were removed at different, though more closely spaced, time points. This hopefully counteracted possible pipetting errors and effects of increasing the digest size. The digest was prepared by adding the following to a 5ml tube:

2000µl of 0.4µg/µl DNA solution
300µl of 10x React buffer 4
695µl of water

The contents were mixed gently by inversion and placed in a 37°C water bath for 30 minutes to equilibrate. Meanwhile, 50µls of 0.5M EDTA were dispensed into each of six, 1.5ml Eppendorf vials, which were then placed on ice.

5µls of Sau3AI (4 units/µl) were added to the reaction, mixed in gently with a glass capillary, then the tube returned to the waterbath. After 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 minutes, a 300µl aliquot was removed from the reaction mix to a tube containing chilled 0.5M EDTA. The tube was immediately incubated at 65°C for 10 minutes to denature the enzyme. 10µls of each timed aliquot were removed, then 2µls of loading dye added. The samples were then run on a 0.4% agarose gel. The gel was stained with ethidium bromide and the sizes and distribution of the DNA fragments examined. Aliquots containing a useful proportion of 20kb fragments (times of 3-7 minutes inclusive) were pooled then stored at 4°C. The pooled volume was ≈1,750µls.

Size Fractionation of the Partial Digest

5 and 25% (w/v) sucrose solutions were made in a buffer containing, 1M sodium chloride, 20mM Tris.HCl pH 8.0, and 5mM EDTA pH 8.0. The sucrose solutions were sterilised by autoclaving at 10psi for 10 minutes before use. 20 and 21mls, of 25 and 5% sucrose solutions

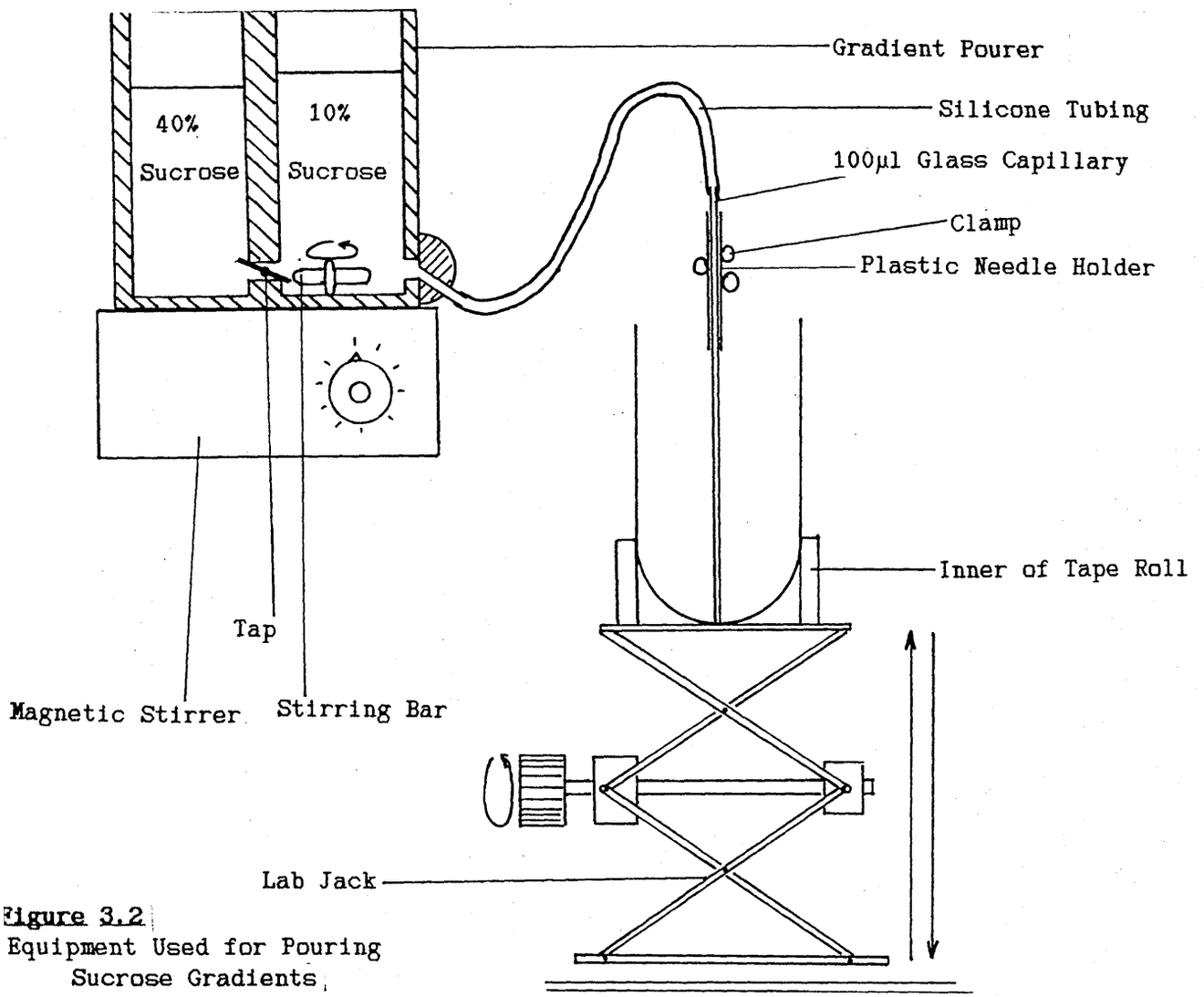


Figure 3.2
Equipment Used for Pouring
Sucrose Gradients

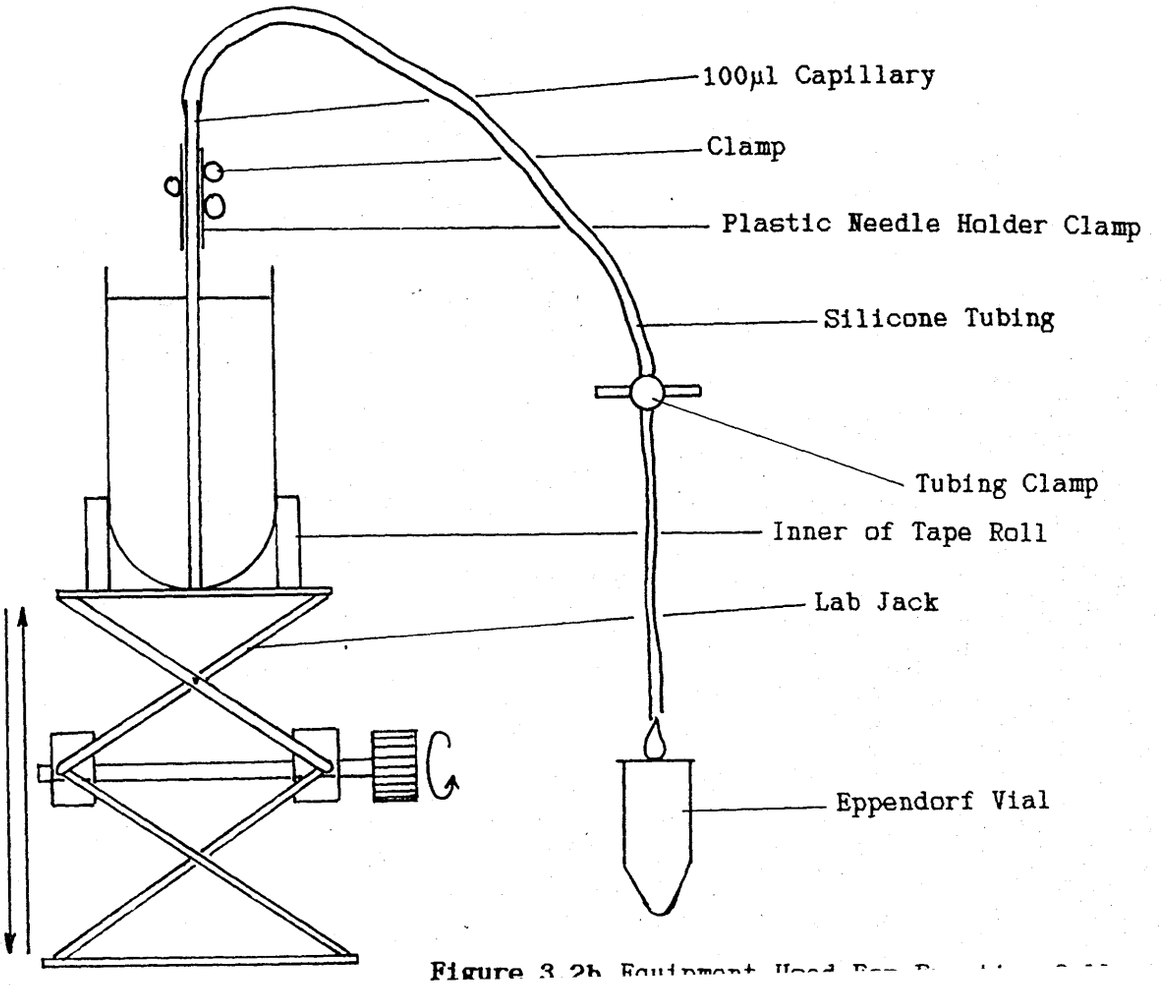


Figure 3.2b Equipment Used For P...

respectively, were added to the appropriate chambers of a gradient former, see figure 3.2. A linear sucrose gradient was then poured into a 40ml Beckman SW.28 polyallomer tube. Note that the combined volumes of the sucrose solutions were greater than the capacity of the centrifuge tube. The volumes used were determined empirically and account for losses caused by sucrose solution becoming trapped in the gradient pourer. The volume of the gradient was 36.5ml.

The pooled DNA digest was heated to 65°C for 10 minutes, to separate sticky ends, then the digest (1,1750µls containing 400µgs of DNA) gently added to the gradient. The volumes of the gradient was made up to ≈39ml with TE (1.20), balanced (with TE), then spun in a Beckman SW.28 rotor, at 104,000 x g, for 18 hours, at 20°C.

The gradient was fractionated using the equipment shown in figure 3.2b. Gradient tubes were placed on the lab jack and gently raised until the free sliding capillary began to move in its holder. The clamp, which prevented a core being taken through the gradient, was then removed and ≈1ml fractions siphoned slowly (≈1drop/sec) into numbered Eppendorf vials.

To determine which fractions contained DNA, 1µl of each aliquot was spotted onto an ethidium bromide plate (2.6). 30µl aliquots were then removed from every other DNA containing fraction. To each aliquot was added, 40µls of TE (to dilute the sodium chloride) and 15µls of loading dye. The samples were heated to 65°C for 15 minutes, to dissociate base paired sticky ends, then run with HindIII cut lambda cI857 DNA, on a 0.4% agarose gel. The gel was stained with ethidium bromide and the DNA visuallised on a transilluminator.

All fractions containing DNA fragments of approximately 15-25kb, were placed in tight dialysis sacs (1.16) and dialysed to remove the sucrose and sodium chloride. Dialysis consisted of, two, 2 litre changes of TE, each change lasting ≈2 hours. The fractions were transferred to separate Eppendorf vials and a 40µl aliquot removed

from each. To each aliquot, 10µls of loading dye were added and the samples heated to 65°C for 10 minutes to separate sticky ends. The samples were run with EcoRI/HindIII cut lambda cI857 DNA, on a 0.4% agarose gel.

Fractions containing fragments just slightly smaller than the 21226bp lambda cI857 EcoRI/HindIII fragment were ethanol precipitated overnight at -20°C. The DNA was pelleted in a 7ml cellulose nitrate tube by centrifugation in a Beckman SW65 rotor, at 120,000 x g, for 30 minutes, at 0°C. The supernatant was decanted and the inside of the tube dried with tissue. The pellet was washed gently with 80% ethanol, then the inside of the tube dried as before. The pellet was then dried by gently blowing air over it with a Pasteur pipette. 10µls of water were added to the pellet, which was then allowed to dissolve overnight at 4°C. The DNA solution was transferred to a clean Eppendorf vial and stored at 4°C until required for cloning.

Preparation of Vector Arms

EMBL3 vector DNA was prepared as described previously (2.3). The enzymes EcoRI and BamHI, and the vector DNA were tested in three reactions:

	Reaction 1	Reaction 2	Reaction 3
water	8.0µls	8.0µls	7.5µls
10x ReAct 3 buffer	1.0µl	1.0µl	1.0µl
EMBL3 DNA (0.5µg/µl)	0.5µls	0.5µls	0.5µls
BamHI (10units/µl)	0.5µl	--	0.5µl
EcoRI (10units/µl)	--	0.5µl	0.5µl

The reaction mixtures were incubated at 37°C for 1 hour, heated to 65°C for 15 minutes, to separate the cohesive ends, then run on a 0.4% agarose gel (2.5).

Vector arms for cloning were prepared by digesting 50µgs (100µls) of vector DNA, with a 3- fold excess of BamHI (150 units), in a 400µl reaction, for 1 hour. To check the reaction had gone to completion,

a sample (≈0.25μg), was run on a 0.4% minigel. 100 units of EcoRI were then added to the digest and the reaction continued for a further 2 hours. A small sample (0.5μg), was run on a large, 0.4% agarose gel to ensure that degradation of the vector had not taken place during the second digestion. The reaction mixture was extracted twice with phenol-chloroform (1.14), and once with chloroform. 40μls of 3M sodium acetate pH 5.2 (1.17) and 264μls of isopropanol (0.6 volumes) were added to the aqueous phase and mixed in by inverting the tube several times. The tube was left at room temperature for 15 minutes, microfuged for 10 minutes, then the supernatant containing the 10bp EcoRI-BamHI polylinker fragments aspirated. The pellet was washed with 80% ethanol, dried, then left to dissolve in 30μls of water overnight.

Ligation of Vector Arms and Size Selected Inserts

In order to generate a viable recombinant DNA molecule, each insert needs to be ligated to a right and left vector arm. If the cohesive ends of the vector arms are annealed (see below) prior to ligation, then the maximum number of viable recombinant molecules are generated by having inserts and annealed arms present in a 1:1 molar ratio. By spotting 0.5μls of arm and insert DNA onto an ethidium plate, the DNA concentrations were found to be 1 and 0.4μg/μl respectively.

The concentration of the vector DNA was 1μg/μl. However, the DNA solution was composed of arms and stuffer fragments. Only the arms were capable of participating in the ligation reaction and their concentration was calculated to be:

$$\frac{\text{Size of the arms=30kb}}{\text{Size of the intact vector=50kb}} \times 1\mu g/\mu l$$

$$= 0.6\mu g/\mu l \text{ of vector arms.}$$

Note that the size of the intact vector was taken to be 50kb, whereas the actual size of the vector is 43kb. This mistake has been shown as it was made during the "real" calculation.

As the inserts were 20kb in size, and the annealed arms were 30kb, then the amount of insert DNA required to give a 1:1 molar ratio of arm:insert with 0.6µgs of arms was:

$$\frac{20kb}{30kb} \times 0.6\mu g = 0.4\mu g.$$

A 1:1 molar ratio of arms to inserts was achieved by using equal volumes of each DNA solution in the ligation mixture.

Before ligation, the cohesive ends of the lambda arms were annealed in the following mixture:

3µl of vector DNA solution (ie ≃1.8µgs of arms)
3µl of insert DNA solution (ie ≃1.2µgs of insert)
1ul of 10x ligation buffer (0.66M Tris.HCl pH 7.5, 0.1M $MgCl_2$)

This mixture, which forms part of the ligation reaction (see below) was then incubated at 42°C for 1 hour.

Annealing increased the number of linear and circular DNA conacatenates which were present after the completion of the ligation reaction. These concatenates, and to a lesser extent linear monomers, are substrates for the subsequent packaging reaction. Monomeric circles are not substrates for packaging. The process of concatenisation was enhanced further by keeping the ligation volume to a minimum, resulting in a high DNA concentration. This favours inter-molecular joining, concatenisation, over intra-molecular joining, monomeric circle formation [Kaiser and Murray; 1985].

The following were then added to the solution of annealed arms:

1µl of 100mM DTT
1ul of 10mM ATP
1µl of T4 ligase (1 unit/µl).

The mixture was overlaid with paraffin oil, incubated overnight at 14°C, then stored at 4°C until required for packaging.

Packaging the Recombinant Molecules

Packaging requires a number of different proteins encoded by the lambda genome. These proteins were purchased as packaging extracts, prepared from cells infected with defective lambda strains. Each strain carries a mutation in one of the components of the protein coat. Infected cells continue to synthesise and accumulate all the other components, even though mature phage particles cannot be assembled. The *in vitro* packaging extracts are prepared by mixing the lysates of the two defective phage strains. The mixture contains all the necessary components to efficiently package recombinant lambda genomes into their protein coats.

The recombinants *were* packaged using a commercially available in *vitro packaging* kit (Amersham International). According to the manufacturer's instructions, the maximum packaging efficiency (ie. greatest number of plaque forming units (pfu)/µg of DNA) would be obtained by using 0.5µg of DNA in a maximum volume of 10µls.

The ligation reaction (volume = 10µls) contained 3µgs of DNA (vector arms and insert DNA). 1.6µls (0.5µgs of vector arms and insert DNA) were removed, the volume made up to 8µls with water, then added to the thawed packaging extracts. The mixture was incubated for 1 hour at 37°C. 500µls of phage buffer were added to the packaged phage which brought the volume up to ≈550µls, then a drop of chloroform added. The packaged phage were stored at 4°C until required.

As a control for packaging efficiency 0.5µgs of intact EMBL3 DNA (in 8µls of TE) were also packaged as described above.

Chapter 3.3 Titering and Amplifying the Library

The packaging efficiency of each packaging extract, was, according to the manufacturer, 10^6-10^7 pfu/µg of recombinant DNA. As only 0.5µgs of DNA was packaged in each reaction, the total number of recombinant phage expected per reaction was 5 x 10^5 - 5 x 10^6pfu. This in turn suggested the titre of the diluted packaging mix was 5 x 10^{5-6}pfu/550µl, ≃10^2-10^3 pfu/µl.

A 20µl aliquot of packaged phage was removed and serially diluted 10-fold by addition to 180µls of phage buffer (1.19). 20µls of the diluted phage were removed and serial dilutions repeated until the expected titre corresponded to ≃10 pfu/180µls of buffer. The dilutions of 10, 100, 1000 and 10,000 pfu/180µls were plated as follows.

Plating cells (2.4) were prepared from *E. coli* strains Q358, Q359 and NM621. A 30µl aliquot of each dilution was added to each of two, 300µl aliquots of each plating cell strain. The mixture was vortexed briefly, then incubated at room temperature for 20 minutes to allow phage attachment. 3mls of BBL-top layer (1.8) were added to the cells which were then poured onto a BBL broth agar plate (1.7). The plates were incubated overnight at 37°C. The number of plaques on each plate was determined and from this the titre of the library calculated.

The packaged EMBL3 DNA, used as a control for determining packaging efficiency, was also titred as described above.

Amplification

In vitro packaging has the disadvantage that the packaged phage have a shorter half-life than those obtained by growth *in vivo*. However, by plating the phage and allowing them to amplify in a host, a stable, high titre, library stock can be obtained. An additional advantage, is that the amplified library can then be distributed to other workers if required. The library was amplified

by adding ≈2,000 pfu to each of nine, 10x10cm plates. Amplification at such low levels, reduced the likelihood of more than one phage simultaneously infecting a bacterium. Such an infection could have led to recombination between repetitive elements present on different recombinant phage inserts. The library was amplified as follows.

The titre of the unamplified library was determined, (see above) and from this calculated the volume of packaged phage containing 2,000 pfu(10µls). This volume was dispensed into nine tubes, each containing 190µls of phage buffer. The mixture was vortexed briefly, then 0.8µls of freshly prepared NM621 plating cells added and the mixture vortexed again. The bacterial suspensions were incubated at room temperature for 20 minutes, to allow attachment of the phage. 5mls of BBL top agar were then added and the contents poured onto a 10 x 10cm BBL agar plate. The plates were incubated at 37°C for ≈12 hours. This incubation time was a compromise between allowing the plaques to grow large enough to be visible, but not so large that they were growing into one another.

15mls of phage buffer were added to each plate then the plates rocked gently for ≈2 hours, to allow tranfer of the phage from the agar to the buffer. The phage solution was poured into a polythene bottle, 0.5mls of chloroform added, mixed in by vortexing and the phage stored at 4°C until required.

3.4 Screening the Library

The strategy for screening the library and the subsequent purification of positively hybridising plaques was as follows. The amplified library was plated at a low density then two duplicate filters lifted from each plate. The filters were then probed with radiolabelled pLHIISB18 DNA (2.7), which encodes the *Rhodobacter sphaeroides* B800-850 structural genes. Positively hybridising

plaques (present on both replica autoradiographs) as opposed to background spots (present on only one autoradiograph) were picked.

The plaques were purified by streaking onto *E. coli* lawns, then picking well separated plaques into two duplicate arrays. One array was screened as above, the other stored until the results of the screening were known. Positively hybridising plaques from the screened array were then matched to those of the unscreened array. The unscreened positive plaques were picked and kept as recombinant phage stocks.

The library was also screened with radiolabelled EMBL3 stuffer fragment and radiolabelled *R. acidophila* genomic DNA. These screenings were used to gauge the level of "background" caused by parental EMBL3 vector. These screenings utilised the plates which were used to titre the amplified library (see below).

Titering the Amplified Library

The titre of the amplified library was determined by plating 10-fold serial dilutions as follows. A 20µl aliquot of the amplified library was removed and added to 180µls of phage buffer (1.19). The mixture was vortexed, 20µls removed then added to 180µls of phage buffer. This dilution process was repeated nine more times to reduce the titre of the phage (from the expected titre of 10^{9}-10^{11} pfu/ml) to a level which could be counted (1-500 plaques) on an 80mm diameter petri dish.

From each of the five dilutions with the lowest titres, two 90µl aliquots of phage suspension were removed and added to 0.4mls of NM621 plating cells (2.4). The mixture was briefly vortexed, then incubated at room temperature for 20 minutes to allow phage attachment. 3mls of BBL top agar (1.8) were then added and the mixture poured onto a BBL broth agar plate (1.7). The plates were incubated at 37°C overnight. The plaques on each plate were counted and the titre of the amplified library determined (6.75 x 10^{9}pfu/ml).

Screening

A 1ml aliquot of the amplified library was removed and diluted 2-fold, twelve times. 1ml of the final dilution was added to 49ml of phage buffer to give a calculated titre of 27,000 pfu/ml. The titre of the diluted diluted library was checked as described above.

The library was plated by preparing 10 plates as follows. 0.1mls of the diluted library (≃2,200pfu) were added to 0.8mls of freshly prepared NM621 plating cells. The mixture was briefly vortexed, then incubated at room temperature for 20 minutes to allow phage attachment. 5mls of warm BBL top layer agarose were then added and the mixture poured onto a 10 x 10cm BBL broth agar plate. The plates were incubated for ≃12 hours, which produced visible but not confluent plaque growth. The plates were then placed at 4°C for 1 hour to allow the agarose top layer to harden.

Duplicate filters were lifted from the plates, and the attached DNA denatured, neutralised and washed as described previously (2.11). The filters were then screened (2.13) with radiolabelled pLHIISB18 (2.7), (2.10). However, as there was a total of twenty filters it seemed prudent to divide these into two groups of ten for hybridisation. Duplicate filter pairs were kept together in each hybridisation bag. In addition 20ngs of probe were labelled in two separate 10ng random priming reactions. The labelled products were run over individual columns then pooled. Half of the probe was then added to each of the hybridisation bags.

Plates which had been used to determine the titre of the amplified library, were also used to determine the level of background due to "non-recombinant" vector. For each library dilution there were two titre plates. Two filters were lifted from each plate as described previously (2.11). One duplicate pair of filters was then probed with radiolabelled stuffer fragment, the other pair with radiolabelled genomic *R. acidophila* DNA (2.13), (2.10).

3.5 Positive Recombinant Plaque Purification

The filters were aligned with their autoradiographs and the asymmetrical orientation holes marked in place on the film. The duplicate autoradiographs were compared and any background spots (present on only one filter) noted. The appropriate plate was laid over the corresponding autoradiograph, then the orientation holes in the plate aligned with the orientation marks on the film.

Isolated, positively hybridising plaques were picked as small agar plugs, using the narrow bore end of a Pasteur pipette. The plugs were transferred into ≈1ml of phage buffer. If the plaque was part of a cluster, the cluster was picked as a large plug using the wide bore end of a pipette. A fresh pipette was used for picking each plaque. The plugs were left at room temperature for ≈1 hour to allow dispersal of the phage.

Sufficient plates were prepared to allow each isolate to be streaked out. The plates were prepared by adding 3mls of BBL top agar (1.8), to 0.4mls of fresh NM621 plating cells (2.4), the mixture then being poured onto a BBL broth agar plate (1.7). A loopful of each phage suspension was then streaked onto the agar surface to give single plaques. The plates were incubated for ≈18 hours at 37°C.

Eight 10x10cm plates, were prepared by adding 5mls of BBL top layer agarose, to 0.8mls of freshly prepared NM621 plating cells and the mixture poured over a BBL broth agar bottom layer plate. Onto these plates were picked two identical arrays of plaques as follows. A photocopied grid was taped to the bottom of each square plate. A well isolated, single plaque was picked with a toothpick. The toothpick was then touched very gently, onto the identical grid coordinate of each plate. A row of plaques was picked in this way, then a new row started with a new plate of single isolated plaques. This process was continued until all the isolated plaque plates had been picked out. The plates were incubated for ≈12 hours then

placed in a fridge at 4°C. After ≃ 1 hour, one of each of the replica plates was removed, a single filter lifted, then probed as described for screening the library.

The plates from which filters had not been lifted were aligned with their replica autoradiographs. The corresponding, positively hybridising plaques, were removed as a plug into ≃ 1ml of phage buffer, a few drops of chloroform added and the phage stocks stored at 4°C.

3.6 Grouping and Mapping the Recombinant Clones

A plate lysate of each purified recombinant clone was made. These yielded sufficient DNA to allow preliminary classification of each phage. The DNA was cut with EcoRI and HindIII, then run on a gel and blotted. The blots were probed with radiolabelled pLHIISB18 DNA. The restriction pattern of the DNA, and sizes of positively hybridising fragments from each individual phage, were then compared with those of all the other phage. This allowed the removal of duplicated isolates (probably generated by library amplification) and retention of those isolates which were unique. Large scale DNA preparations were made from each of the unique isolates. The phage were then grouped on the basis of restriction pattern and members of these groups subsequently mapped.

Isolating DNA from Recombinant Clones Using Plate Lysates

For each phage stock, 3mls of L-broth top layer agar (1.5) were added to 0.4mls of NM621 plating cells (2.4) and the mixture poured onto an L-broth agar plate (1.6). A loopful of phage stock was streaked onto the surface of the agar and the plate incubated overnight at 37°C.

A well isolated plaque was picked into 500µls of phage buffer (1.19) and incubated at room temperature for 1 hour. 100µls of the phage suspension were added to 0.4mls of NM621 plating cells, briefly vortexed, then incubated at room temperature for 20 minutes, to allow phage attachment. 3mls of L-broth top layer agarose were added to the cells and the mixture poured onto a thick, wet, L-broth agar plate.

The plates were tapped on the shelf of the incubator (lid side up) so that the condensation would fall and help spread the phage. The plates were incubated for ≈18 hours at 37°C, then 5mls of phage buffer added. The plates were rocked gently for ≈2 hours to allow transfer of the phage to the buffer. The buffer was removed, care being taken not to remove any agarose, and transferred to a 10ml centrifuge tube. The phage suspension was spun in a JA-21 rotor, at 10,000 x g, for 10 minutes at 4°C, to pellet agarose particles and bacterial debris. The cleared phage lysate was transferred to a clean 10ml centrifuge tube and 5µls of a solution containing a mixture of 10mg/ml DNase I and 10mg/ml RNase added. The lysate was incubated at 37°C for 30 minutes, to degrade bacterial DNA and RNA. An equal volume of 20% w/v PEG 6000 in 2M NaCl was then added. The mixture vortexed, then the tubes left in ice water (0°C) for 1 hour to precipitate the phage.

The phage were pelleted by centrifugation in a JA-21 rotor at 45,000 x g, for 20 minutes, at 0°C. The supernatant was decanted, the tubes briefly respun and remaining supernatant removed with a pipette tip. The phage pellet was allowed to dissolve in 500µls of phage buffer for 1 hour, then the suspension transferred to a 1.5ml Eppendorf vial. 5µls of 10% w/v SDS and 5µls of 0.5M EDTA pH 8.0 were added, mixed in by vortexing, then the solution incubated at 70°C for 15 minutes to lyse the phage. The solution was then extracted twice with an equal volume of phenol-chloroform (1.14) then once with an equal volume of chloroform. 50µls of 3M sodium acetate pH 5.2 (1.17) and 1ml of ethanol were added to the supernatant, mixed in by gentle inversion and the tube placed on ice

for 1 hour. The DNA was pelleted by microcentrifugation for 15 minutes, the supernatant decanted, then the pellet washed twice with 80% ethanol. The pellet was dried briefly in a vacuum desiccator, then dissolved in 10µls of TE.

The DNAs (≃0.5µgs) were cut with EcoRI and HindIII (10 units of each) and the digestion products run on a 0.8% agarose gel (2.5). The gel was blotted (2.12) and the filters probed (2.13) with radiolabelled pLHIISB18 (2.7),(2.10).

The restriction patterns and filter autoradiographs of the phage were compared. DNA was then prepared from one isolate of each of the phage which were duplicated and from all of the isolates which were unique.

Large Scale Preparation of Recombinant Phage DNA

The recombinant phage DNA was prepared essentially as described for the preparation of EMBL3 vector DNA (2.3). However there were two differences in this protocol: i) The recombinant phage were grown on *E. coli* strain NM621 rather than Q358, and ii)) The recombinant phage were only banded once in CsCl.

Grouping and Mapping the Clones

The DNAs were digested with the following enzymes and enzyme combinations: HindIII, SalI, EcoRI, HindIII+EcoRI, HindIII+SalI and EcoRI+SalI. Where one enzyme was used in a digest the reaction conditions of:

12µls of water
2µls of 10x ReAct buffer 3 (1.33)
5µls of DNA solution (≃0.5µgs)
1µl of enzyme (10 units)

Where two enzymes were used the reaction conditions consisted of:

11µls of water
2µls of 10x ReAct buffer 3
5µls of DNA solution (≈0.5µgs)
1µl of each enzyme (10 units of each)

The samples were incubated at 37°C for ≈2hrs, then 4µls of loading dye added. The samples were heated at 65°C before being run on a 0.8% agarose gel (2.5). The gel was blotted (2.12) and the filters probed (2.13) with radiolabelled pLHIISB18 (2.7),(2.10). Comparison of the restriction and autoradiography patterns permitted the clones to be more accurately grouped. The members of three of these groups were then mapped using the restriction and autoradiography data.

3.7 Generation of M13 Clones for Dideoxy Sequencing

Strategy

Mapping studies suggested that the smallest positively hybridising fragment produced by digestion of the clone DNAs with the enzymes SalI, EcoRI and HindIII was ≈2kb. The position of possible light harvesting structural genes within this fragment was unknown. Rather than attempt to subclone and sequence such a fragment in the hope of uncovering the light harvesting genes the following strategy was used.

A representative clone from each of groups 1, 2 and 3 were digested with a variety of restriction enzymes. The enzymes used were chosen on the basis of the "ends" which their action generated; all of the enzymes produced ends which were compatible with those generated by cleavage of the M13 mp19 polylinker with the same or other enzymes. The digests were run on a gel, blotted and probed with radiolabelled pLHIISB18. Fragments were then chosen for cloning on the basis of

the following; It was assumed from the mapping studies (see previously) that fragments which hybridised positively did so because they contained a pair of genes, each gene encoding an alpha or a beta polypeptide. If the gene pair could be cleaved internally this would generate at least two positive restriction fragments. Each fragment would contain part of the gene pair, but more importantly, the internal cleavage site would make that gene immediatley accessible to sequencing. If each of these two fragments were cloned, in opposite orientations, then sequencing would result, on average, in two out of four clones generating immediate sequence from part of a light harvesting gene. This would occur because the cleavage site within the light harvesting gene fragment would be ligated directly to the polylinker of the vector. However, only fragments smaller than ≃3kb could be subcloned due to the instability of larger inserts in M13.

An alternative approach was to search for small positively hybridising fragments which could be sequenced from opposite ends with a central sequenced overlap. Both of these strategies were used.

Materials and Methods

Representative clones from groups 1, 2 and 3 (clones 9, 16 and 6 respectively) were digested in the appropriate restriction buffer with each of the following enzymes;

Enzyme	10x ReAct Buffer
ClaI, AccI, AluI(*), RsaI(*).	1 (1.32)
PstI, TaqI(*), XbaI, SstI, SstII, XhoI, HaeIII, HinfI.	2 "
KpnI, HincII, Sau3AI(*), HpaI.	4 "
SphI, PvuII.	6 "
HpaII(*)	8 "

The enzymes marked (*) recognise tetranucleotide sequences, the remainder hexanucleotide sequences.

The samples were run on a 0.8% agarose gel (2.5), photographed and blotted (2.12). The filters were probed with radiolabelled pLHIISB18 (2.7), (2.10), (2.13) then positively hybridising bands examined for size and the presence of internal cleavage sites. Suitable (as defined above) genomic clone fragments were subcloned into M13 mp19 RF as follows.

Approximately 5µgs of DNA from each representative genomic clone were digested with the the appropriate restriction enzyme. Approximately 5µgs of mp 19 DNA were digested with the enzyme which generated compatible ends for subcloning the genomic clone fragments, see overleaf.

Genomic clone	Genomic clone enzyme	mp19 enzyme
9	HpaII	AccI
9	RsaI	SmaI
9	HincII	SmaI
9	AluI	SmaI
6	SstI	SstI
16	SstI	SstI

The DNA was digested for 2 hours, heated to 65°C, to denature the enzymes then a small aliquot run on a 0.8% gel to check that digestion had taken place. The digests were extracted with phenol-chloroform (1.14), chloroform, then ethanol precipitated. The DNA was pelleted by microcentrifugation, the pellets washed in 80% ethanol then resuspended in 10µls of TE (1.20).
The genomic clone fragments were "shotgun cloned" into M13 by mixing equal quantities of the two DNAs and ligating. The ligation reactions were composed of:

4µl of water.
2µl of 5x T4 ligase buffer (supplied by BRL).
1µl of insert DNA (0.5µg).
1µl of vector DNA (0.5µg).
1µl of 10mM ATP.
1µl of T4 ligase (1 unit/µl).

In addition, a series of ligation reactions were set up which were identical to that described above with the exception that they lacked insert DNA, the lost volume being made up with water. These acted as controls for testing the efficiency of the ligation reaction. The ligation reactions were incubated at 14°C for ≈18 hours.

90µls of TE were then added to each ligation reaction to give a volume of 100µls. Four aliquots, each of 25µls (250ng of DNA) were then added to four tubes, each containing 0.3mls of JM101 competent cells (2.17). In addition, the following acted as controls for determining the transformation efficiency and the level of uncut mp19 vector:

a) 10ngs of uncut DNA were added to 0.3mls of JM101 competent cells, to determine the efficiency of transformation.

b) 0.5µgs of mp19 cut with AccI, SmaI and SstI, were each added to 0.3mls of JM101 competent cells, to gauge the level of uncut vector in the ligation reaction.

The competent cells were then transformed and plated as described previously (2.17). The plates were incubated for ≈18 hours at 37°C then placed at 4°C for 1 hour, to harden the top agar. Filters were then lifted (2.11) and screened with with radiolabelled pLHIISB18 DNA (2.8), (2.10), (2.13). Positively hybridising plaques were picked into 1ml of 2x YT broth (1.9), vortexed, then stored at 4°C.

3mls of 0.6% top agar (1.12), were added to 0.4mls of lawn cells (2.17)then the mixture poured onto an H agar plate (1.10). A loopful of the stored phage suspension was streaked out onto the 0.6% agar surface to give single plaques, then the plate incubated for 12 hours at 37°C. White, well isolated plaques were toothpicked into two duplicate arrays on JM101 lawns (see also 3.5). Each lawn was prepared by adding 5mls of 0.6% top agar, to 0.8mls of JM101 lawn cells, the mixture then being poured onto a 10 x 10cm H agar

plate. The plates were incubated overnight at 37°C for 12 hours, then placed at 4°C. After 1 hour, one of the plates was removed and duplicate filters lifted and screened as described above. Plaques from the stored array were picked after being matched to those which were shown to positively hybridise on the screened array. The plaques were picked into 2x YT broth and stored at 4°C until required for the preparation of sequencing templates (2.18).

3.8 Verifying the Source of the Genomic Clone Inserts

The purpose of this experiment was to show that the genomic clone inserts were derived from *R. acidophila* strain 7050 DNA. This involved digesting all of the clone DNAs and *R. acidophila* genomic DNA with EcoRI+SalI, running the products on a gel which was then blotted and probed with radiolabelled pLHIISB18. If the insert DNAs were derived from *R. acidophila* strain 7050, positively hybridising clone bands would be seen which would have counterparts of identical size in the genomic lanes.

Each of the clone DNAs were cleaved with EcoRI+SalI as described in 3.6. *R. acidophila* strain 7050 DNA was cleaved in the following digest:

1µl of water
5µls of 10x ReAct buffer 3
40µls of genomic DNA soln. (0.2µgs/µl)
2µls of SalI (10 units/µl)
2µls of EcoRI (10 units/µl)

The digest was incubated at 37°C for ≈2 hours, 10µls of loading dye added, then the sample heated to 65°C for 10minutes. The digest was quenched on ice, then 15µls of digest loaded into each of four wells of a 0.8% agarose gel (2.5). The clone DNAs were loaded in the remaining wells. The gel was run, blotted (2.12) then the filter

probed (2.13). Note that the genomic DNA used in this experiment was not the same isolate used to construct the library.

The result of this experiment, at the time, suggested that the library was contaminated with inserts derived from another bacterial strain. I now believe that this initial hypothesis was incorrect.

However, the following experiment was then carried out to determine which strain had contaminated the library. *R. acidophila* strain 7750 was the most likely contaminant as it could grow in the low pH of Pfennig's medium which was used to culture the cells for the library inserts. However, not wishing to confine my search to the preferred candidate, the other species of photosynthetic bacteria which were in the lab at the time of library construction were also included. These were: *R. palustris* 'DSM8252', *R. palustris* 'French', *Rhodospirillum rubrum*, and *Rb. sphaeroides* 2.4.1. It was hoped, that by a combined genomic and phage Southern blot analysis as described above, that the "alien" genomic clones could be traced to one of these photosynthetic bacteria.

These species and both strains of *R. acidophila* were streaked out for single colonies on agar plating medium (1.2). The colonies were grown under a high light regime at 28°C for approximately 2 weeks in an anaerobic GasPak jar. A GasPak hydrogen and carbon dioxide generator envelope was used to produce the anaerobic atmosphere. For each species, an isolated colony was used to innoculate an agar stab (1.2) which was grown under a high light regime for 2 weeks. The stabs were then overlaid with the appropriate medium; Pfennig's medium (1.1) for *R. acidophila* strains and Fullers medium (1.2) for all other species. Then the cultures were placed under high light for 7-14 days to allow the cells to grow out of the agar into the liquid. The liquid was transferred to a clean McCartney bottle, the bottle filled with the appropriate medium and the culture returned to the growth room for 3-4 days. The cultures were transferred to 500ml flat sided bottles and the bottles filled with the appropriate medium (see chapter 2.1 for treatment of *R. acidophila* strain 7050)

then all other cultures were grown as for *R. acidophila* strain 7750 (see 2.1). DNA was prepared from each culture as described previously (2.2).

In addition DNA was isolated from *E. coli* strain Q358. This DNA may have contaminated the vector DNA during isolation, been ligated to the vector arms and was then cross hybridising with the probe.

The genomic DNAs and representative lambda clones were cleaved with EcoRI+SalI. The digested samples were run on a 0.8% agarose gel (2.5), which was then blotted (2.12) and probed (2.7), (2.10), (2.13) with radiolabelled pLHIISB18.

The result of this experiment suggested that, in terms of size, the positively hybridising clone bands could be accounted for in full, by strain 7050 genomic DNA. However, strain 7050 shared common positive bands with strain 7750 and in terms of band intensity the picture was less clear. In order to remove the ambiguity which existed the following experiment was carried out.

A new isolate of DNA (from a single colony starting culture) was prepared for strain 7050 and 7750. 30µgs of genomic DNA from each strain were digested with EcoRI+SalI. 1µg of each representative genomic clone was also digested with these enzymes. The DNA in the digests were ethanol precipitated, washed twice with 80% ethanol, then resuspended in TE. The concentration of the DNA solutions were determined using an ethidium plate (2.6). 5µgs of genomic DNA and 50ngs of clone DNA were loaded onto a 0.8% agarose gel. The samples were loaded so that the maximum number of clones were adjacent to both *R. acidophila* DNAs ie.

Clone 7050 DNA	Clone 7750 DNA	Clone 7050 DNA	Clone 7750 DNA	Clone 7050 DNA	etc.
DNA	DNA	DNA	DNA	DNA	DNA

This would allow a more accurate comparison of the relative sizes of the positively hybridising genomic and clone bands to be made.

Their relative intensities could also be compared more accurately. In addition the different amounts of the genomic and clone DNA loaded would allow the entire filter to be used for autoradiography. If differential loading had not been carried out the filter would have to have been cut into strips and the genomic and clone strips exposed for different times.

The gel was then blotted and probed as described previously.

CHAPTER 4 LIBRARY CONSTRUCTION (RESULTS)

In order to isolate the genes encoding the light harvesting (LHII) polypeptides of *Rhodopseudomonas acidophila* strain 7050 a genomic DNA library was constructed. This involved the isolation of genomic DNA of high molecular weight, which was then partially digested with the restriction endonuclease Sau3AI. The resulting fragments were size selected by velocity gradient centrifugation, then ligated to the arms of the lambda replacement vector EMBL3. *In vitro* packaging was used to encapsidate the recombinant DNA molecules before their introduction into *E. coli*.

During the course of this work a number of *R. acidophila* genomic "libraries" had been constructed. However, all of them, when plated, had failed to generate recombinant plaques. The reason these libraries had failed was not apparent. It had been shown that a genomic library of *R. palustris* strain 1e5 DNA could be successfully constructed using the EMBL3 vector [Tadros *et al.*, 1987]. Therefore, there appeared no obvious reason why an *R. acidophila* library should not also be constructed using this vector. In order to determine the cause of failure of previous *R. acidophila* libraries, *R. acidophila* and *R. palustris* libraries were constructed in a series of parallel experiments. These experiments were performed using the same batches of Sau3AI, vector arms, ligase and packaging mixes.

The data presented within this chapter, are the results of each step in the construction of the *R. acidophila* strain 7050 library in this parallel experiment. However, these results are also typical of those which failed to lead to the production of recombinant plaques in other library constructions. Likewise, the materials and methods (Chapter 3) are specific to the construction of this particular library, but are typical of those used to construct the other failed libraries.

4.1 Isolation of High Molecular Weight Genomic DNA from *Rhodopseudomonas acidophila* strain 7050.

The isolation of size selected, 20kb inserts required high molecular weight genomic DNA as a starting material. Ideally this DNA would have consisted of intact genomes, however in reality, DNA of ≥50kb in size, had to suffice as the substrate. DNA of a smaller starting size than 50kb, after partial digestion, would have yielded an unacceptably small proportion of 20kb fragments with two BamHI site compatible sticky ends.

The isolation of genomic DNA proved considerably more difficult than initially envisaged. The main problem appeared to be that the cell wall of *R. acidophila* was refractory to digestion with lysosyme. Evidence for this came from microscopic examination of cells which had been bathed in a lysosyme solution. It was expected that the cell walls of lysosyme treated cells would degrade and (spherical) sphaeroplast would form. Instead, the appearance of the treated cells was indistiguishable from untreated cells.

Microscopic examination of cells, which had been negatively stained with nigrosin, showed that each cell was surrounded by what appeared to be a transparent envelope. This gave each cell an appearance reminiscent of a piece of frog spawn. It was thought likely that the transparent layer was a capsule or slime layer which many prokaryotes form by the deposition of organic polymers outwith the cell [Stanier *et al.*, 1975]. These exopolymers vary widely in composition. A few *Bacillus* species produce exopolypeptides, composed entirely of glutamic acid residues. With this exception, the bacterial exoploymers are polysaccharides. However, apart from the simpler homopolymeric exopolysaccharides (for example, cellulose, synthesised by *Acetobacter xylinum*), the detailed structures have rarely been elucidated [Stanier *et al.*, 1975]. It was thought possible that the presence of such a layer could interfere with the action of the lysozme by preventing the enzyme access to the wall.

In order to attempt removal of the envelope, small batches of cells were treated with solutions of: 1mg/ml α-amylase, 1mg/ml cellulase, 1mg/ml pectinase, 1mg/ml proteinase K, 1mg/ml lipase and 0.5M E.D.T.A pH 8.0. These solutions were made in 50mM glucose, 25mM Tris HCl pH 8.0, 10mM E.D.T.A pH 8.0. The solutions were tested individually and as cocktail of all six solutions. The batches of cells were bathed in these solutions for 1 hour at 37°C. The cells were then negatively stained and examined under the microscope.

0.5M EDTA pH 8.0 had not removed the envelope, but had reduced its thickness. The other solutions had no apparent effect. The subsequent addition of lysosyme followed by the addition of 10% w/v SDS solution to a final concentration of 1% w/v to each batch of cells did not result in cell lysis.

A fresh batch of cells were boiled for 15 minutes in 0.5M EDTA pH 8.0 which removed the cell envelope. However, subsequent treatment with lysosyme and SDS (as before) did not result in cell lysis. This suggested that the structure of the cell wall and not the presence of the cell envelope was preventing lysozyme action.

Lysosyme acts by passing through the wall and degrading the wall polymers from within (Ian Hunter, personal communication). It was considered possible that the enzyme was capable of degrading the wall but was too large a molecule to enter the cell through the wall matrix and carry out its action. The following method was tried to force the lysozyme into the cell. A batch of cells were partially dehydrated by soaking in 5M NaCl for 1 hour. The cells were then placed in a solution of 25mM EDTA pH 8.0, 100μg/ml lysosyme for 30 minutes. It was hoped that the influx of water caused by rehydration of the cells would result in the lysozme being "dragged" into the cell. However, subsequent treatment of the cells with SDS as before did not result in cell lysis. Whether this failure was due to the refractile nature of the wall, or because the lysosyme never managed to enter the cell by this method is unknown.

The only method which has been found to be reproducible for lysing *R. acidophila* cells is that described in chapters 2.1. The key to the success of this method appears to lie in the age of the cells. Cells which are young and growing logarithmically, as judged by looking at the density of the suspension by eye, in general, lyse with relative ease. Though it must be said that they never lyse as readily as cells from other bacterial species. Cells which are in late log or stationary phase will not lyse by this method. A method which I have tried once for older cells, and which appears to work, is to treat the cells as described for young cells but leave them at 37°C for 4 days before adding the 10% w/v SDS.

It is still unclear why older *R. acidophila* cells are so refractile to lysosyme action. During the course of this work all other photosynthetic bacteria, irrespective of age, proved easy to lyse by the method described. It may simply be that old cell walls are a poor substrate for lysosyme action, some molecular change occuring within the walls with age. It may also be that the method used to force lysosyme into the cells does not work and that indeed the wall does form an immpenetrably tight mesh which prevents lysosyme entering the cells. Whatever the reason(s), young cells appear the most suitable substrate for lysosyme action.

Genomic DNA was isolated and the concentration determined spectrophotometrically, to be 0.4µg/µl. The size of the uncut genomic DNA was determined on a 0.4% gel (data not shown) then 5µgs digested with each of EcoRI, HindIII, BamHI, SalI and PstI. The digests were run with an uncut sample (0.8µgs) on a 0.4% agarose gel, see figure 4.1.

Discussion

From figure 4.1 it can be seen that the uncut genomic DNA is of high molecular weight (Lane 1). Electrophoresis on a 0.4% gel suggested that the vast majority of the fragments were greater than 50kb but probably less than 75kb.

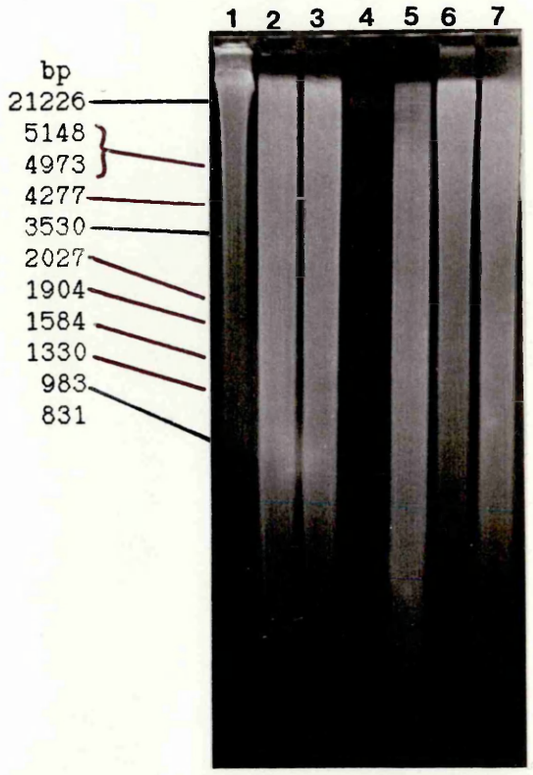


Figure 4.1

Genomic DNA isolated from *R. acidophila* strain 7050.

Lane 1: Uncut genomic DNA.

Lane 2: Genomic DNA cut with EcoRI.

Lane 3: Genomic DNA cut with HindIII.

Lane 4: λ cI857 DNA cut with EcoRI+HindIII.

Lane 5: Genomic DNA cut with BamHI.

Lane 6: Genomic DNA cut with SalI.

Lane 7: Genomic DNA cut with PstI.

During the course of this work genomic DNA was isolated from *R. acidophila* cells on a number of occasions. Although every care was taken to minimise mechanical shearing and nuclease digestion, the DNA isolated was always of smaller size than that expected (≈100kb). Whether this small size was due to mechaical and/or enzymatic degradation is unclear but it may be a side effect of the poor cell lysis described earlier.

As the DNA fulfilled the requirements of size (>50kb) and ability to be cleaved with a range of restriction enzymes (Lanes 2, 3, 5, 6 and 7) it was used as the starting material for generating 20kb inserts.

4.2 Optimisation of the partial digestion conditions

Inserts, 20kb in size, were generated by partial cleavage of genomic DNA with Sau3AI. The optimal conditions for generating fragments of this size were determined by calculation, then experimentally.

Calculation (chapter 3.2) suggested that digestion of 8µgs of DNA with 0.2 units of Sau3AI for 30 minutes would yield the maximum number of 20kb fragments. These conditions were tested experimentally by digesting 8µgs of DNA with 0.2 units of enzyme for between 3 and 120 minutes. The result of this digestion is shown in figure 4.2.

Discussion

After 3 minutes of digestion the average size of the DNA fragments has decreased from >50kb (Lane 1) to between ≈20-30kb (Lane 2). After 7 minutes the maximum number of 20kb fragments had been generated (Lane 2). 20kb fragments are also found after 15 and 30 minutes but in considerably lower amounts (Lanes 3 and 4). Note that after 120 minutes (Lane 7) all of the DNA fragments were smaller than 20kb. This suggested that inserts of 20kb could be generated from all regions of the *R. acidophila* genome. If this were not the

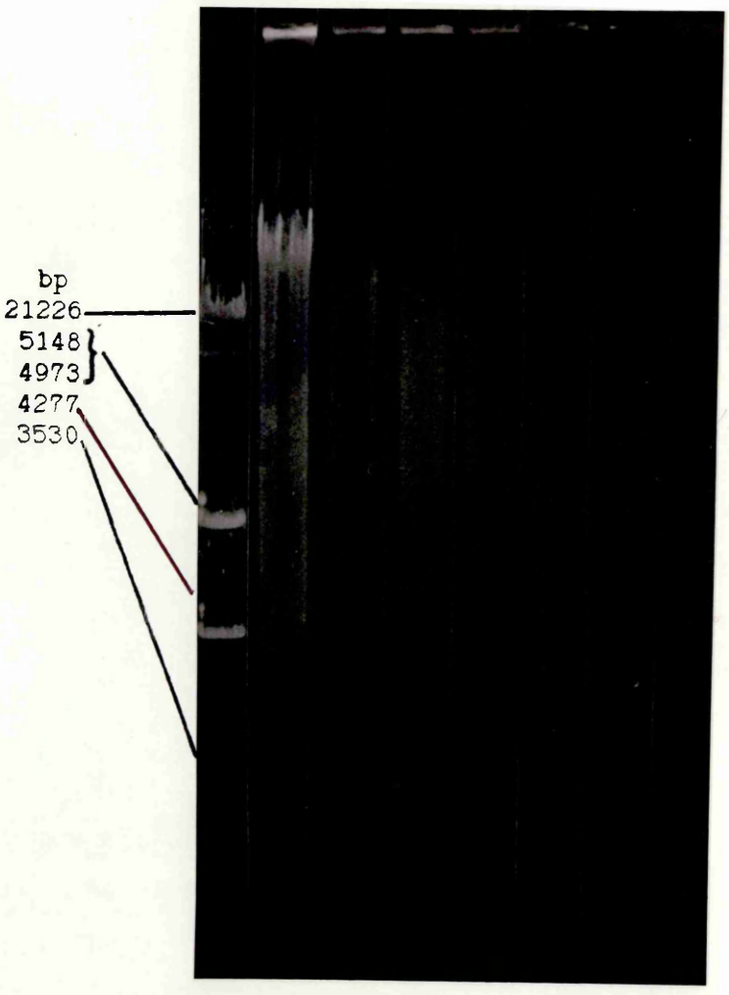


Figure 4.2

Small scale Sau3AI digest to optimise partial digestion conditions.

Lane 1: λ cI857 DNA cut with EcoRI+Hind III.
Lane 2: Uncut *R. acidophila* genomic DNA.
Lane 3: Genomic DNA after 3 minutes digestion.
Lane 4: Genomic DNA after 7 minutes digestion.
Lane 5: Genomic DNA after 15 minutes digestion.
Lane 6: Genomic DNA after 30 minutes digestion.
Lane 7: Genomic DNA after 60 minutes digestion.
Lane 8: Genomic DNA after 120 minutes digestion.

Digestion ratio was 8μgs of DNA and 0.2 units of Sau3AI

Note this gel was originally loaded in the incorrect order. The gel photograph has been re-ordered to correct this. The gel was originally loaded with the lane order; 1 7 8 3 4 5 6 2.

case a complete library could not have been constructed using this strategy.

The reaction had been carried out using 0.2 units of enzyme to digest 8μgs of DNA. This suggested that a large scale digestion, using the ratios of 1 unit of enzyme to 40μgs of DNA, would yield the maximum number of 20kb fragments if the reaction were carried out for 7 minutes.

Calculation had suggested that the optimal digestion time was 30 minutes, but experiment showed that for this isolate of DNA the time was 7 minutes. Why should there be such a large discrepency between the calculated and experimental values?

Leaving the more obvious reasons, such as pipetting inaccuracies aside, it should be remembered that certain assumptions were made during the calculation. It was assumed, for example, if 1 unit of Sau3AI cuts 1μg of lambda DNA to completion in 1 hour that it would have an identical effect on 1μg *R. acidophila* DNA. However there is no reason to believe that there is the same number of Sau3AI sites per unit of DNA length in both lambda and *R. acidophila* DNA. Likewise, it was stated that, on average, Sau3AI will cut a piece of random sequence DNA once every 250bp. However, *R. acidophila* DNA is not a piece of random sequence DNA. A piece of random sequence DNA would have a guanine+cytosine content of 50%, whereas that of *R. acidophila* is 65% [Pfennig, 1969]. In addition when the DNA concentration was measured, it was the concentration of all the fragments irrespective of size which was measured. As a proportion of the fragments were already too small to generate 20kb fragments then the measured DNA concentration was greater than the concentration of DNA capable of generating 20kb fragments. It is in part for these reasons that the calculated conditions can only be used as a rough guide for determining the optimal conditions by experiment.

4.3 Large scale genomic partial digest

The experimentally determined partial digestion conditions were used to cleave 800µgs of genomic DNA with 20 units of Sau3AI (chapter 3.2). However, to counteract pipetting errors and effects of scaling up the digest, 500µl aliquots were removed at times close to that determined as the optimum. A sample of each timed aliquot was run on a 0.4% agarose gel, shown in figure 4.3.

Discussion

Aliquots taken at 3, 4, 5, 6 and 7 minutes (Lanes 3-7) were considered to contain useful proportions of 20kb fragments and were pooled. Aliquots taken after this time had very few fragments which were equivalent in size to the 23.130kb λHindIII fragment (Lanes 1 and 13).

The variations in the optimal time for digestion when scaling up the reaction can be seen by comparing the size distribution of fragments in the 15 minute aliquot of the test digest (figure 4.2, Lane 5) with the 12 minute aliquot of the large scale digest, Lane 12. These differences were probably a result of pipetting errors when adding the enzyme.

4.4 Partial Digest Size Fractionation By Velocity Gradient Centrifugation.

The partially digested DNA fragments were resolved on the basis of size, by passage down a sucrose velocity gradient, as described in chapter 3.2. The gradient was fractionated and aliquots from every other fraction run on a 0.4% agarose gel, see figure 4.4a.

Discussion

The agarose gel shown in figure 4.4a suggested that fractionation had been successful. The DNA bands were relatively "tight" with

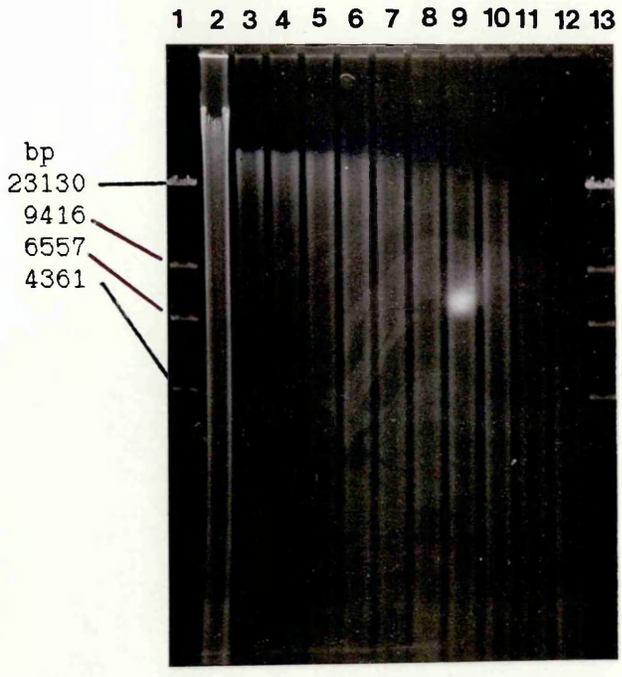


Figure 4.3

Large scale partial digestion of *R. acidophila* genomic DNA with Sau 3AI.

Lane 1: λ cI857 DNA cut with Hind III.
Lane 2: Uncut *R. acidophila* genomic DNA.
Lane 3: Genomic DNA after 3 minutes digestion.
Lane 4: Genomic DNA after 4 minutes digestion.
Lane 5: Genomic DNA after 5 minutes digestion.
Lane 6: Genomic DNA after 6 minutes digestion.
Lane 7: Genomic DNA after 7 minutes digestion.
Lane 8: Genomic DNA after 8 minutes digestion.
Lane 9: Genomic DNA after 9 minutes digestion.
Lane 10: Genomic DNA after 10 minutes digestion.
Lane 11: Genomic DNA after 11 minutes digestion.
Lane 12: Genomic DNA after 12 minutes digestion.
Lane 13: λ cI857 DNA cut with HindIII.

Digestion ratio was 1 unit of enzyme to 40μgs of genomic DNA.

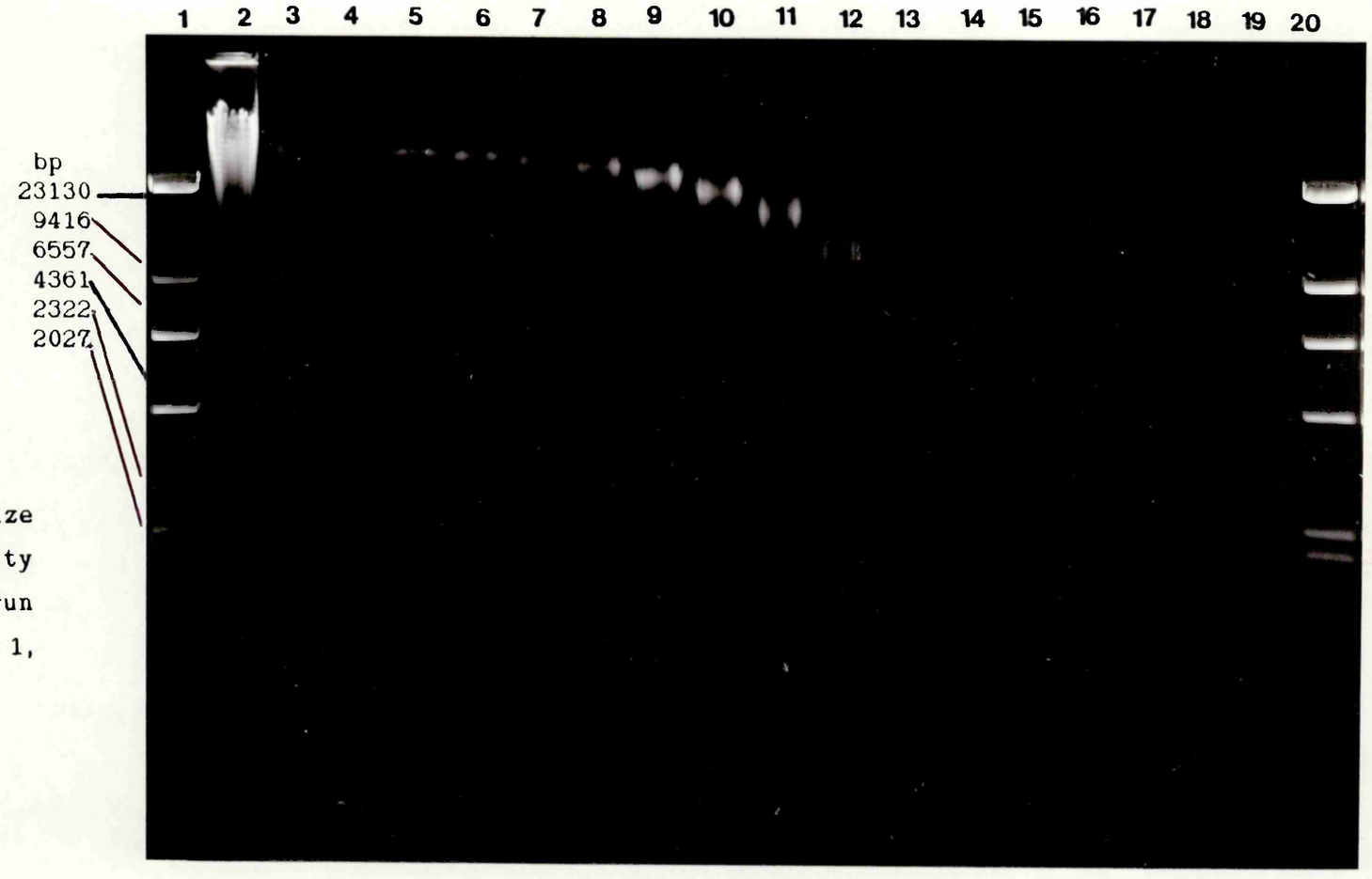


Figure 4.4a

DNA solution fractions after size resolution over a sucrose velocity gradient. Every other fraction was run on this gel. ie. lane 3 = fraction 1, lane 4 = fraction 3 etc.

Lane 1+20: λ HindIII.

Lane 2: Uncut *R. acidophila* genomic DNA.

Lanes 3-19: Samples of every other fraction (fraction numbers 1-33) after size resolution.

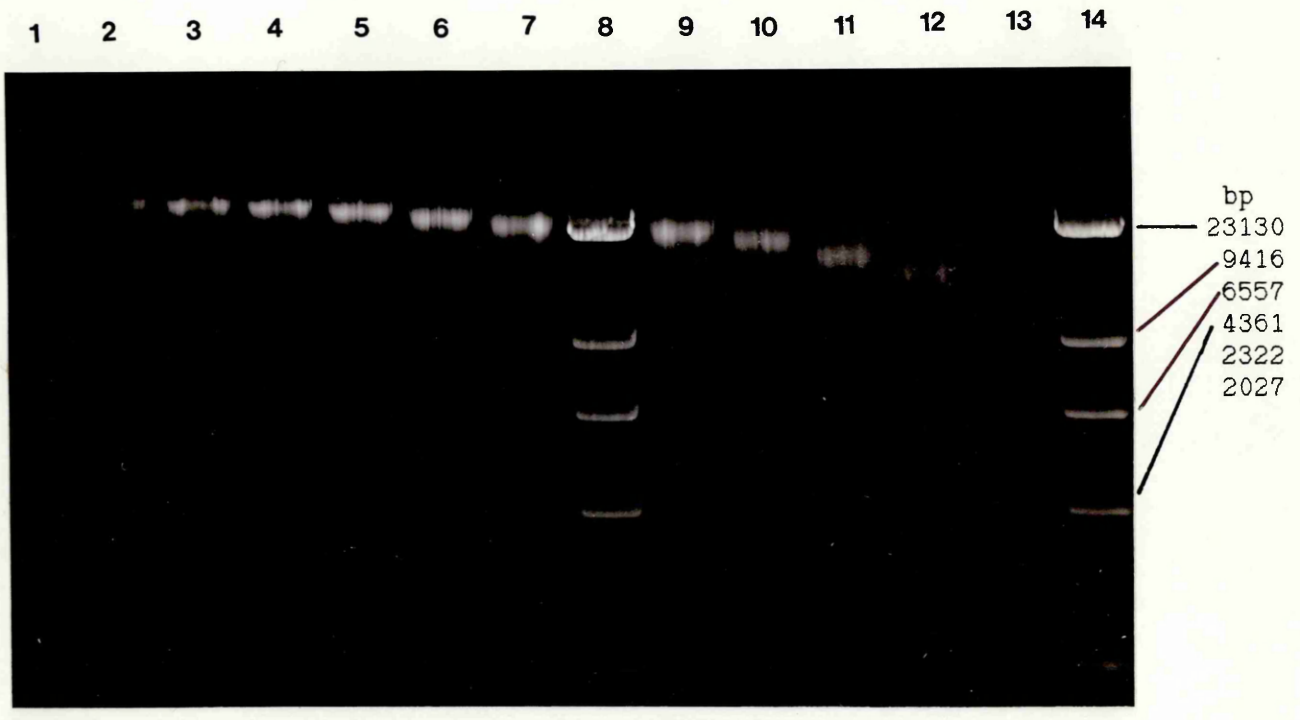


Figure 4.4b

Fractions 9-19 shown in Figure 4.4a after dialysis.

Lanes 1,8 & 14: λ cI857 DNA digested with HindIII.
Lanes 2 - 7 : Fractions 9 - 14.
Lanes 9 - 13 : Fractions 15 -19.

Fraction 16 (Lane 10) was kept and used as a source of insert DNA.

little smearing above or below the bands of greatest intensity. This suggested that each fraction consisted of a population of fragments which had little size variation. The gel acted as a rough guide as to which fractions contained suitable sized fragments for ligation to the vector. The gel was not more accurate in this respect because the samples and the lambda size markers contained differing concentrations of sodium chloride. Such differential salt concentrations may have resulted in indentical sized genomic and lambda DNA fragments migrating at different rates.

Fractions 9-19 (Lanes 7-12), which contained fragments in the correct size range were dialysed against TE. An aliquot of each was removed and sufficient 100mM NaCl added to give the samples the same concentration of sodium chloride as the lambda markers. The samples were then loaded onto a 0.4% agarose gel, see figure 4.4b.

On this gel fraction 16 (Lane 10) appeared to consist of a population of fragments of ≈20kb ie. slightly smaller than the 23.13kb lambda size marker. The DNA within this fraction was ethanol precipitated and used as inserts for the construction of the library.

4.5 Preparation of vector arms

Construction of the library required the preparation of EMBL3 vector arms. The vector was prepared as described in Chapter 2.4 and the arms prepared as described in chapter 3.2.

The enzymes EcoRI, BamHI and BamHI+EcoRI were tested in separate digests each containing 0.25µgs of vector DNA. Figure 4.5 shows the result of this digest. Note that the integrity and restriction pattern of the vector DNA had been checked before this digest was carried out. This data is not shown.

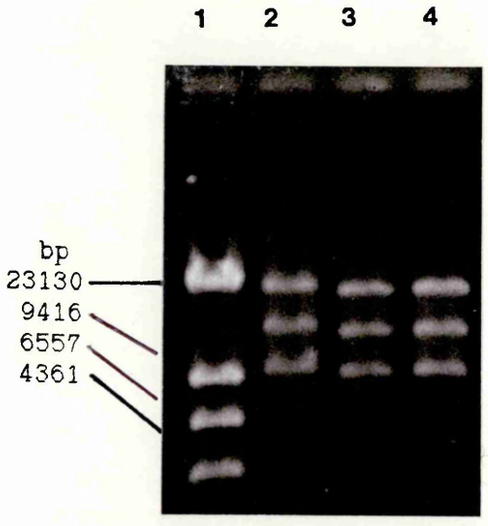


Figure 4.5

Testing the restriction enzymes EcoRI and BamHI by digesting λ EMBL3 DNA.

Lane 1: λ cI857 DNA digested with HindIII.

Lane 2: λ EMBL3 DNA digested with EcoRI.

Lane 3: λ EMBL3 DNA digested with EcoRI+BamHI.

Lane 4: λ EMBL3 DNA digested with BamHI.

Discussion

Comparison of the cleaved vector DNA and the lambda size markers suggested that the approximate sizes of the vector fragments were 20, 15 and 9kb. These sizes compare favourably with the predicted sizes of 20,280bp (left arm), 13,727 (stuffer fragment) and 9218 (right arm) for BamHI cleaved vector [Frishauf *et al.*, 1983; Kaiser and Murray, 1985]. The apparent differences in size between the fragments generated by single and double digestion were considered to be an anomaly of the gel. The differences did not reflect the loss of the 10bp polylinker through double digestion. If this had been the case it would have led only to size differences in stuffer fragment bands and would not have affected the sizes of the vector arms. It is also highly unlikely that a 20bp (two, 10bp polylinkers) difference in size, would have led to a visible difference in the migration rates of DNA fragments of the sizes considered here.

4.6 Ligation of vector arms to genomic inserts

The size selected genomic inserts were added to the annealed vector arms then ligated. A small aliquot of the completed ligation reaction was then removed. This aliquot, BamHI+EcoRI digested EMBL3 DNA and uncut lambda cI857 DNA was loaded onto a 0.4% agarose gel, see figure 4.6.

Discussion

It was noticed that when the aliquot of the ligation reaction had been removed, that the solution was viscous. This suggested that the ligation reaction had been successful and had generated concatenates of recombinant EMBL3 molecules.

The BamHI+EcoRI cleaved EMBL3 vector DNA shown in figure 4.6 was an aliquot taken from the large scale digestion described in chapter 3.5. Both the cut vector and the lambda cI857 DNA have undergone a

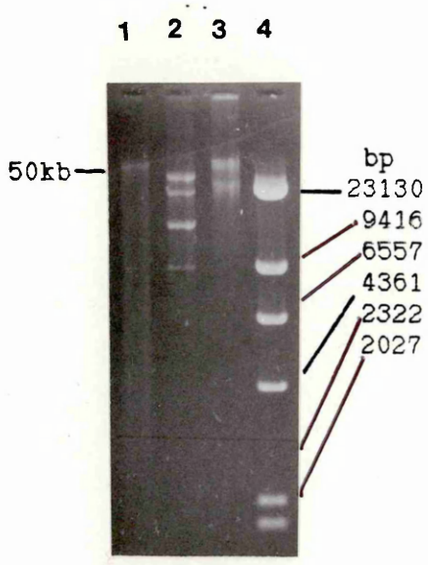


Figure 4.6

Minigel showing ligation of λ EMBL3 arms to genomic inserts.

Lane 1: Uncut λ cI857 DNA.

Lane 2: λ EMBL3 DNA digested with BamHI+EcoRI (See also Figure 4.5).

Lane 3: Ligation of λ EMBL3 vector and genomic insert DNA.

Lane 4: λ cI857 DNA digested with HindIII.

small amount of degradation during the running of the gel. The cause of this degradation is not known, however, when these samples were run on another gel (data not shown) degradation was not apparent.

It is also apparent that the cleaved EMBL3 has either not been heated to 65°C for long enough to separate the cohesive ends of the molecule, or not placed on ice quickly enough after the heating process. This accounts for the four restriction bands found rather than the three expected. The smallest and faintest band corresponds to the 9kb right arm. Its faintness probably reflects the fact that a proportion of the arms have reannealed to the 20kb left arm and have given rise to the largest (29kb) band.

The ligation reaction has generated DNA concatenates, signified by the high molecular weight DNA present in the well slot. The remainder of the DNA appears to have formed into either ≈50kb or ≈30kb molecules. The former are probably probably monomeric vector-insert recombinants (53kb), insert dimers (40kb) and left arm-insert recombinants (≈40kb). The latter ≈30kb molecules are probably stuffer dimers (≈27kb), left-right arm recombinants (≈29kb) and insert-right arm recombinants (≈29kb).

As the ligation reaction appeared to have been successful an aliquot of the recombinant DNA was packaged as described in chapter 3.2.

Chapter 5 TITERING THE *R. acidophila* LIBRARY

After the recombinant DNA had been packaged, 500µls of TE were added to the packaging mixture. This gave the packaged phage solution a total volume of 550µls. The phage titre of the solution was then determined by plating serial dilutions of the solution onto *E. coli* strains Q358, Q359 and NM621. Both the *R. acidophila* and *R. palustris* libraries were diluted then plated in this way (chapter 3.3). In addition, the packaged EMBL3 phage, used as a control for packaging efficiency, were also diluted, plated and the titre of the solution determined.

At the time the library was constructed, the recommended *E. coli* strains for the propagation of the EMBL3 vector and its recombinant derivatives were Q358 and Q359 respectively. Both of these strains had been used previously to plate "failed" *R. acidophila* libraries. Strain NM621 was provided by Noreen Murray [Dept. of Molecular Biology, Edinburgh University] and was used because it was reputed to permit enhanced recovery or recombinant phage sequences.

The three phage types were plated on these three strains as described in chapter 3.3, then the number of plaques present on each plate determined. Photographs of the plates are shown in figure 5.1. However, please note that only replica plate 1, with the higher concentrations of plated phage is presented ie. V%=0.5 for *R. acidophila*; V%=0.05 for *R. palustris* and EMBL3 are shown. The V%age values are the plated volumes of the diluted phage solutions expressed as a percentage of 550µls (the final volume of the diluted packaging reaction). For example V%=0.05 is: 0.05% of 550µls, equivalent to ≈0.3µls of the diluted phage solution.

The numbers of plaques found on both replica plates at the lowest concentrations are also presented in table 5.1.

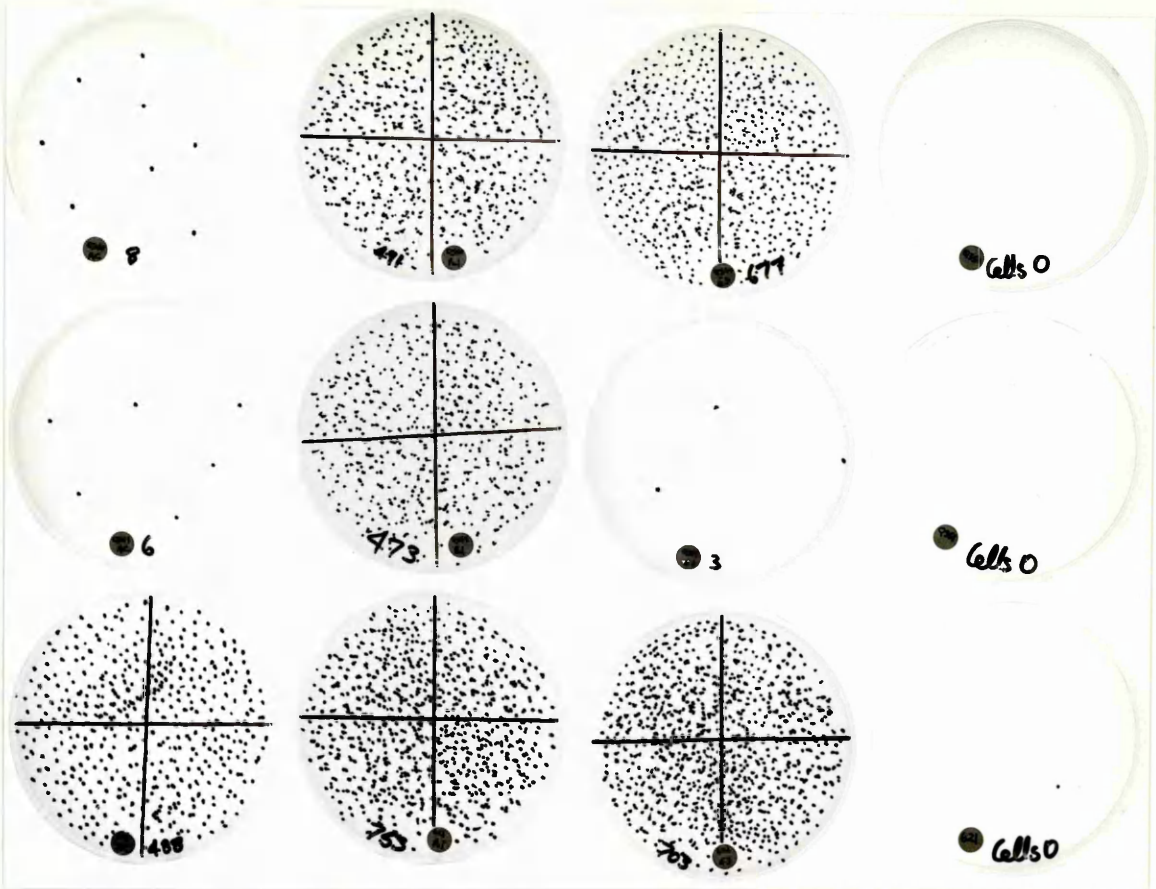


PLATE KEY

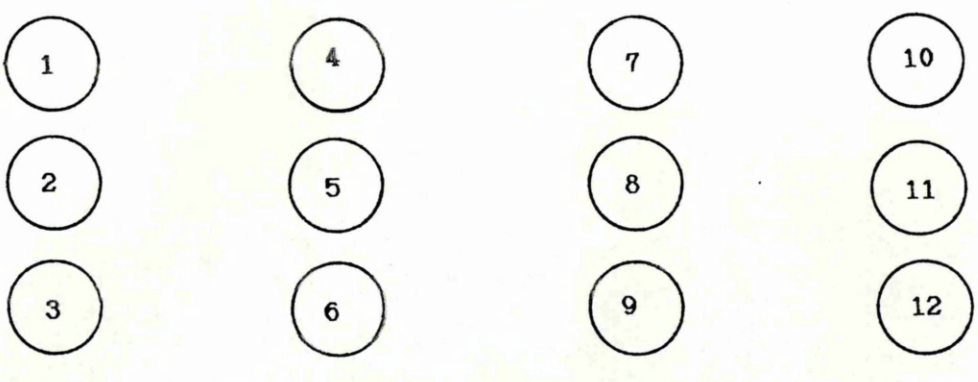


Figure 5.1

Plates 1, 2 & 3: *R. acidophila* strain 7050 genomic library plated on *E. coli* strains Q358, Q359 and NM621 respectively.

Plates 4, 5 & 6: *R. palustris* strain French genomic library plated on *E. coli* strains Q358, Q359 and NM621 respectively.

Plates 7, 8 & 9: Wild-type λ EMBL3 vector.

Plates 10, 11 & 12: *E. coli* strains Q358, Q359 and NM621 plated as a control for phage contamination.

The plaques on each plate have been marked with black felt-tip pen. The numbers of plaques on each plate are also shown.

	Source of Phage Insert											
	R. acidophila				*R. palustris*				EMBL3			
V%	0.5		0.05		0.05		0.005		0.05		0.005	
Replica No.	1	2	1	2	1	2	1	2	1	2	1	2
E. coli strain												
Q358	8	9	0	0	491	525	23	29	677	813	86	110
Q359	6	3	0	0	473	503	31	31	3	0	0	0
NM621	488	573	65	39	753	779	57	83	703	665	81	85

Table 5.1

Shown above are the numbers of plaques found when the three types of phage were plated on the three bacterial strains.

V% is the plated volume of the packaged phage expressed as a percentage of 550µls.

Replica Nos. 1 and 2 are duplicate plates containing the same volume of packaged phage solution.

Discussion

The *R. acidophila* library when plated on strain NM621 yields ≈100 fold more plaques than when plated on either strain Q358 or Q359. It is probably fair to say that if strain NM621 had not been available, then this *R. acidophila* library would have been considered as another failure.

The numbers of *R. acidophila* recombinants recovered on strains Q358 and Q359 were approximately the same as those previously recovered when other *R. acidophila* libraries ("failures") had been plated on these strains. This suggested that these previous failures may in fact have been useful, representative libraries. Fortunately, a small quantity of one of these libraries had not as yet been consigned to the dustbin. When plated on strain NM621 the plaque yield was approximately the same as that of the newly constructed library (data not shown). This suggested that the low titre of previous libraries had not been caused by library construction errors but was instead due to the phage being plated on an unsuitable host.

Although not quite so dramatic, it can also be seen from table and figure 5.1, that the numbers of *R. palustris* recombinants recovered when plated on NM621 is almost twice as great compared to when plated on either of the other two strains. The difference in plating efficiency of the vector on strains Q358 and NM621 was, as expected, negligible. Non-recombinant phage should not, and to all intents and purposes did not, plate on strain Q359.

R. acidophila recombinant phage are not unique in their inabilty to be propagated on particular *E. coli* strains. There is a growing amount of evidence that the problem of "missing clones" in genomic DNA libraries is relatively widespread. It is only the degree of the problem which appears to vary. For example, there is an inability of 9% of recombinant phage carrying human genomic DNA fragments and over 80% of recombinants carrying *Physarum polycephalum* DNA

fragments to be propagated on wild type *E. coli* hosts (such as Q358 and Q359). It is believed that these losses are due, in part, to the mechanisms *E. coli* posesses for identifying foreign DNA and destroying it. Although these mechanisms may provide the cell with a useful defence against invading DNA molecules, it clearly can be a significant problem in creating representative genomic libraries.

The host strains which were used to plate the libraries are derivatives of *E. coli* K12. Wild type *E. coli* K12 produces a type I restriction enzyme, EcoK, which cleaves DNA containing unmodified recognition sequences. *E. coli* K12 protects its own DNA from cleavage by producing a methylase which modifies susceptible restriction sites [Adams and Burdon, 1985]. The *hsd* region in *E. coli* K12 encodes the products neccessary for this restriction modification activity and consists of three genes *hsd*R, *hsd*M and *hsd*S [Sain and Murray, 1980]. Unmodified phage DNA will, however, be subject to cleavage as will the inserts of recombinant phage DNA molecules. Restriction of recombinant molecules will occur even if the vector has been grown in a modifying strain. This can be avoided by propagating the recombinant phage in a restriction deficient (r_K^-) host. *hsdR*$^-$ strains are defective in restriction but not modification (r_K^- m_K^+).

In addition to EcoK, two further restriction systems, McrA and McrB, have been described [Raleigh and Wilson, 1986; Raleigh, 1987]. The two Mcr (for modified cytosine restriction) systems each restrict DNA containing 5-methylcytosine but function independently of each other as defined by their specificity for different methylated sequences. Since it has been reported that the DNA of many organisms contains 5-methylcytosine, [Ehrlich and Wang, 1981] attempts to clone this type of DNA in an Mcr$^+$ strain would result in low recombinant yields. The Mcr systems have been shown to be functional in many common laboratory strains of *E. coli* including Q358 and Q359 [Whittaker *et al.*, 1989]. This suggests that plating recombinant phage on *E. coli* strains defective in both Mcr functions would enhance recovery.

The relative genotypes of the *E. coli* strains used to propagate the recombinant phage are presented below.

Q358	hsdR-k	hsdM+	mcrA+	mcrB+
Q359	hsdR-k	hsdM+	mcrA+	mcrB+ P2.
NM621	hsdR-k	hsdM+	mcrA-	mcrB- recD-.

All of these strains lack the EcoK restriction system (hsdR-k), but possess the EcoK site modifying methylase (hsdM+). This permits recombinant phage grown in these strains to be tranferred to, and propagated in, *E. coli* strains possessing EcoK restriction activity.

It is not known if *R. acidophila* DNA is methylated, however, if it were, the greater phage recovery when plated on NM621 could, in part, be accounted for by the lack of Mcr restriction activity. The methylation status of *R. acidophila* DNA could have been determined by digestion with restriction enzymes which recognise the same piece of sequence but differ in their restriction activities when the DNA is methylated eg. HpaII and MspI both recognise and cleave the sequence 5'-CCGG-3'. However, HpaII and MspI will not cleave when the central cytosine and external cytosine are methylated respectively.

In addition to lacking restriction activity strain NM621 is also recD-. recD is one of three genes (*recB, recC* and *recD*) which codes for Exonuclease V (ExoV) an enzyme with multiple activities which promotes genetic recombination. ExoV binds to double stranded DNA terminii then rapidly moves along the DNA unwinding it and producing single stranded loops. During this unwinding which requires ATP hydrolysis, the enzyme cuts the DNA and releases single stranded DNA fragments. Cleavage occurs at high frequency at Chi sites, 5'-GCTGGTGG-3', which stimulate genetic recombination in their vicinity. The DNA-dependent ATPase, recombination and nuclease activities reside with the RecB, RecC and RecD activities respectively.

In strain NM621 the exonuclease activity has been removed (recD-), therefore the phage DNA can replicate by the σ or rolling circle mode. This system is much more efficient for generating viable phage as a) linear concatenates are more efficient substrates for the packaging machinery than di- or multimeric circles, and b) it does not rely on the host recombination systems to form multimeric concatenates.

Though these factors undoubtedly enhance recombinant phage recovery over a recD$^+$ strain, it does not explain why there appears to be a differential recovery of *R. acidophila* and *R. palustris* recombinant phage when both are plated on strain NM621. This can be seen by looking at the total number of phage present in the diluted packaging mix, see table 5.2.

In this table the approximate titre of the packaged phage suspensions were determined from the data presented in table 5.1. The method of calculation used is given in the following example;

The numbers of *R. acidophila* plaques on NM621 replica plates 1 and 2 when 0.5% (V%=0.5) of the library was plated were 488 and 573.

$\Rightarrow$ 1% of the phage solution contained 488+573=1061pfu.

$\Rightarrow$The total number of phage in the solution = 1061x100 $\simeq 1.0 \times 10^5$pfu.

Note, the total numbers of phage were determined using only the %V values of 0.5% for the *R. acidophila* library, and 0.05% for the *R. palustris* library and EMBL3 vector.

Comparison of the numbers of *R. acidophila* and *R. palustris* recombinants recovered when plated on strain NM621 suggest that the recovery of R. acidophila recombinants is at least tenfold lower than their *R. palustris* counterparts.

	Source of phage insert		
	R. acidophila	*R. palustris*	EMBL3
E. coli strain			
Q358	1600	1.0×10^6	1.5×10^6
Q359	900	9.7×10^5	3000
NM621	1.0×10^5	1.5×10^6	1.3×10^6

Table 5.2

This table shows the projected number of phage present in the three packaged phage solutions when plated on the three different bacterial strains.

It has been reported that inserts containing palindromic sequences, or inverted repititions are unstable or completely inviable in lambda. It is not clear, however, whether palindromes are inherently unstable, or if they impose severe problems which provide a strong selection for deleted derivatives. In order to determine if palindromic insertions were responsible for the reduced recovery of *R. acidophila* recombinants, an aliquot of the R. acidophila library was plated (as described previously) onto *E. coli* strains NM621 and DL491. The latter strain is a derivative of *NM621, which contains an sbcC* mutation, this mutation being relevant to the stabilisation of palindromic sequences.

The number of plaques generated on the two strains was virtually the same (data not shown). It seemed unlikely, therefore, that the presence of palindromic sequences had resulted in the lower titre of the *R. acidophila* library.

This result presented a dilemma: did this mean that the library was representative but of low titre, or was it of high titre, but 90% of the recombinants were, for some reason not being propagated? If the latter was the case then a second dilemma was presented: were the phage which formed plaques representative of the library, because the the proportion which did not grow were "(destroyed)" in a random fashion, or did they consist of an unrepresentative sub-population of the library?

A low titre, representative library or a high titre non-representative library could have been generated because of a number of factors, these include:

a) The ratio of arm:insert DNA not being optimal during the ligation reaction. The DNA concentrations of the arms and inserts were determined using ethidium bromide plates. This is not the most accurate method of determining DNA concentrations but is the only type of method available for measuring such small quantities. Judgemental errors in determing these concentrations may have

resulted in a sub optimal ratio of these two reacting species. Whereas the judgement may have produced an optimal ratio for the construction of the *R. palustris* library.

b) The *in vitro* packaging mixes were made from E. coli strains which possessd both EcoK+ and Mcr+ restriction activity. The levels of the restriction activity in the packaging mixes was unknown, but any activity is likely to have led to a reduction in the phage titre. Whether this led to misrepresentation or simply a lower titre would depend on the distribution of the restriction sites around the R. acidophila genome. For example, if susceptible sites were not distributed evenly throughout the genome, particular inserts containing these sites would be more likely to be restricted resulting in misrepresentation of these species. However, if the sites were evenly distributed among inserts there would be a general reduction across the entire recombinant population and representaion may be maintained.

c) Although strain NM621 lacked EcoK and Mcr functions it is possible that an as yet undefined restriction system is functional in this strain. Such a system could result in the restriction of recombinant *R. acidophila* phage. Whether the reduction in phage would be representative or unrepresentative would again depend on the distribution of susceptible sites within the *R. acidophila* genome.

d) The inserts were smaller/larger in reality than suggested by the gel data. If this was the case the packaging constraints of lambda would have resulted in a lower titre of library than expected. Though such a library would probably be representative. Ideally the sized fractions above and below that selected would also have been ligated to the vector and packaged.

Please note that the order in which I have presented these possibilities does not reflect the likelyhood of them being responsible for the reduced library titre. Indeed, which, if any,

of these resulted in the lower titre of the *R. acidophila* library, was and is still unclear.

However, recall that the titre of the new and the "failed" R. acidophila libraries were approximately the same when plated on strain NM621. This sugested that the titre achieved perhaps was as high as was possible given the type of packaging mixes, host bacteria and method of selecting inserts available at that time. It was not known if the library was representative of the entire *R. acidophila* genome. If it was then there were more than sufficient clones to encompass the genome as only ≃1000 independant clones were required for a complete library and 1.0×10^5 phage had been generated. If the library was unrepresentative it could only be hoped that the light harvesting genes were not encoded in those phage which were "missing". Evidence will be presented in later chapters which suggests that the light harvesting genes were recovered after plating the *R. acidophila* recombinants on strain NM621.

Chapter 6 SCREENING THE *R. acidophila* LIBRARY

The amplified library was plated at low density on strain NM621 and duplicate filters lifted from the plates. The filters were probed with radiolabelled pLHIISB18, as described in chapter 3.3. An example of the result of this screening is shown in figure 6.1. Approximately 10,000 plaques were screened, and of these 81 were picked as positives.

The positives phage were streaked out for single plaques then picked into duplicate arrays. A filter was lifted from one of the duplicates of each array, then screened as described above. An example of the result of screening one of these filters is shown in figure 6.2. This secondary screening reduced the number of genuine positive plaques to 45. This reduction occurred because some of the plaques picked as positives in the primary screening did not hybridise positively in the secondary screening, see rows 1, 3, 4, 5, 7 and 10 of figure 6.2. Representative positive plaques from the rows of the unscreened duplicate plates were then picked and served as purified phage stocks.

To determine the level of background plaques due to non-recombinant phage, filters were lifted from the *R. acidophila* library and probed with either radiolabelled stuffer fragment, or radiolabelled *R. acidophila* genomic DNA. As a control filters were also lifted from known non-recombinant EMBL3 plaques and probed with radiolabelled EMBL3 stuffer. This gave a standard against which to gauge the appearance of EMBL3 probed back with its own stuffer. Results typical of those found during these screenings are shown in figure 6.3. As an additional control radiolabelled *R. acidophila* genomic DNA was used to probe EMBL3 plaques. As expected, the EMBL3 plaques did not hybridise positively, and therefore the autoradiographs have not been presented.

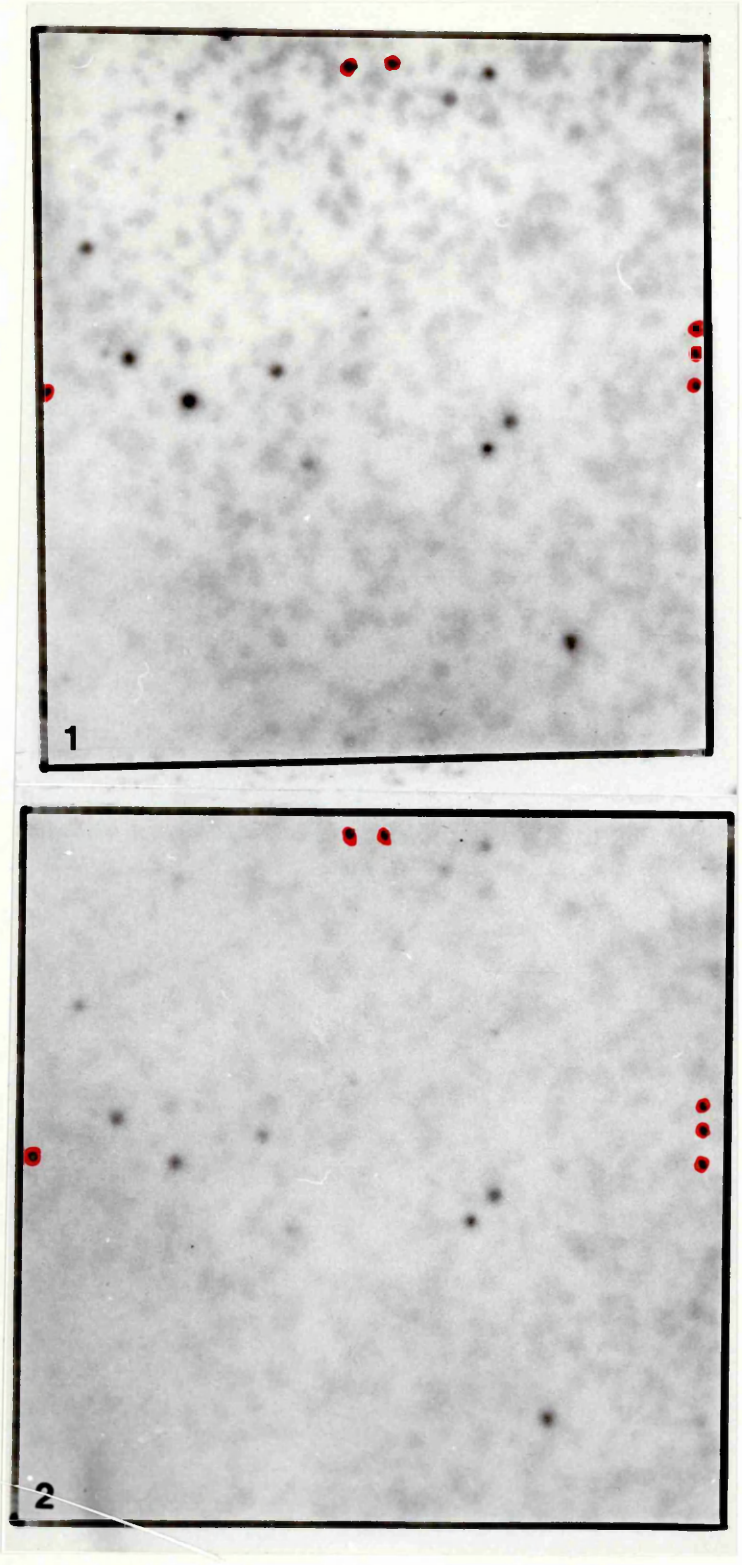


Figure 6.1

The library was plated, replica filters lifted and screened with radio-labelled pLHIISB18.
Above are shown autoradiographs of replica filters 1 + 2 taken from the same plate. Genuine positive plaques are duplicated on both autoradiographs. . = orientation marks on the autoradiographs.

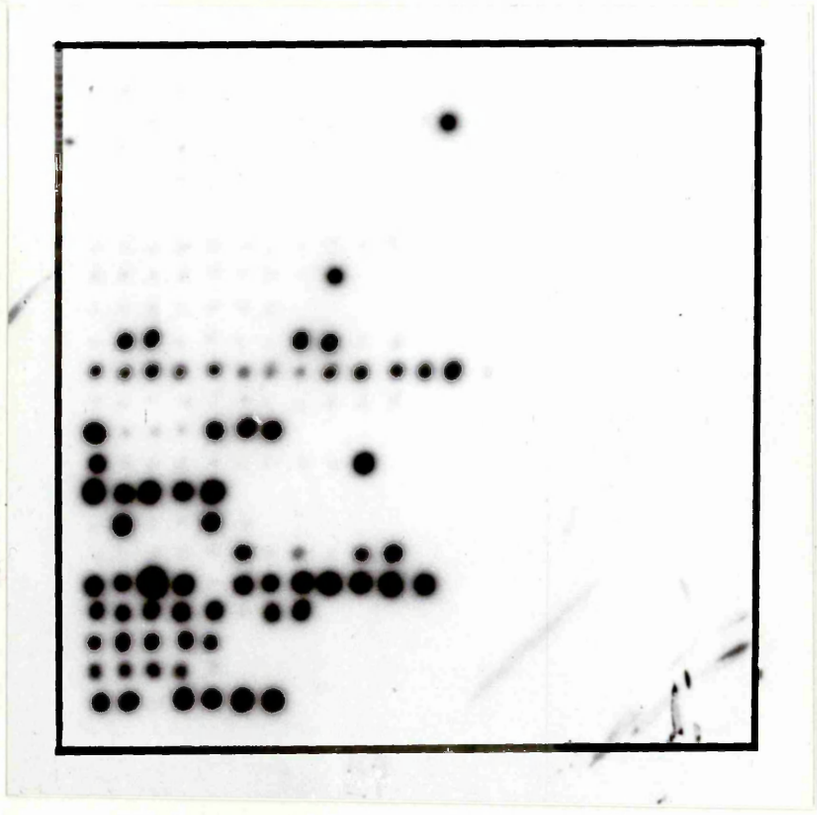


Figure 6.2

Autoradiograph of a filter lifted from one of the duplicates of an array for secondary screening. The very faint signals are from non-hybridising plaques. Rows 1, 3, 4, 5, 7 and 10 did not contain positively hybridising plaques on re-screening. An example of a positively hybridising plaque can be seen at the end of row 2.

Plate 1: Autoradiograph of filter from the genomic library screened with radiolabelled λEMBL3 stuffer fragment. Note all plaques are positive. However only the one arrowed is believed to be wild-type vector. Evidence for this comes from Plate 2 below.

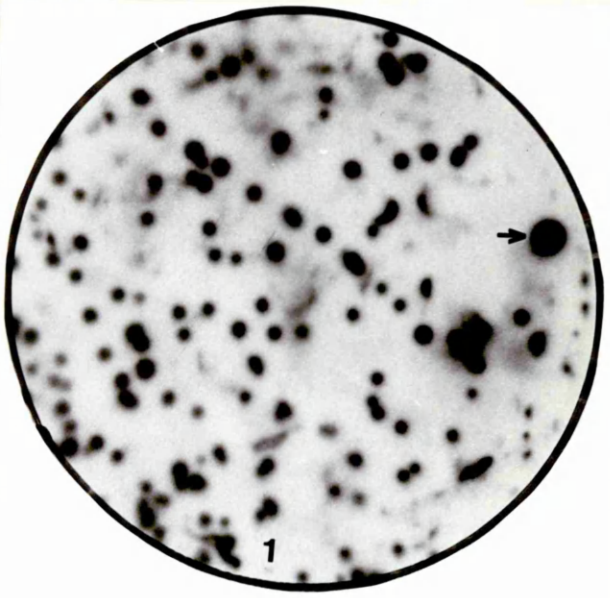


Plate 2: Autoradiograph of filter lifted from EMBL3 plaques and screened with radiolabelled λ EMBL3 stuffer fragment. Note similarity in signal strength between the arrowed plaque on Plate 1 and the plaques on this plate.

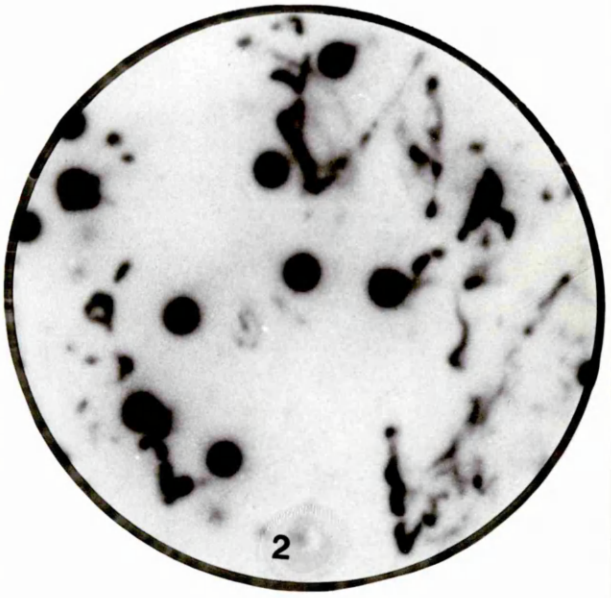


Plate 3: Autoradiograph of filter lifted from the genomic library and screened with radiolabelled *R. acidophila* strain 7050 genomic DNA.

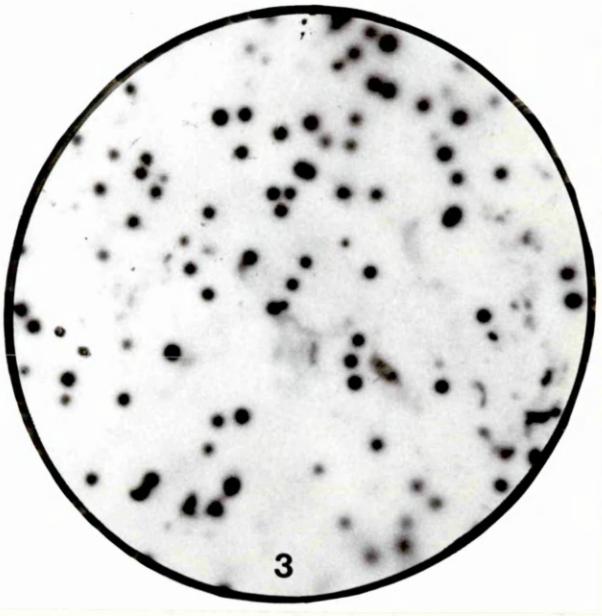


Figure 6.3

Discussion

When the *R. acidophila* recombinants were probed with EMBL3 stuffer fragment all the plaques generated a positive hybridisation signal, see figure 6.3, plate 1. However, in the case of the vast majority of these the signal was relatively weak. This weak signal was assumed to be caused by contamination of the stuffer fragment DNA with other parts of the vector during isolation. It could, however, have been caused by cross reaction of the labelled stuffer fragment with the arms of the recombinant molecules.

The most strongly hybridising plaque (arrowed) is thought to be non-recombinant vector. Evidence for this comes indirectly by comparison with the signal intensity of the known EMBL3 vector plaques (plate 2) which were also probed at the same time. The similarity between these strongly hybridising species suggested that the arrowed positive on plate 1 was an EMBL3 non-recombinant. Perhaps this could have been checked further by isolating the presumptive EMBL3 plaque, making a plate lysate and checking the restriction pattern of the DNA. However, I felt sufficiently confident of this result not to carry out this test.

The difference in signal intensity of the plaques (arrowed *vs* unarrowed) on plate 1 probably reflected, not only contamination or cross reaction of the probe but also differential growth rates of the recombinant and non-recombinant phage. It is possible that non-recombinant phage would be the more vigorous and may have a larger burst size. The additional DNA generated could result in a stronger hybridisation signal.

The level of background ie. non-recombinants/total number of plaques x 100 was calculated to be ≃ 3%. Evidence for this came from other filters which were also screened but have not been shown.

For comparison the *R. acidophila* library was screened with radiolabelled genomic DNA. The idea behind this was, if the

background was calculated as 3%, then 97% of the remaining plaques should generate a positive signal with a genomic DNA probe. Part of the *R. acidophila* library was plated, the plaques counted (≈121/plate) and a filter lifted and probed (plate 3). In addition EMBL3 vector plaques were also probed to ensure that the genomic DNA was not hybridising to vector sequences. As stated earlier, this hybridisation did not generate a positive signal and the result is therefore not presented. Note the EMBL3 and *R. acidophila* recombinants were hybridised together. This acted as an internal control for hybridisation.

The vast majority of plaques generated a positive signal, see plate 3. However, it was impossible to tell if two adjacent plaques were giving a positive signal or if only one. This meant that the background level could not be determined accurately by this method. In addition the number of plaques on the plate were counted but only some of these became bound to the filter because the filter was by necessity smaller than the plate. Therefore, the filter outline had to be marked on the plate and the plaques within the filter area determined. This was difficult to do accurately and there were also plaques which straddled the filter boundary. These were not problematic if they gave a positive signal, but if they did not, then it was impossible to tell if the lack of signal was due to a genuine non-recombinant or because the border had been redrawn in not quite the correct place.

As these flaws in the experimental design could not be overcome I determined the total number of positive signals on the autorads (≈115) and used this figure to determine approximately the number of background plaques ie. 115/121 x 100= 95% non-background, 5% background plaques. This figure is in approximate agreement with that determined by probing the library with the stuffer fragment as described above (≈3%).

This experiment would given a more useful result if the filters had been probed first with radiolabelled genomic DNA then reprobed with

the radiolabelled stuffer fragment. In this way a more accurate level of background could have been determined as a fixed number of plaques would have been screened each time.

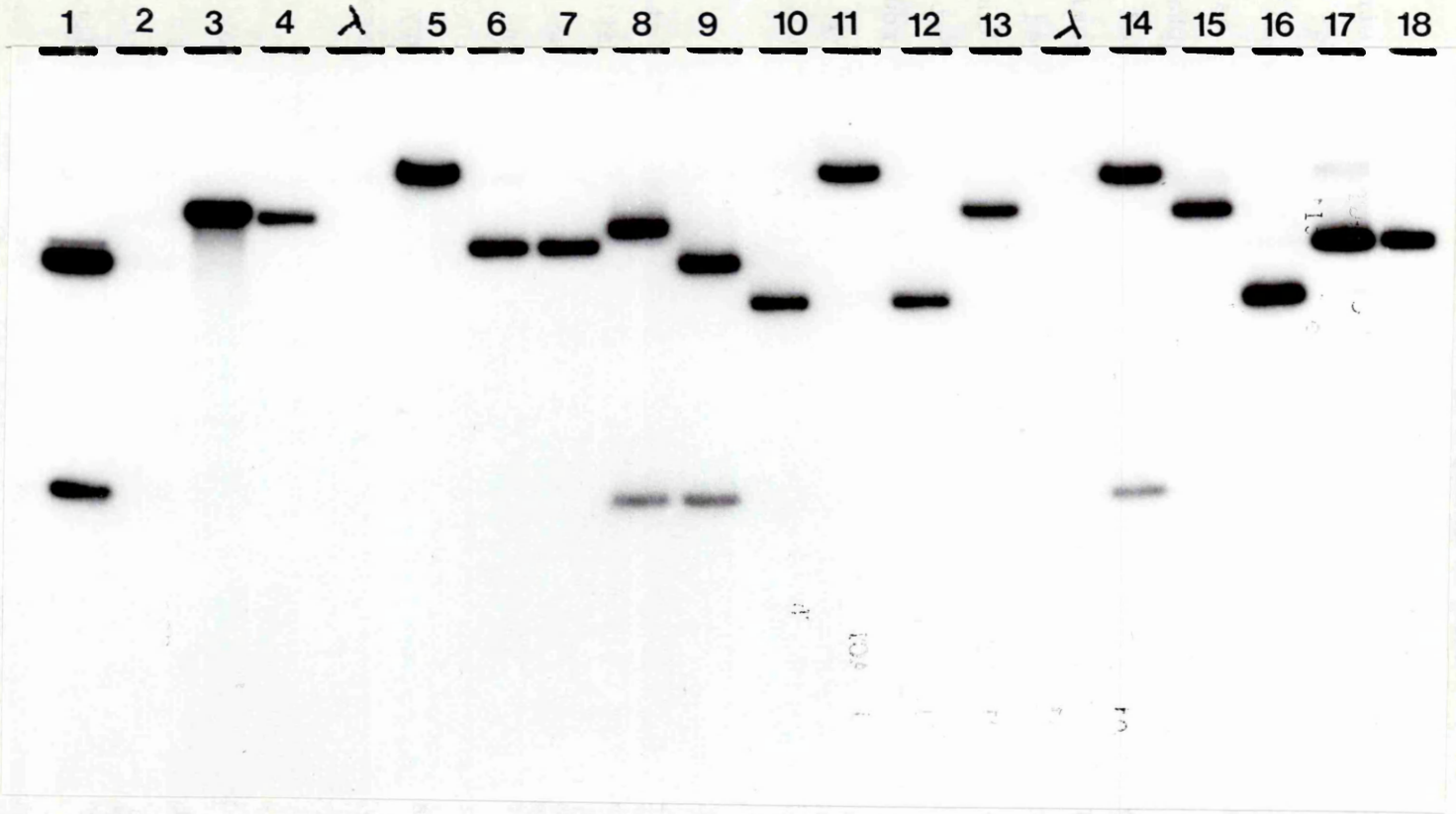
Chapter 7 PHAGE GROUPING AND RESTRICTION MAPPING

Spectrophotometric studies have shown that *R. acidophila* strain 7050 is capable of synthesising at least 3 different types of LHII antenna complex ie. B800-850 types I and II, and B800-820 [Cogdell *et al.*, 1983]. In all purple bacteria examined each light harvesting minimal unit is composed of, at least, one alpha and one beta polypeptide. These polypeptides, in turn, are encoded within a single operon. It seemed reasonable to assume that, by analogy, the light harvesting genes of *R. acidophila* would be clustered into, at most, three different genomic segments, each segment encoding a single alpha-beta polypeptide pair. It was envisaged that each of the 45 positive genomic clones could be placed in three groups, with each group encoding an alpha-beta polypeptide pair and up to ≈40kb of flanking DNA. The groups could then be assigned to segments found within the genomic DNA.

The process of organising the clones into groups was started by making a plate lysate of each of the 45 clones and isolating their DNA. Each of the DNAs were cut with EcoRI+HindIII then run on a 0.8% agarose gel. Comparison of the restriction patterns of each clone with every other clone, showed that 10 of the clones were present as two or more copies, data not shown. One copy of each clone type was kept, the remainder discarded. These copies presumably arose from amplification of the library.

The purpose of amplification is to make multiple copies of each lambda clone. However, when the amplified library is then screened more than one copy of each clone may be picked. This problem could have been overcome by screening a primary (unamplified) library.

The remaining 35 clones were digested as before, run on a 0.8% gel, blotted, then the filters probed with radiolabelled pLHIISB18. The gel photographs and the Southern blot autoradiographs are shown in figures 7.1 and 7.2.



b

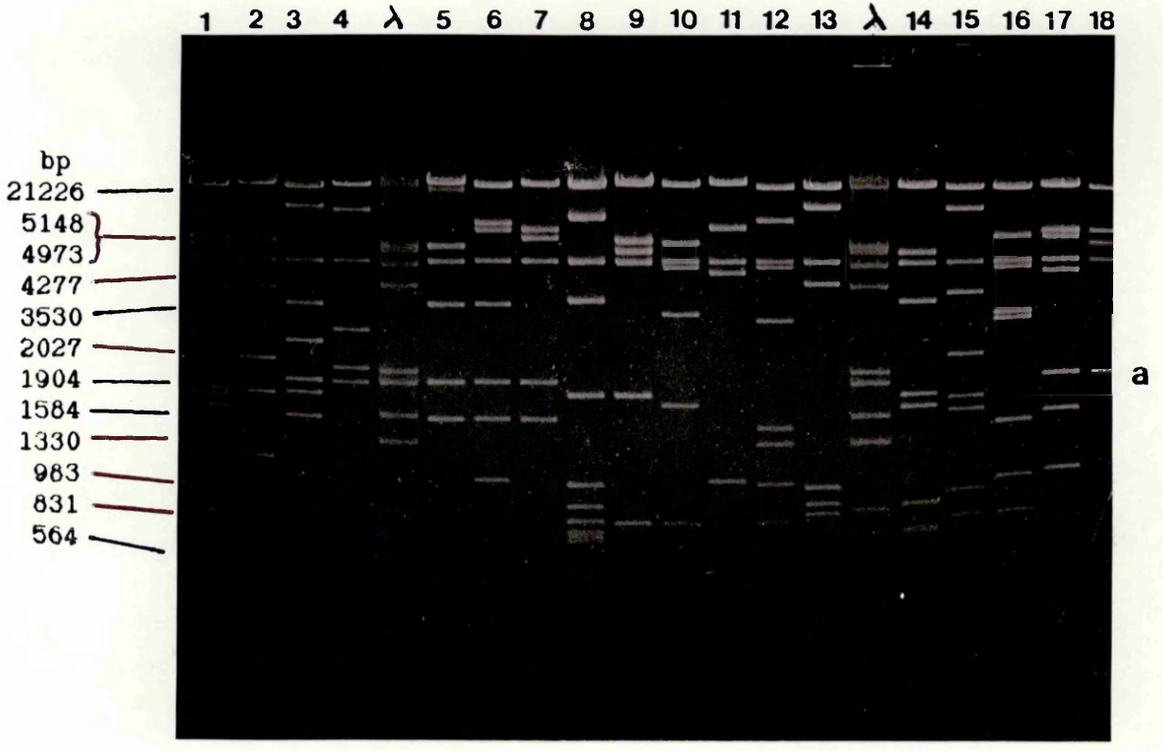


Figure 7.1

Genomic clones 1-18 were digested with EcoRI+HindIII then run on a 0.8% agarose gel (a). The gel was blotted then the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown below (b).

λ= λcI857 DNA cut with EcoRI+HindIII.

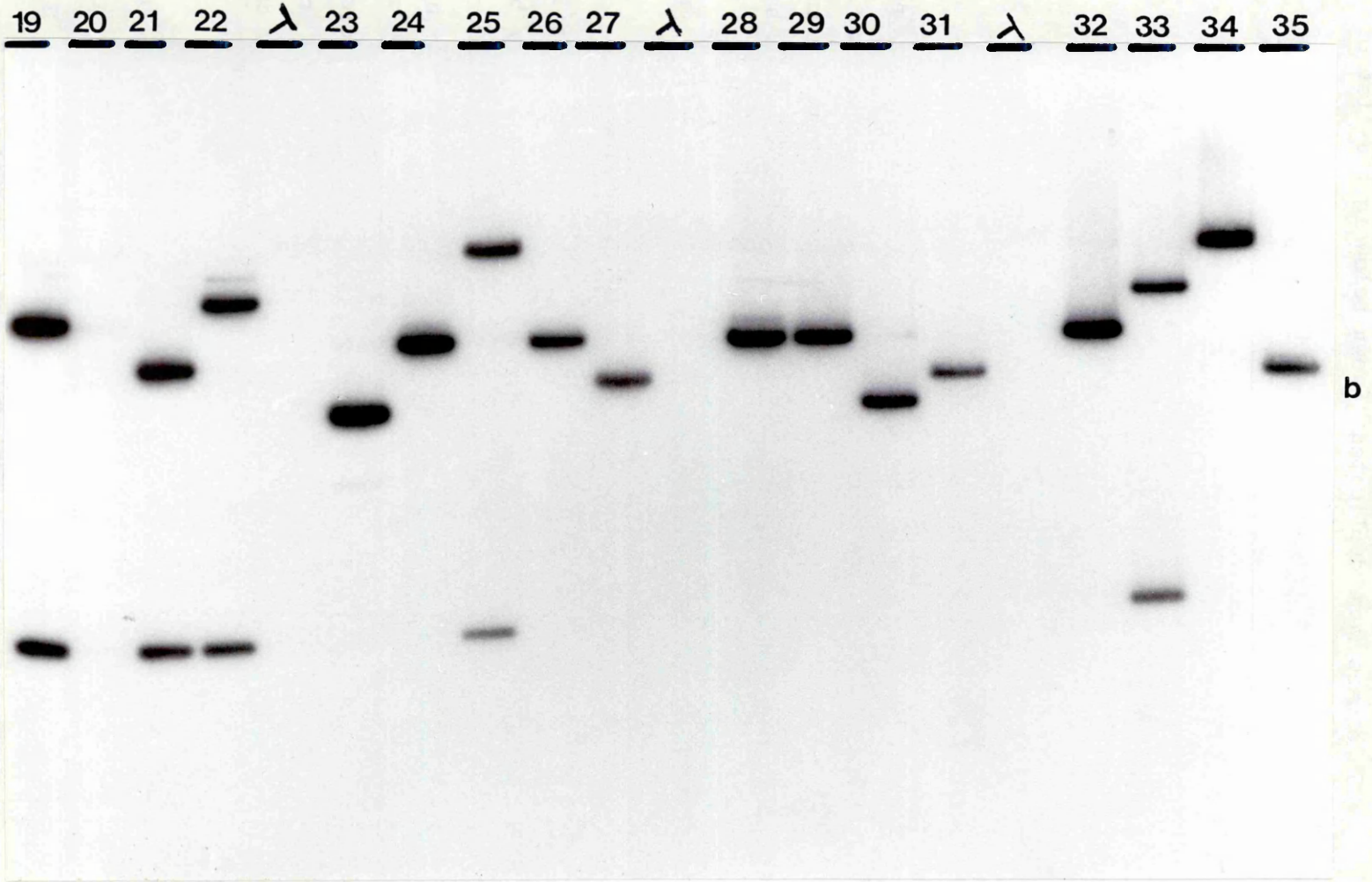
This second restriction screening revealed that 14 of the remaining 35 clones were duplicated. Shown below is the list of the 6 clones which were kept and the 8 copies which were discarded:

KEPT		DISCARDED
7	=	32.
8	=	22, 33 and 19.
10	=	30
14	=	25
18	=	29
27	=	35

This left 27 clones which were unique. Two of these, numbers 2 and 20 failed to give a positive hybridisation signal when probed, see figures 7.1 and 2. These clones were not examined further. Unfortunately two clones, numbers 24 and 26 failed in batch culture, so they were also removed from further examination. This left 23 unique clones and it was from these that phage groupings and restriction maps were determined.

By comparing the restriction pattern of each of the EcoRI+HindIII digested clones, a picture of those clones which may be related to one another began to emerge. This picture, however, was often contradicted by the data which emerged from Southern blot analysis. Clones 3, 5, 6, and 7 for example, all possess bands of ≃1.5 and 1.9kb, see figure 7.1a. This suggested that these clones may form part of an overlapping series. However, Southern blot analysis suggested that only 6 and 7 were related as only they had positively hybridising bands of the same size, see figure 7.1b.

This apparently contradictory evidence arose through the choice of enzymes used for the digestion of the clones and can be explained by the following model. Imagine that the gene of interest lies between two EcoRI sites on the insert of a genomic clone. Digestion of this clone with EcoRI would release an EcoRI fragment of size X which would hybridise positively. However, let us now assume that


19
20
21
22
λ
23
24
25
26
27
λ
28
29
30
31
λ
32
33
34
35
b

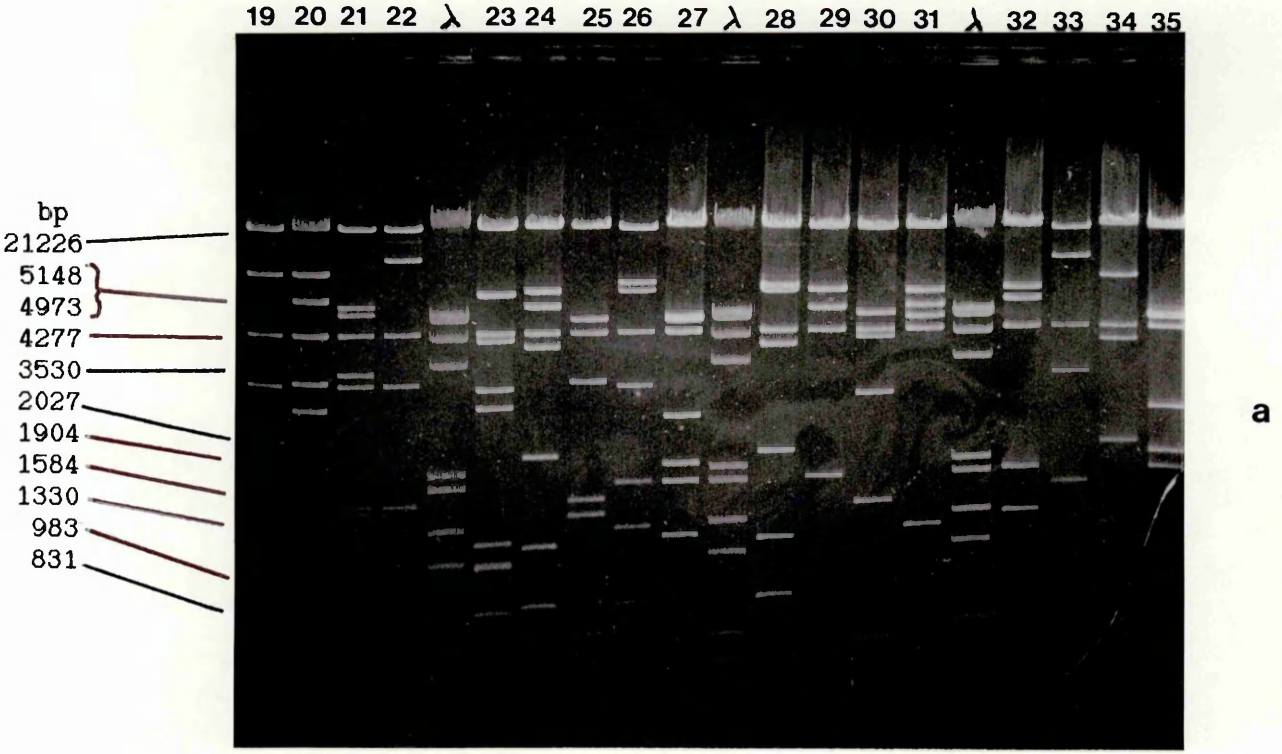


Figure 7.2

Genomic clones 19-35 were digested with EcoRI+HindIII then run on a 0.8% agarose gel (a). The gel was blotted then the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown below (b).

λ= λcI857 DNA cut with EcoRI+HindIII.

there is a genomic clone that overlaps with the first. The insert of this clone also contains the gene of interest, however, it only contains a single EcoRI site, the other site having been lost during the construction of the library. As there are no sites for EcoRI in the vector arms then digestion of the clone with this enzyme would result in a positively hybridising fragment the size of which would be; ≃X + 20 or 9kb ie. the size of the vector arms. Therefore, overlapping clones could have similar restriction patterns but very different Southern blot patterns. This anomaly meant that groupings of phage were made initially on the basis of their restriction pattern.

Southern blot patterns did however prove useful, for example, the two positively hybridising bands found in each of clones 1, 8, 9, 14, 19, 21, 22, 25 and 33 suggested that they may belong to a single related group. However, as described earlier, clones 8, 19, 22 and 33 had an identical restriction pattern and only number 8 of these four was retained. From the photographs this does not appear to be the case because clone 8 has a number of bands smaller than ≃1.5kb and these are not visible in the other three clones. However, this is photographic artefact which was not apparent with the original gel was examined. Likewise clones 14 and 25 had an identical restriction pattern and only 14 was retained. The remaining five clones 1, 8, 9, 14 and 21 did appear to form an overlapping series.

Another group of candidates for forming an overlapping series were clones 10, 12, 16, and 23. These clones had the same sized positively hybridising restriction fragments and common restriction banding patterns.

To investigate the restriction patterning further a large scale DNA preparation was made from each of the unique phage isolates. As stated earlier clones 24 and 26 failed at this stage and were not investigated further.

Each of the remaining 23 clone DNAs were digested with; HindIII, HindIII+SalI, SalI, SalI+EcoRI and EcoRI. SalI was used as to overcome the Southern blot anomalies described earlier. This enzyme separates the insert DNA from the vector arms without cleaving the arms themselves. As there are no EcoRI sites in the vector arms, double digestion with SalI+EcoRI generates the same vector fragments as with SalI alone.

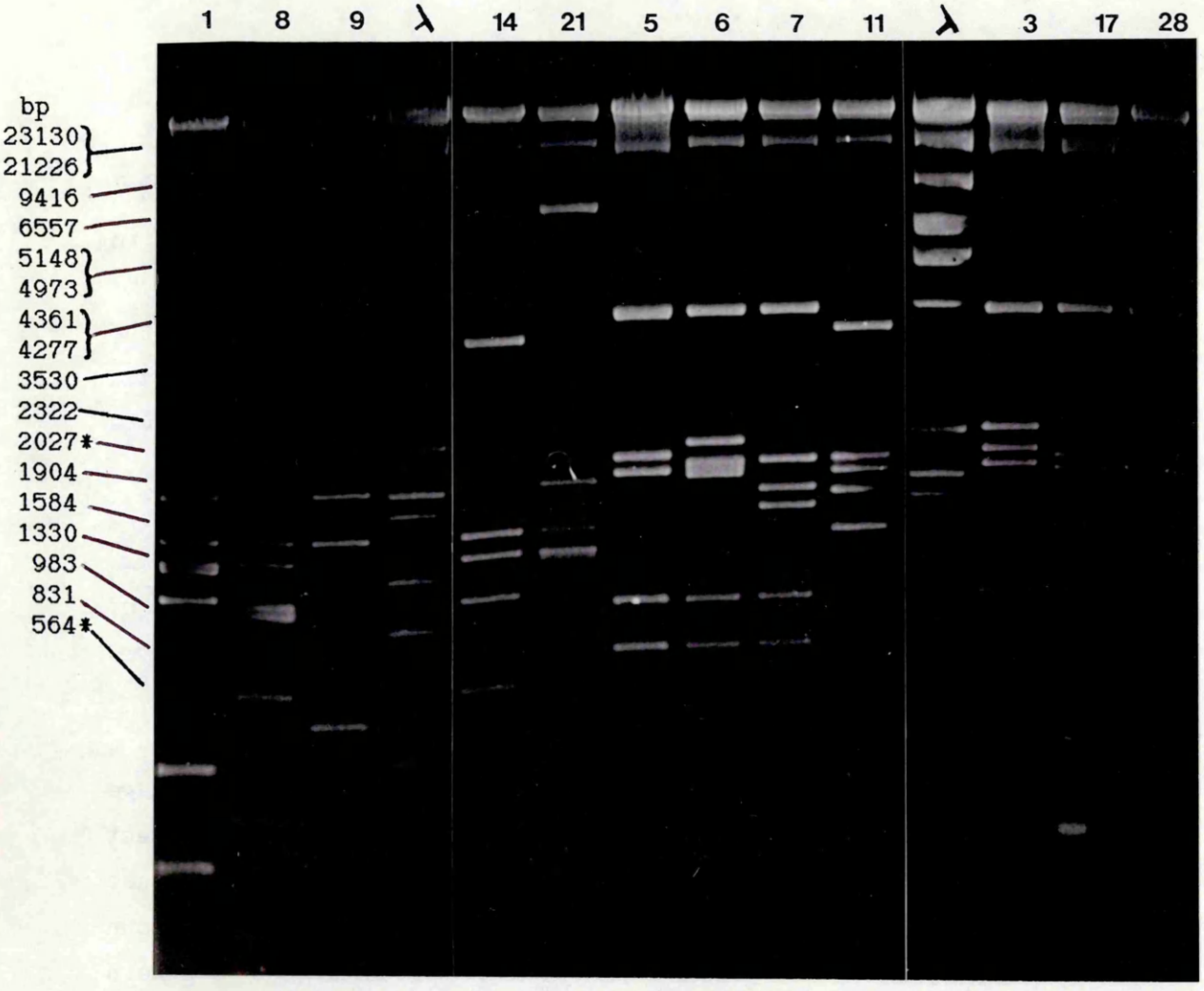
HindIII cleaves the right arm of the vector at ≃4.5kb from its end. Double digestion with SalI and HindIII releases two fragments of ≃4.5kb from the right arm of the vector, the arm being ≃9kb in size. There are no HindIII sites in the left arm of the vector.

On the basis of the previous groupings each of the digested phage DNAs were run next to what was considered to be their closest relative. The results (not shown) suggested that the lambda clones could be placed in four distinct groups. These groupings are shown as a combination of SalI+EcoRI restriction patterns and Southern blot patterns in figures 7.3 and 7.4.

Group 1	Group 2	Group 3	Group 4
1	5	10	31
8	6	12	18
9	7	16	27
14	11	23	4
21	3		13
	17		15
	28		34

The clones within each of groups 1, 2, and 3 formed an overlapping series probably derived from the same genomic segment. The clones within any one particular group were not related to, and did not appear to overlap with, any members of the other two groups.

Group 4 consisted of a collection of clones which did not appear related to clones in groups 1-3. In addition any particular clone



Figures 7.3 & 7.4

To illustrate the groupings of the genomic clones, the DNAs have been digested with SalI+EcoRI and run on 0.8% agarose gels, see also Figure 7.4 overleaf.

The gel shown above (Figure 7.3a) has two groups of clones:-

Group 1 clones; 1, 8, 9, 14 & 21.

Group 2 clones; 5, 6, 7, 11, 3, 17 & 28.

Notes The marker fragments (shown as λ on the gels) are λ cI857 DNA digested with EcoRI+HindIII, mixed with λ cI857 DNA digested with HindIII alone. Doublets are marked with *.

The gel photographs have been cut to remove unwanted lanes at the following positions:

Figure 7.3a: Between λ and clone 14, and between λ and clone 11.

Figure 7.4a: Between λ and clone 10, and between λ and clone 13.

The joins are clearly visible in both photographs.

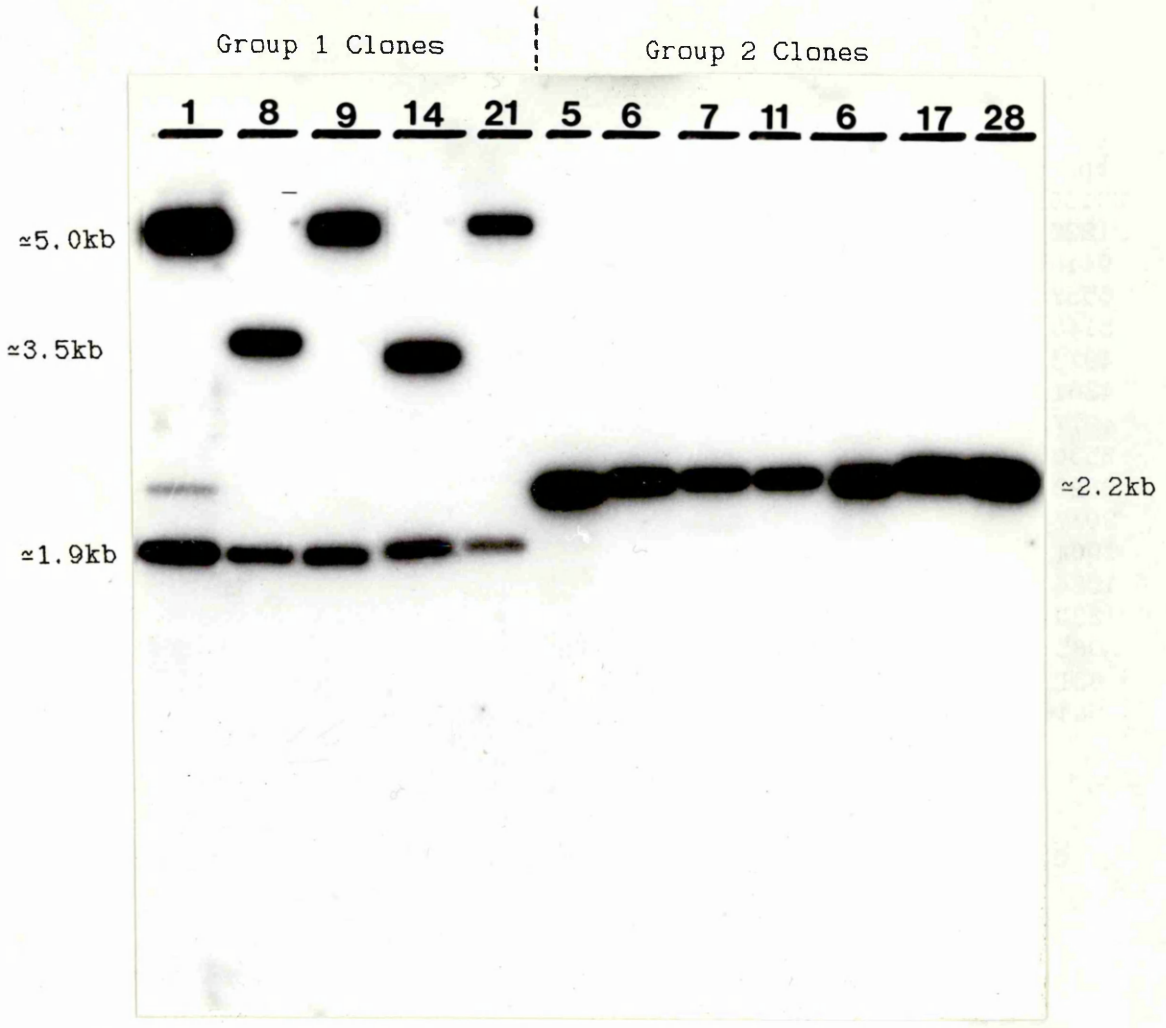


Figure 7.3b

The gel shown in Figure 7.3a was blotted and the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown above, Figure 7.3b. To emphasise the positively hybridising common bands in each group the λ marker lanes were removed before the filter was used for autoradiography. Size markers are now placed at the ends of the autoradiograph.

Clones in Group 1 have two positively hybridising bands: A band of approximately 1.9kb common to all of the clones, and probably representative of the genomic state; and a higher molecular weight band of either ≃ 3.5kb or ≃ 5kb. The latter is thought to be representative of the genomic state, whereas the former was probably generated by Sau3AI digestion of the larger 5kb fragment during construction of the library.

Clones in Group 2 share a common positively hybridising 2.2kb band.

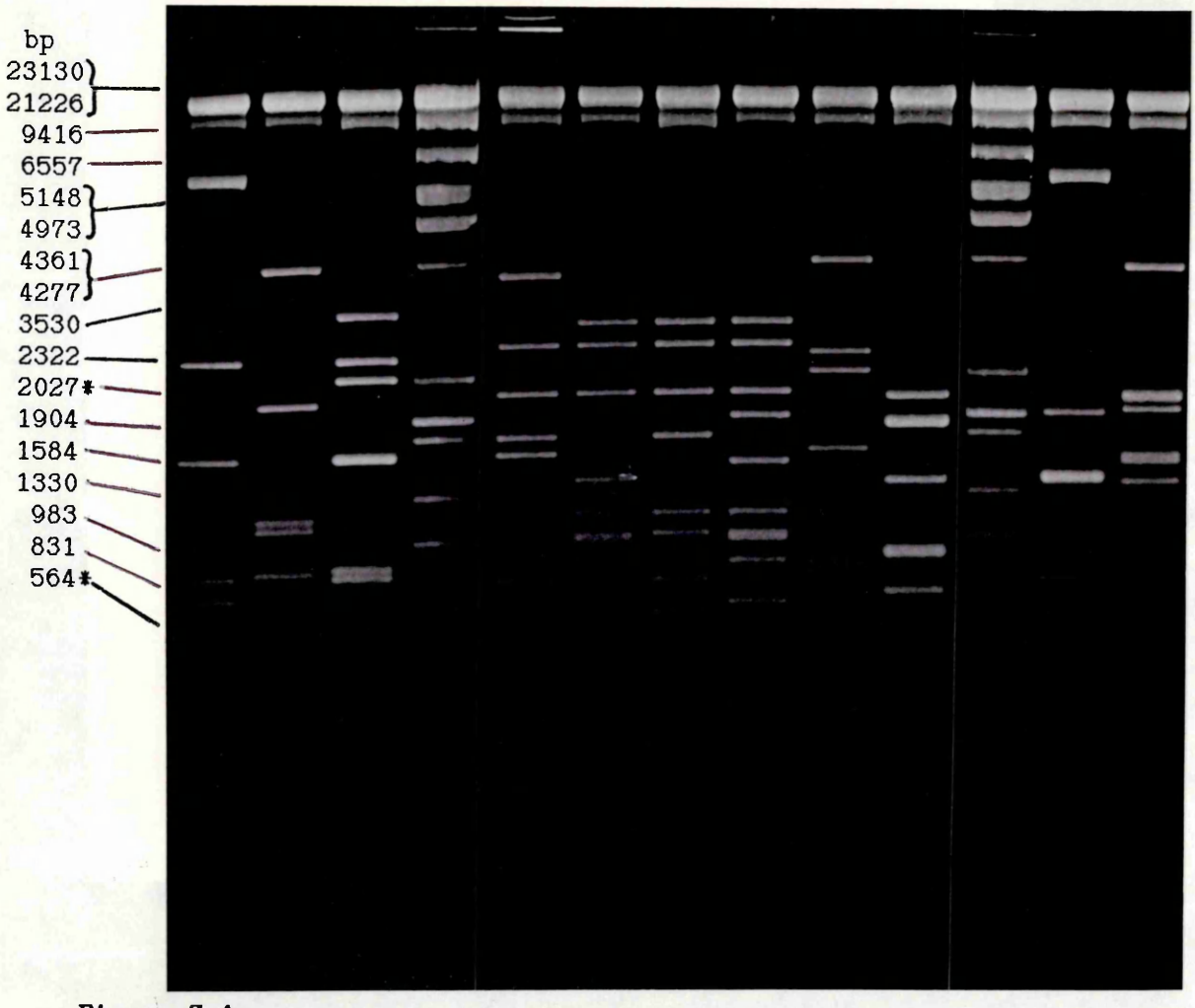


Figure 7.4a

To illustrate the groupings of the genomic clones, the DNAs have been digested with SalI+EcoRI and run on 0.8% agarose gels, see also Figure 7.3 on previous page.

The gel shown above (Figure 7.4a) has two groups of clones:-

Group 3 clones; 10, 12, 16 & 23.

Group 4 clones; 31, 18, 27, 4, 13, 15 & 34.

Notes The marker fragments (shown as λ on the gels) are λ cI857 DNA digested with EcoRI+HindIII, mixed with λ cI857 DNA digested with HindIII alone. Doublets are marked with *.

The gel photograph has been cut to remove unwanted lanes between λ and clone 10, and between λ and clone 13. The joins are clearly visible in the photograph.

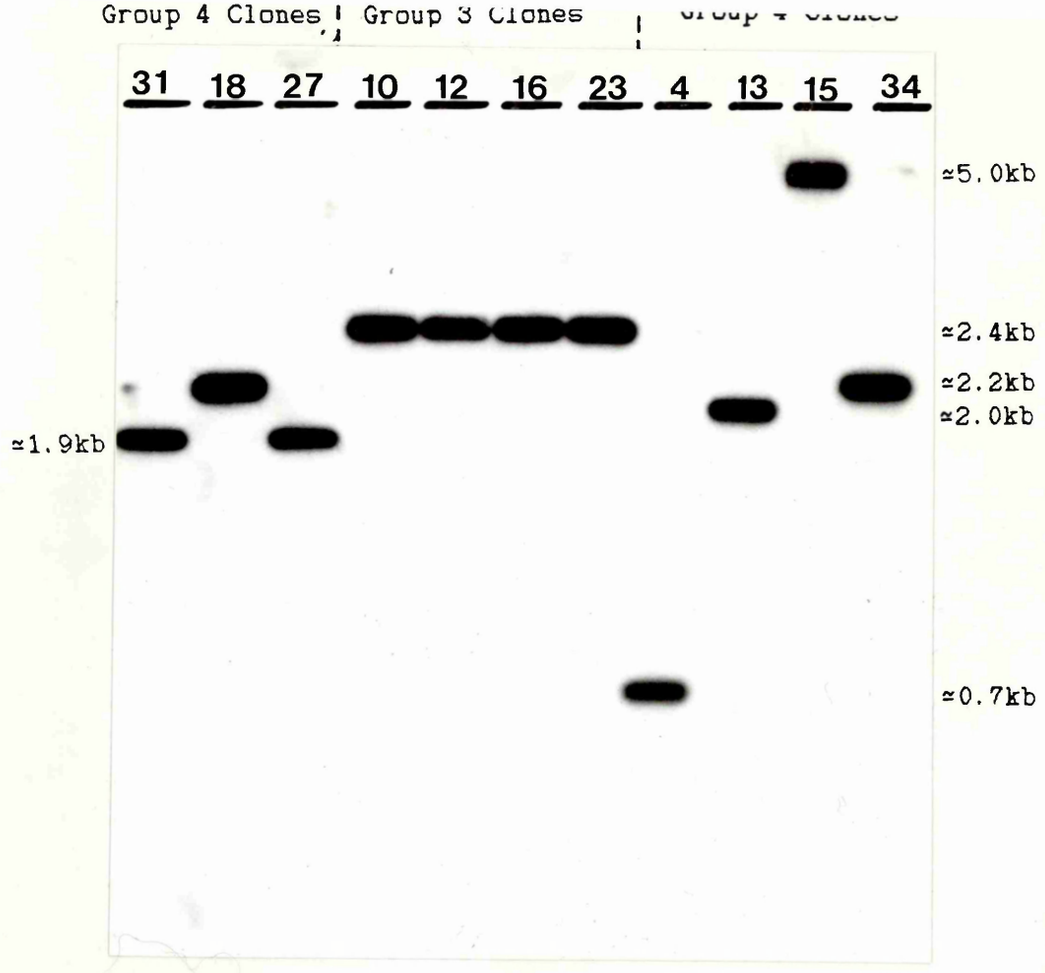


Figure 7.4b

The gel shown in Figure 7.4a was blotted and the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown above, Figure 7.4b. To emphasise the positively hybridising common bands in each group the λ marker lanes were removed before the filter was used for autoradiography. Size markers are now placed at the ends of the autoradiograph.

Clones in Group 3 share a positively hybridising band of ≈2.4kb.
Clones in Group 4 have postively hybridising bands of the following sizes:.

Clone 31 ≈1.9kb
Clone 18 ≈2.2kb
Clone 27 ≈1.9kb (doublet)
Clone 4 ≈0.7kb
Clone 13 ≈2.0kb
Clone 15 ≈5.0kb
Clone 34 ≈2.2kb

within group 4 did not appear to be related to any other group 4 clone, with the possible exceptions of clones 31 and 27 which may be related. Some of these group 4 clones ie. 31, 18, 27, 18 and 34 contained positively hybridising fragments which were of the same size as those found by genomic Southern blotting (see chapter 10). It is therefore likely that these clones are poorly represented within the the library and have not been picked with sufficient frequency to form an overlapping series.

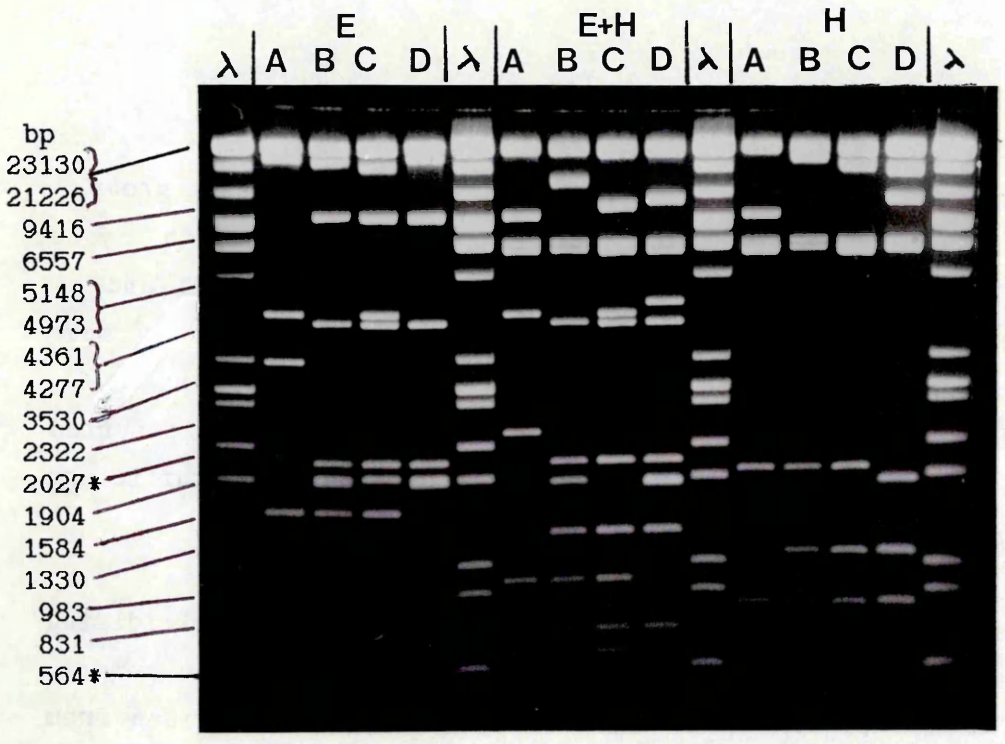
Some of the group 4 clones, however, contain positively hybridising fragments which do not appear to have counterparts within genomic Southern blots, clones 4 and 13. These have possibly arisen from cloning a Sau3AI partial fragment which contains only a part of a presumptive light harvesting gene.

As the clones in groups 1-3 were overlapping and provided a relatively easier prospect for mapping only these groups were examined further. However, though neither mapped nor sequenced, clones within group 4 may yet provide a useful source for the isolation of further light harvesting genes.

Clones in groups 1, 2 and 3 were mapped by digesting the clones from each group with HindIII, EcoRI, SalI, HindIII+EcoRI, HindIII+SalI and EcoRI+SalI. The digests were run on a 0.8% gel, blotted then probed with radiolabelled pLHIISB18. An example of one such series of digests is shown for the clones in group 3, figures 7.5-7.7. From such data restriction maps were constructed for the clones within the three groups see figures. 7.8-7.10.

It should be emphasised that these maps have not been confirmed by partial digestion or by subcloning and that the sizes of the fragments shown are approximate.

Positively hybridising fragments from representative clones in groups 1-3 were then subcloned into mp19 before sequencing.


E
E+H
H
λ A B C D λ A B C D λ A B C D λ
bp
23130
21226
9416
6557
5148
4973
4361
4277
3530
2322
2027*
1904
1584
1330
983
831
564*

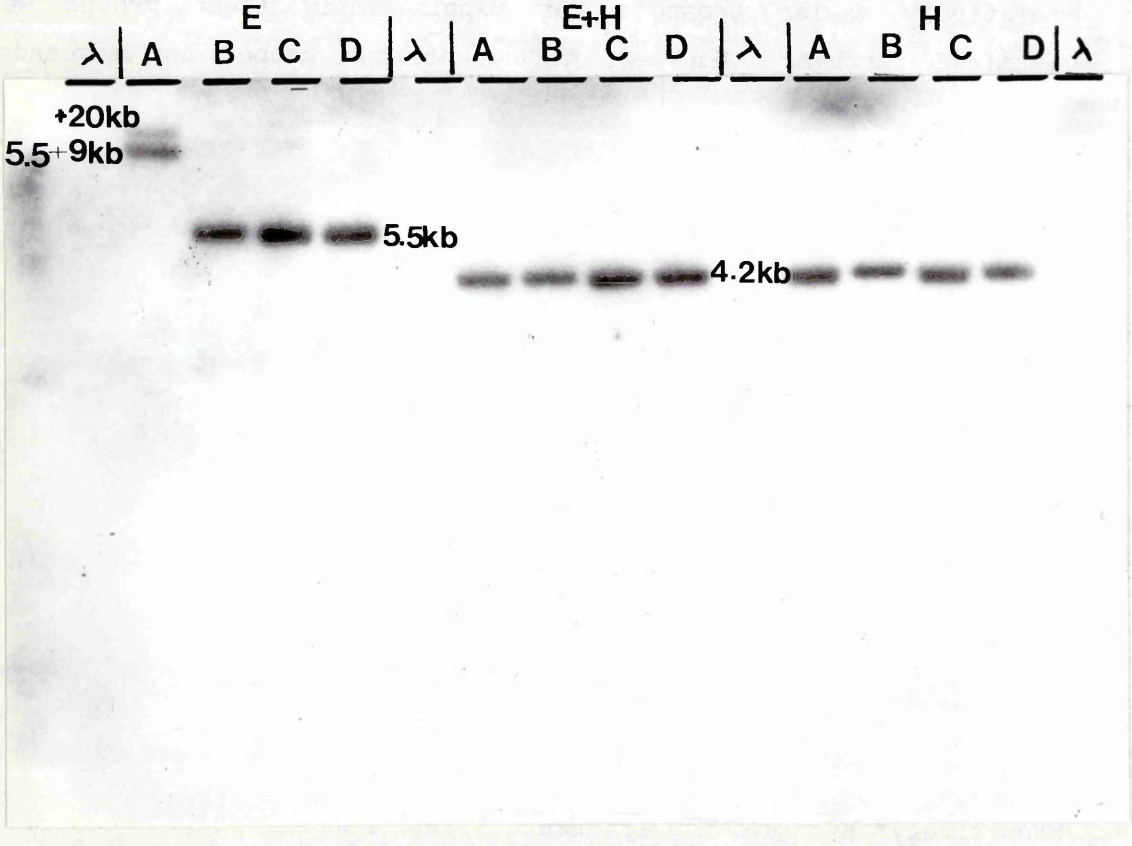
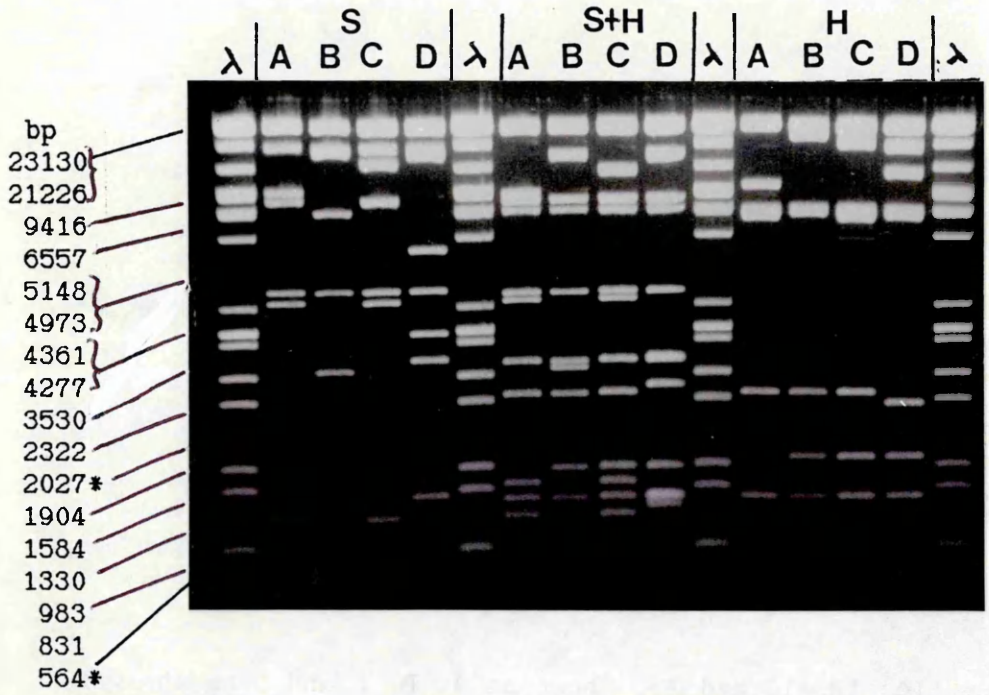

E
E+H
H
λ A B C D λ A B C D λ A B C D λ
+20kb
5.5+9kb
5.5kb
4.2kb

Figure 7.5

Genomic clones 10, 12, 16 and 23, shown as A, B, C and D on the gel photograph, were digested with EcoRI (E), HindIII (H) and a combination of EcoRI+HindIII (E+H). λ= λ cI857 DNA which has been digested with EcoRI+HindIII mixed with λ cI857 DNA digested with HindIII.

The gel was then blotted and the filter probed with radiolabelled pLHIISB18 DNA. The resulting autoradiograph is also shown. The sizes of the positively hybridising bands have been marked on the autoradiograph. Note the genomic clone 10 (A) when digested with EcoRI generates two positively hybridising bands, one of >20kb and and one of ≈14.5kb (shown as 5.5+9kb). A possible explanation for this is given in the text.


S
S+H
H
λ A B C D λ A B C D λ A B C D λ
bp
23130
21226
9416
6557
5148
4973
4361
4277
3530
2322
2027*
1904
1584
1330
983
831
564*

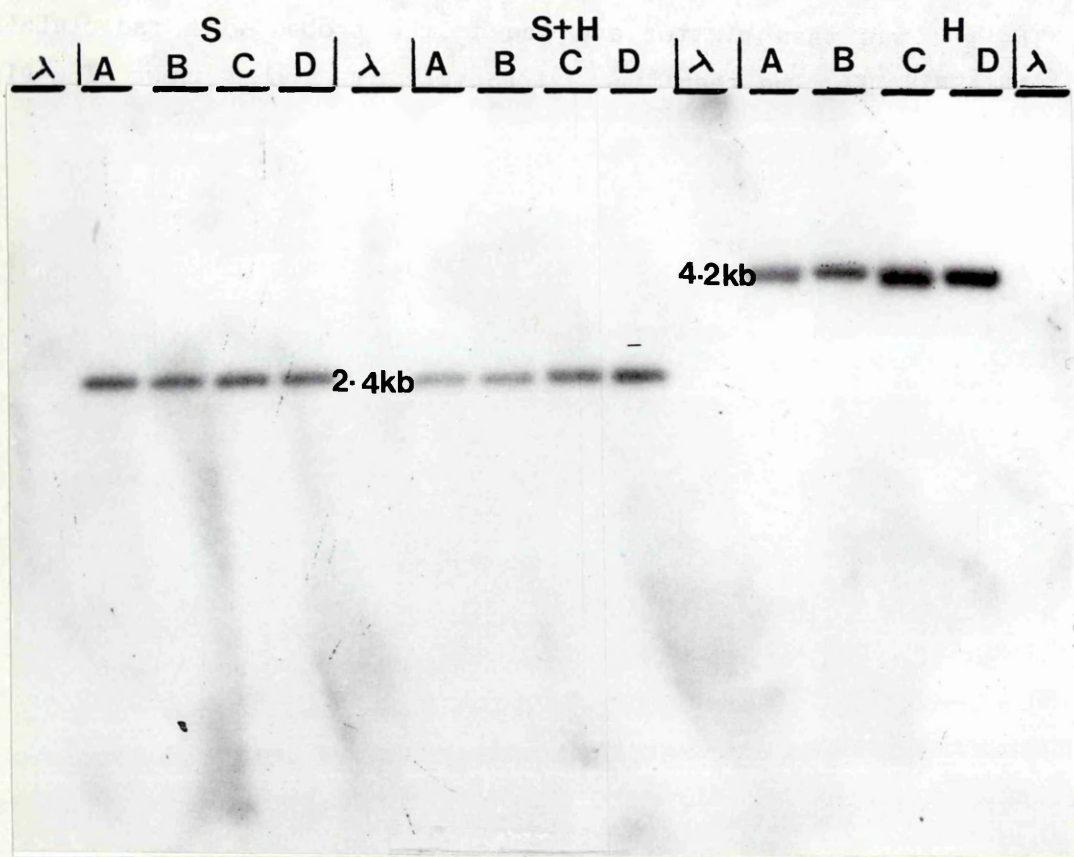
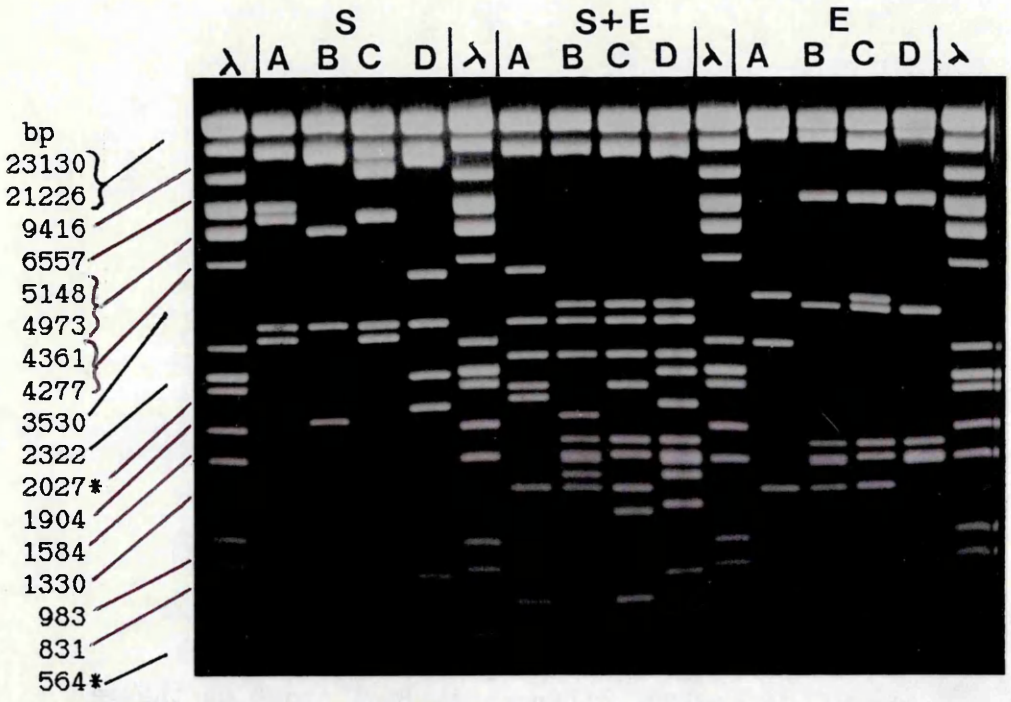

S
S+H
H
λ A B C D λ A B C D λ A B C D λ
4·2kb
2·4kb

Figure 7.6

Genomic clones 10, 12, 16 and 23, shown as A, B, C and D on the gel photograph, were digested with SalI (S), HindIII (H) and a combination of SalI+HindIII (S+H). λ= λ cI857 DNA which has been digested with EcoRI+HindIII mixed with λ cI857 DNA digested with HindIII.

The gel was then blotted and the filter probed with radiolabelled pLHIISB18 DNA. The resulting autoradiograph is also shown. The sizes of the positively hybridising bands have been marked on the autoradiograph.


S
S+E
E
λ A B C D λ A B C D λ A B C D λ
bp
23130
21226
9416
6557
5148
4973
4361
4277
3530
2322
2027*
1904
1584
1330
983
831
564*

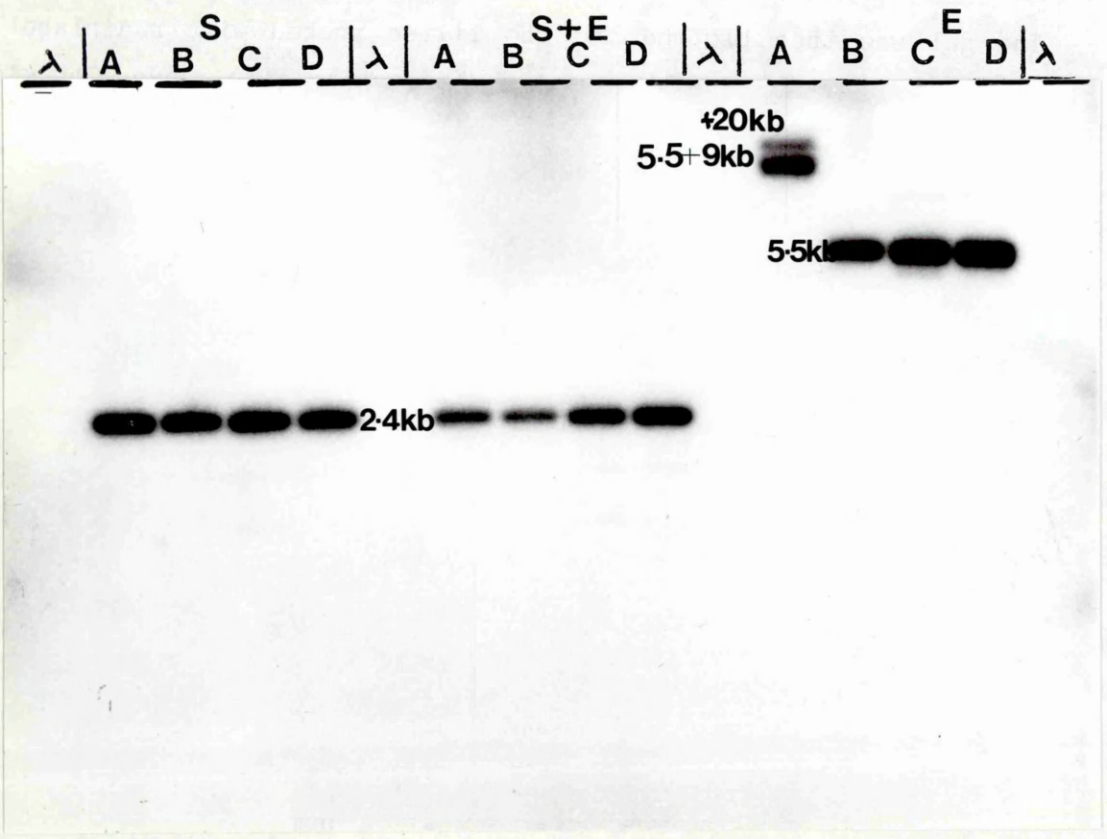

S
S+E
E
λ A B C D λ A B C D λ A B C D λ
+20kb
5·5+9kb
5·5kb
2·4kb

Figure 7.7

Genomic clones 10, 12, 16 and 23, shown as A, B, C and D on the gel photograph, were digested with SalI (S), EcoRI (E) and a combination of SalI+EcoRI (S+E). λ= λ cI857 DNA which has been digested with EcoRI+HindIII mixed with λ cI857 DNA digested with HindIII.

The gel was then blotted and the filter probed with radiolabelled pLHIISB18 DNA. The resulting autoradiograph is also shown. The sizes of the positively hybridising bands have been marked on the autoradiograph. Note the genomic clone 10 (A) when digested with EcoRI generates two positively hybridising bands, one of >20kb and and one of ≈14.5kb (shown as 5.5+9kb). A possible explanation for this is given in the text.

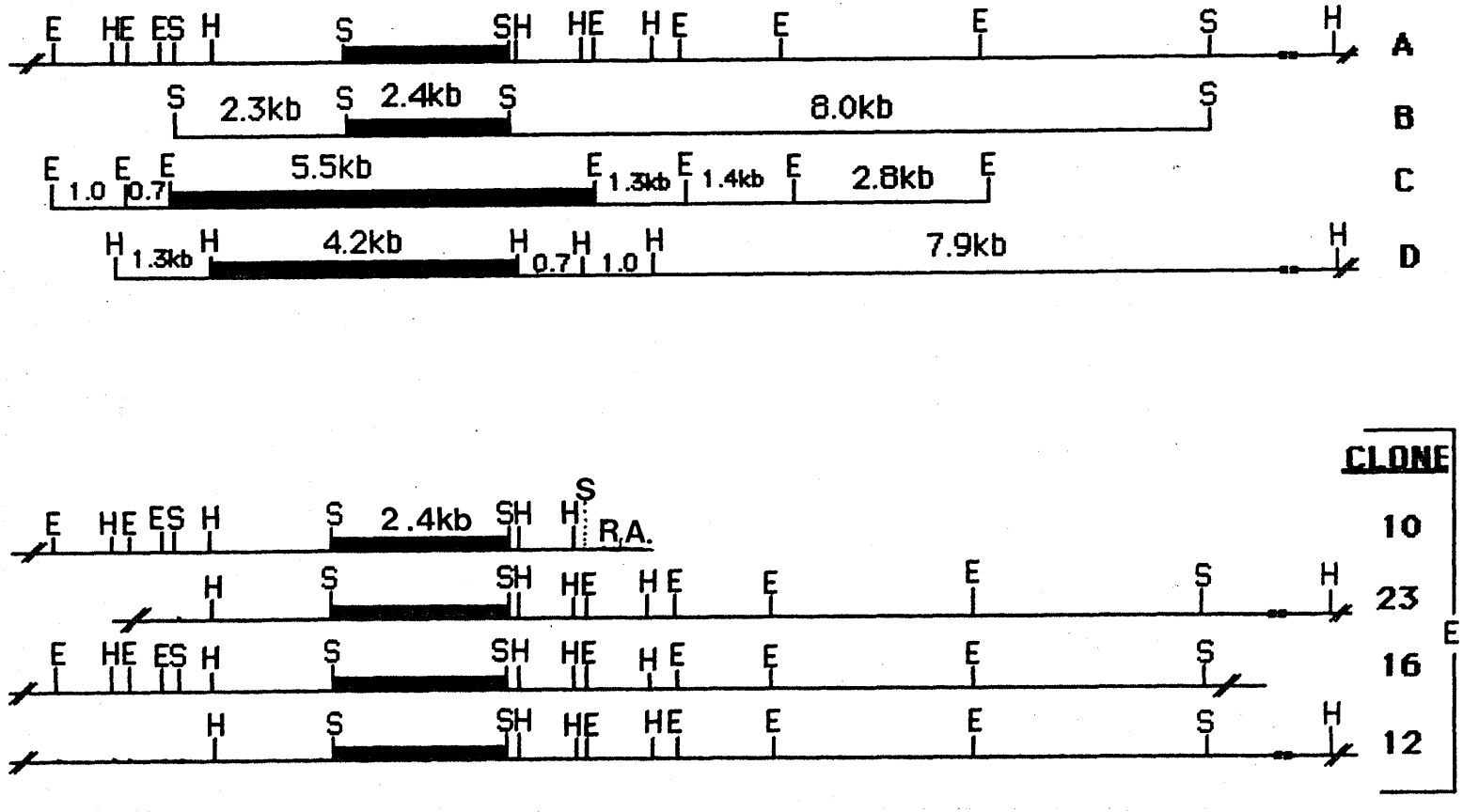

E HE ES H S SH HE H E E E S H
A
S 2.3kb S 2.4kb S 8.0kb S
B
E 1.0 E 0.7 E 5.5kb E 1.3kb E 1.4kb E 2.8kb E
C
H 1.3kb H 4.2kb H 0.7 H 1.0 H 7.9kb H
D
CLONE
E HE ES H S 2.4kb SH H S R.A.
10
H S SH HE H E E E S H
23
E HE ES H S SH HE H E E E S
16
H S SH HE HE E E S H
12
E

Figure 7.8

A=The predicted restriction map of the genomic region of DNA encompassed by lambda clones 10, 23, 16, and 12.

B, C, and D are the predicted restriction maps of this genomic region when cut with Hind III, SalI and EcoRI respectively. The sizes of the fragments generated are are shown in kilobases.

E=The restriction maps of the individual genomic clones.

-■ ■-Signifies a restriction fragment which is non-scale on the map and the size of which is either presented or can be determined from the available information.

-//-Signifies the end of the region mapped within both the genomic segment and lambda clones. Note that unless otherwise stated restricion fragments exist outwith these areas but have not been mapped.

S
: These are SalI sites which are present in the EMBL vector. They are not found in the insert. R.A=Right arm of vector.

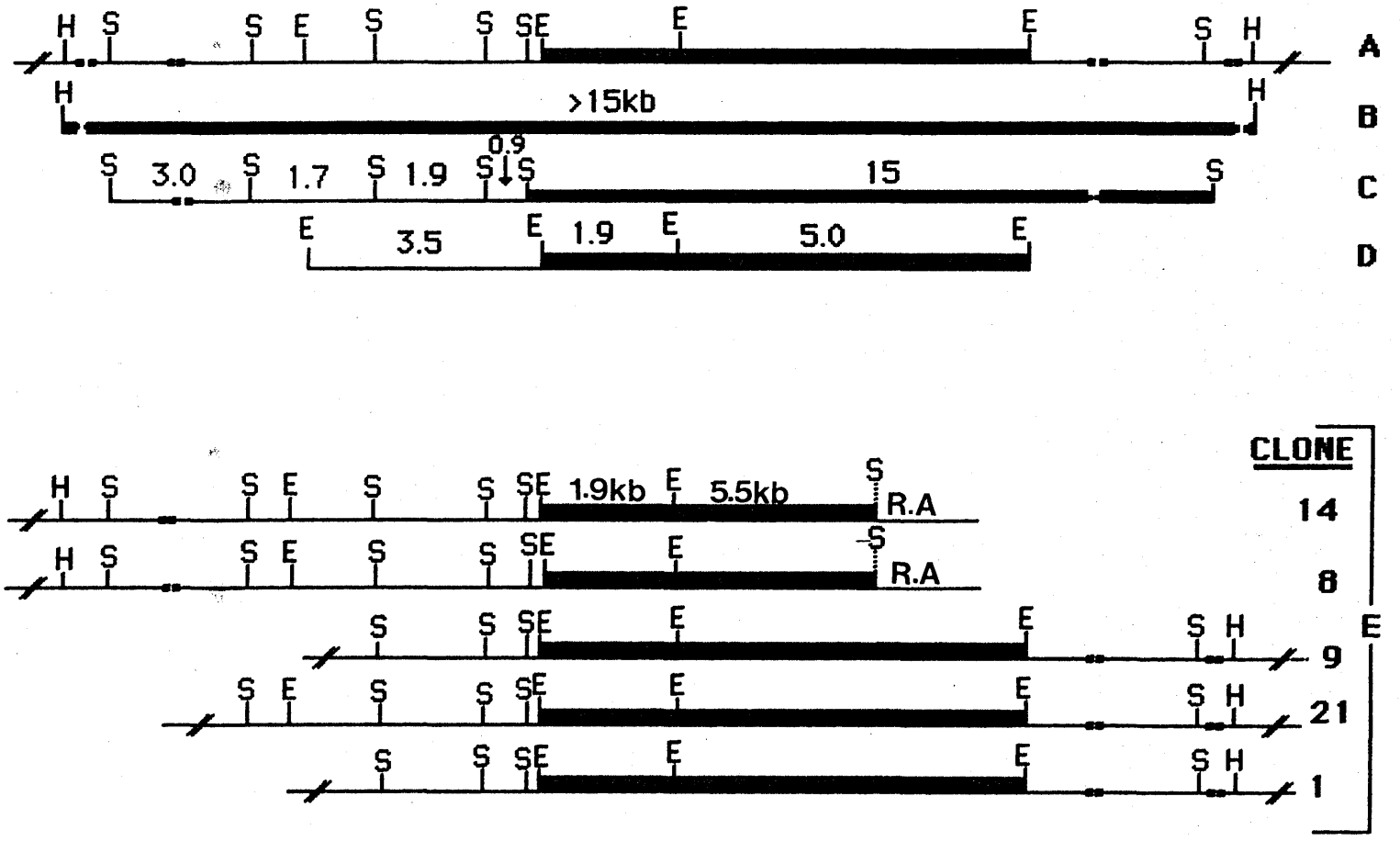

H S S E S S SE E E S H
A
H >15kb H
B
S 3.0 S 1.7 S 1.9 S 0.9 S 15 S
C
E 3.5 E 1.9 E 5.0 E
D
CLONE
H S S E S S SE 1.9kb E 5.5kb S R.A
14
H S S E S S SE E S R.A
8
S S SE E E S H
9
S E S S SE E E S H
21
S S SE E E S H
1
E

Figure 7.9

A=The predicted restriction map of the genomic region of DNA encompassed by lambda clones 14, 8, 9, 21, and 1.
B, C, and D are the predicted restriction maps of this genomic region when cut with Hind III, SalI and EcoRI respectively. The sizes of the fragments generated are are shown in kilobases.
E=The restriction maps of the individual genomic clones.
—▮▮—Signifies a restriction fragment which is non-scale on the map and the size of which is either presented or can be determined from the available information.
—//—Signifies the end of the region mapped within both the genomic segment and lambda clones. Note that unless otherwise stated restricion fragments exist outwith these areas but have not been mapped.
S
⋮ These are SalI sites which are present in the EMBL vector. They are not found in the insert. R.A.=Right arm of vector

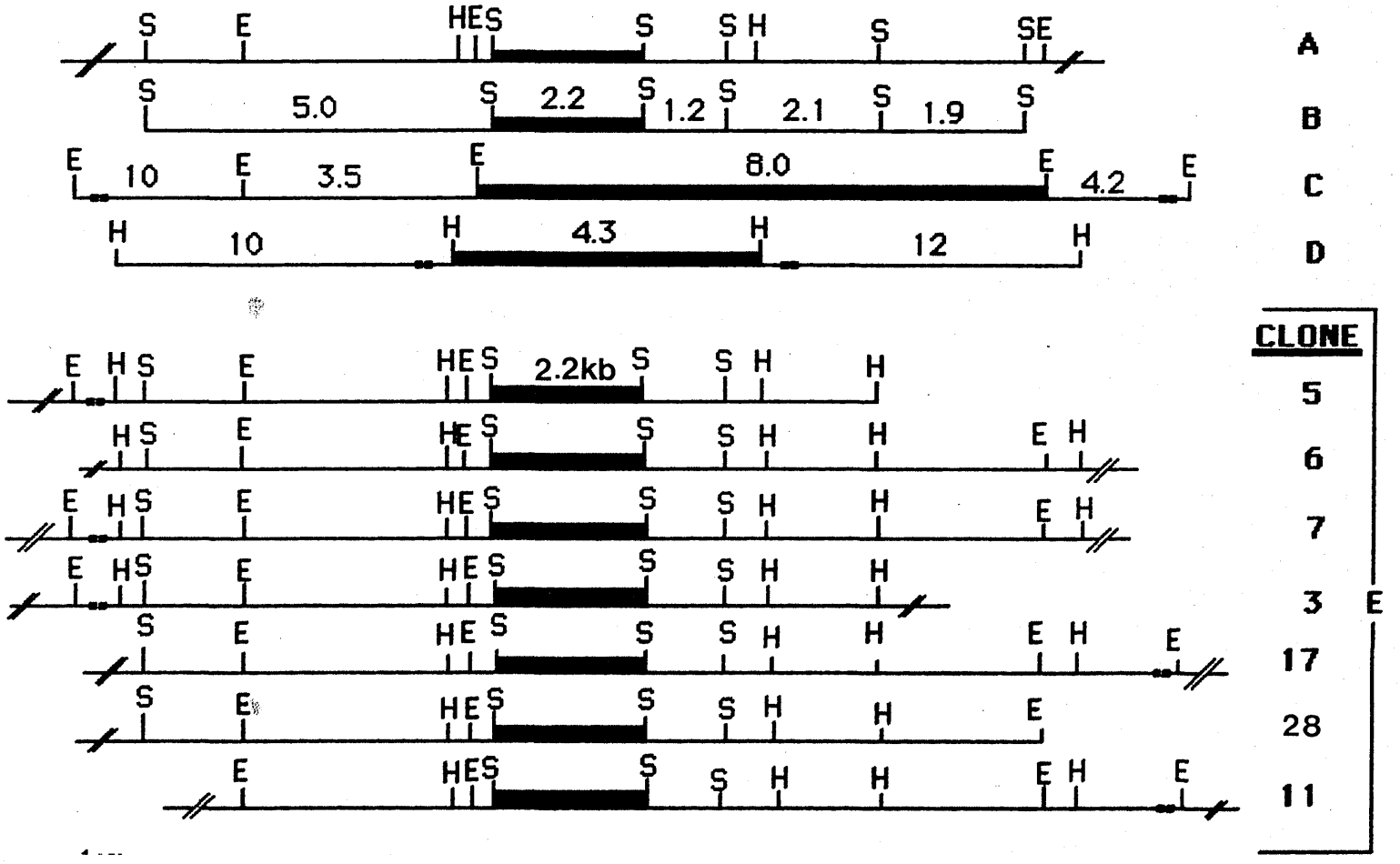

S E HES S S H S SE
A
S 5.0 S 2.2 S 1.2 S 2.1 S 1.9 S
B
E 10 E 3.5 E 8.0 E 4.2 E
C
H 10 H 4.3 H 12 H
D
CLONE
E H S E HES 2.2kb S S H H
5
H S E HE S S S H H E H
6
E H S E HE S S S H H E H
7
E H S E HE S S S H H
3
S E HE S S S H H E H E
17
S E HE S S S H H E
28
E HES S S H H E H E
11
E

Figure 7.10

A=The predicted restriction map of the genomic region of DNA encompassed by lambda clones 5, 6, 7, 3, 17, 28, and 11.
B, C, and D are the predicted restriction maps of this genomic region when cut with Hind III, SalI and EcoRI respectively. The sizes of the fragments generated are are shown in kilobases.
E=The restriction maps of the individual genomic clones.
—▮▮—Signifies a restriction fragment which is non-scale on the map and the size of which is either presented or can be determined from the available information.
—//—Signifies the end of the region mapped within both the genomic segment and lambda clones. Note that unless otherwise stated restricion fragments exist outwith these areas but have not been mapped.

Chapter 8 SEQUENCING: Generation of M13 mp19 recombinants

M13 mp19 recombinants were generated to provide single stranded template DNA for sequencing. The recombinants contained positively hybridising restriction fragments, obtained from a representative genomic clone, from each of groups 1, 2 and 3 (clones 9, 6 and 16 respectively).

The choice of the representative was made on the basis of information obtained from the restriction mapping studies. The positively hybridising region of the representative lay approximately centrally within the insert. The choice of such a clone reduced the possibility of sequencing a region of, for example, a structural gene, only to determine that the task could not be completed because the last few bases were "missing". In addition the use of such a clone would permit, at a later date, the subcloning and sequencing of upstream and downstream regulatory elements.

In order to generate small positively hybridising fragments for sequencing, the representative clone from each group was digested with a variety of restriction enzymes. The digested DNA was run on a gel, blotted, then probed with radiolabelled pLHIISB18. The following describes the process used to determine which fragments were suitable for subcloning.

Generation of fragments for sequencing from genomic clone 16

The phage DNA was digested with a variety of restriction enzymes, run on a gel, blotted and probed as described in chapter 3.8. The resulting gel photograph and blot autoradiograph are shown in figures 8.1 and 8.2.

Although the background on the Southern blot autoradiograph is higher than desirable, it is still clear that lanes 3, 4, 6, 7, 9,

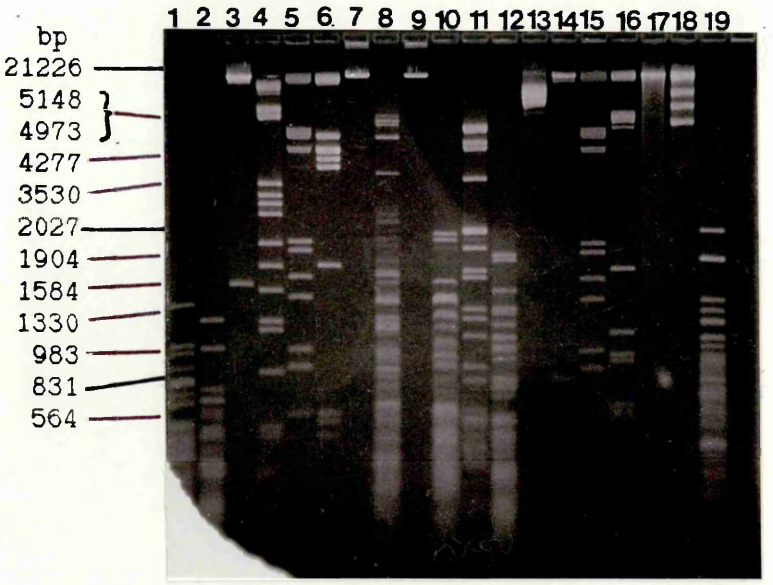


Figure 8.1

Genomic clone 16 DNA was digested with the following restriction enzymes.

Lane 1: HaeIII
Lane 2: HpaII
Lane 3: KpnI
Lane 4: PstI
Lane 6: PvuII
Lane 7: SalI
Lane 8: TaqI
Lane 9: XbaI
Lane 10: AluI
Lane 11: HincII
Lane 12: Sau3AI
Lane 13: SphI
Lane 14: SstI
Lane 16: SstII
Lane 17: XhoI
Lane 18: SmaI
Lane 19: HinfI

Lanes 5 and 15: λcI857 DNA cut with EcoRI+HindIII.

Note "single" ≈800bp SstI fragment in line 14.

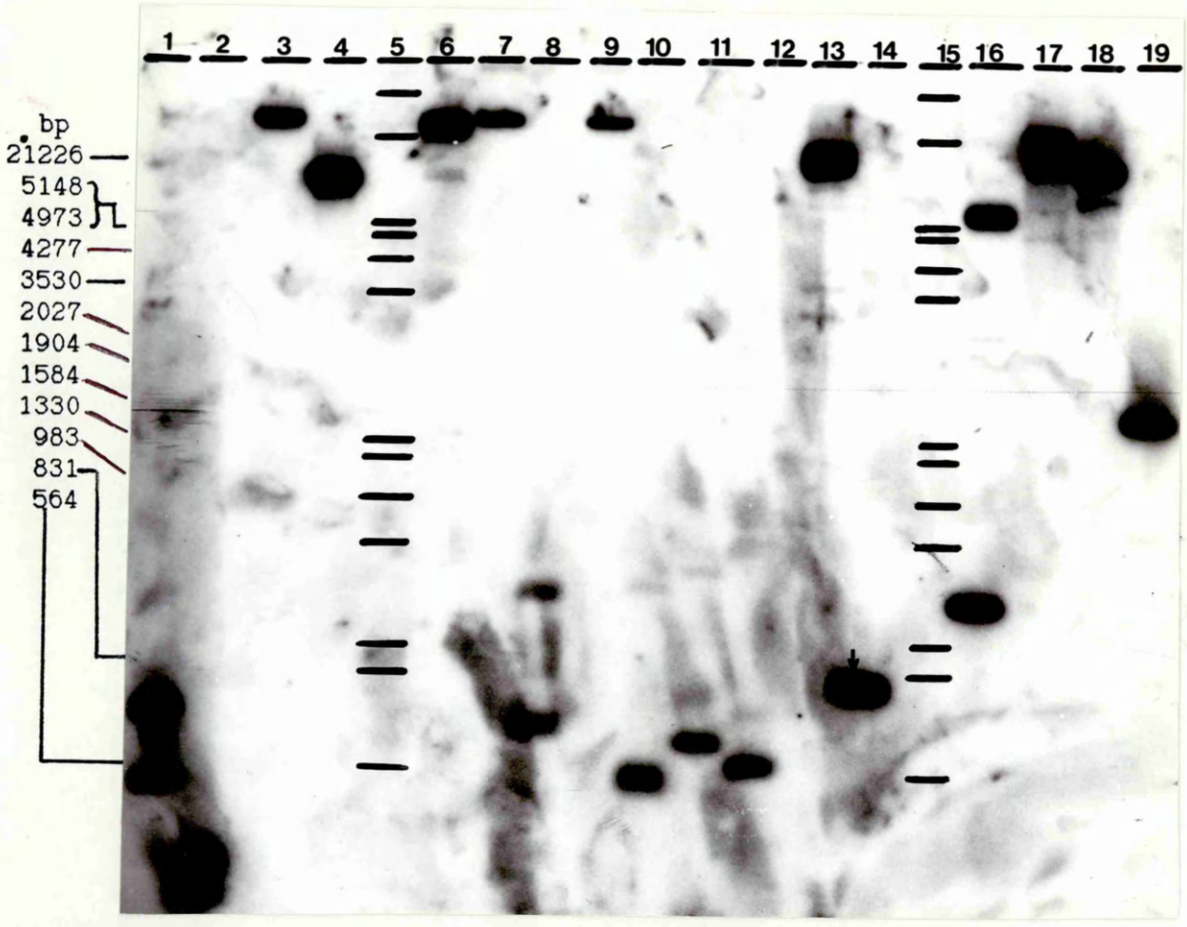


Figure 8.2

The gel shown in Figure 8.1 was blotted and the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown above.

Lane 1: HaeIII
Lane 2: HpaII
Lane 3: KpnI
Lane 4: PstI
Lane 6: PvuII
Lane 7: SalI
Lane 8: TaqI
Lane 9: XbaI
Lane 10: AluI
Lane 11: HincII
Lane 12: Sau3AI
Lane 13: SphI
Lane 14: SstI
Lane 16: SstII
Lane 17: XhoI
Lane 18: SmaI
Lane 19: HinfI

Lanes 5 and 15: λcI857 DNA marker fragments are marked in felt-pen.

Positively hybridising ≈800bp SstI fragment is shown arrowed in Lane 14.

13, 16, 17, and 18 each contain a positively hybridising band greater than 5kb in size. The DNA in these lanes was digested by the enzymes KpnI, PstI, PvuII, SalI, XbaI, SphI, SstII, XhoI and SmaI respectively. These fragments were not subcloned into mp19 because they were considered too large to form stable inserts. The enzymes used in these digests could therefore be dissmissed from the search of enzymes generating suitable fragments. It was assumed that the DNA in lane 7 had not cut with SalI because the enzyme was denatured.

Lanes 1, 2, 8, 10, 11, 12 and 14 contain DNA cut with HaeIII, HpaII, TaqI, AluI, HincII, Sau3AI and SstI respectively. With the exception of the SstI digested DNA, all the other digests appear to generate two or more positively hybridising fragments. In the cases of lanes 1 and 2 (HaeIII and HpaII respectively) the positively hybridising fragments are within the "ideal" size range for subcloning, see chapter 3.8. Digestion with each of these two enzymes generates two positively hybridising bands. By subcloning both the HaeIII and HpaII generated fragments, overlapping sequence data from both fragment sets could have been obtained. This would have allowed the order of the fragments to be determined.

Lanes 8, 10, 11 and 12 also contain double or triple positively hybridising bands. However, in these cases the larger of the positivly hybridising fragments are >1kb in size. If these relatively large fragments had been subcloned they could not be sequenced from opposite ends to give a central sequenced overlap.

However, the most suitable enzyme for generating a fragment for sequencing was SstI (Lane 14). This enzyme appeared to generate a single positively hybridising band of ≈800bp (later evidence will show this is not the case). A fragment of this size was considered just within the size limit which would allow sequencing from both ends with a central sequenced overlap. With hindsight this limit was too optimistic, 600bp being closer to reality!

In addition to size, examination of the gel, figure 8.1, suggested that this was the only SstI fragment genenerated by SstI digestion of this clone. Therefore, if the SstI cleaved clone DNA was mixed with and ligated to the vector, only this fragment would give rise to positively hybridising mp19 recombinants. This hypothesis could be tested by screening the resulting mp19 recombinants with radiolabelled pLHIISB18. If the hypothesis was correct all of the mp19 recombinants should give a positive hybridisation signal.

M13 mp19 RF and clone 16 DNAs were digested with SstI and an aliquot run on a minigel to check for cleavage. The DNAs were then mixed, ligated then used to transform JM101 competent cells. The cells were then plated (onto four plates) see chapter 3.8. Filters were then lifted from the plates and screened with radiolabelled pLHIISB18. The autoradiograph of one of these filters is shown in figure 8.3.

On this plate there were twenty recombinant plaques (white). All twenty gave a positive hybridisation signal, however, ten of these were strong positives, and ten weak positives see figure. In addition there were fifty seven non recombinant plaques (blue). They did not give a visible hybridisation signal. This result was typical of the other three plates.

As predicted, all of the recombinants hybridised positively. However the reason for the difference in signal intensity between positives was unclear. Initially it was thought this may be a result of having the one fragment in two possible orientations within the vector. This could have led to different growth rates of the recombinants and would have accounted for the 1:1 ratio of weak:strong positives. Later experiments (see chapter 9) however, suggested that the SstI "fragment" was in fact two, virtually unresolvable fragments. Therefore, the difference in plaque intensity may reflect the differences in the degree of hybridisation of two different inserts with the probe, or the rate of growth of different recombinant phage types.

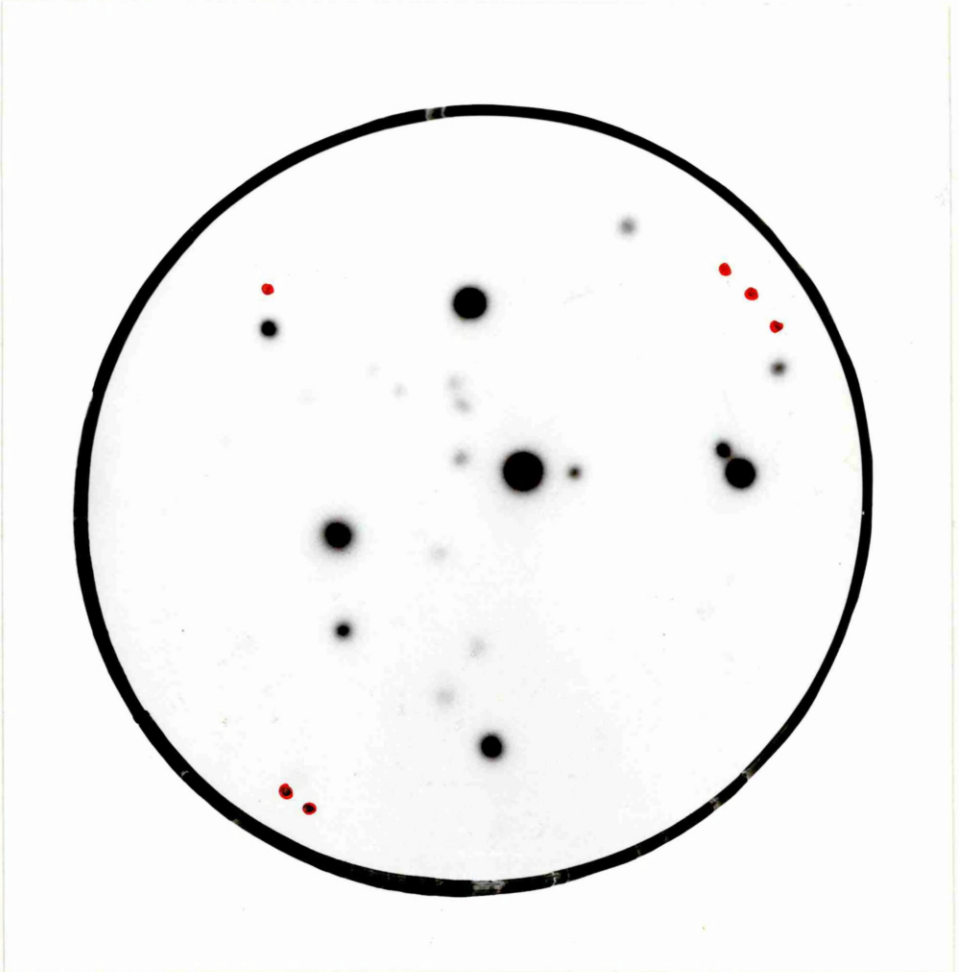


Figure 8.3

M13 mp19 and genomic clone 16 DNAs were digested with SstI, the DNAs mixed, ligated, then used to transform JM101. This gave 57 non-recombinant plaques (blue) and 20 recombinant plaques (white). Filters were lifted from the plates and screened with radiolabelled pLHIISB18. An autoradiograph of one of these filters is shown above.

All twenty of the recombinant plaques hybridised positively. However, they gave either strong or weak hybridisation signals. This result was considered inconsistent with the idea of there being a single type of positively hybridising SstI fragment.

● = filter orientation marks.

The positive plaques were picked, rescreened, then used to generate single stranded templates for sequencing.

Generation of fragments for sequencing from genomic clone 6.

Lambda clone 6 was digested with a series of enzymes, then the digests were run on a 0.8% gel, see figure 8.4. The gel was blotted. and the filter probed with radiolabelled pLHIISB18. The filter autoradiograph is shown in figure 8.5. The high background on this blot has unfortunately obscured some of the weaker, positively hybridising bands.

Lanes 1, 2, 6, 10, 11, 12, 16, and 17 all contain single positively hybridising bands $\geq$ ≈3kb in size. Fragments of this size were considered too large to be stable in mp19 and were therefore deemed unsuitable for subcloning. The enzymes used in these digests were ClaI, AccI, PstI, SstII, XhoI, KpnI, HpaI, and SphI respectively.

Lane 8 (XbaI) also contains a positively hybridising, high molecular weight, band. However, this reflects the fact that the phage DNA did not cut with the enzyme in this reaction. A digest of genomic clone 9 carried out at the same time produced a similar result. This suggested that the enzyme was denatured at the time of use.

Lane 4 (RsaI) contains a positively hybridising band of ≈500bp, but unfortunately also contains a positively hybridising band of ≈20kb.

Lanes 3 (AluI) and 13 (HincII) each contain two positively hybridising bands, a low molcular weight band of <500bp and a higher molecular weight band of ≈700 and ≈800bp respectively. These fragments would have been within the suitable size range as previously outlined. However, to simplify the sequencing the positively hybridising, 700bp SstI fragments (lane 9) were selected for subcloning into SstI cut mp19. It was hoped that this clone may have a similar genomic organisation to clone 16. If this were the

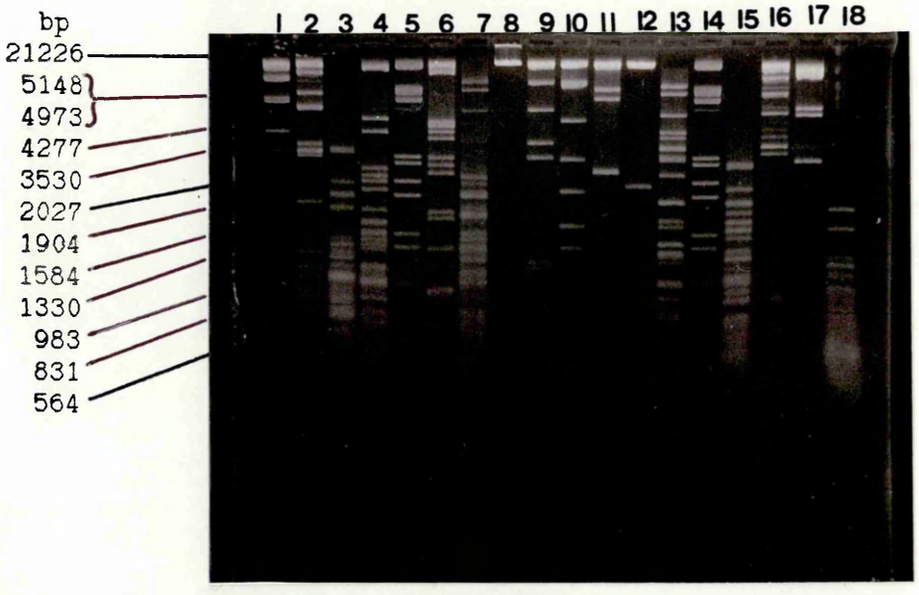


Figure 8.4

To generate suitable fragments for sequencing, genomic clone 6 DNA was digested with the following restriction enzymes.

Lane 1: ClaI	Lane 10: SstII
Lane 2: AccI	Lane 11: XhoI
Lane 3: AluI	Lane 12: KpnI
Lane 4: RsaI	Lane 13: HincII
Lane 6: PstI	Lane 15: Sau3AI
Lane 7: TaqI	Lane 16: HpaI
Lane 8: XbaI	Lane 17: SphI
Lane 9: SstI	Lane 18: HpaII

Lanes 5 and 14: λcI857 DNA cut wuth EcoRI+HindIII.

Note ≃ 700 bp SstI fragments in Lane 9.

Lane 1: ClaI
Lane 2: AccI
Lane 3: AluI
Lane 4: RsaI
Lane 6: PstI
Lane 7: TaqI
Lane 8: XbaI
Lane 9: SstI
Lane 10: SstII
Lane 11: XhoI
Lane 12: KpnI
Lane 13: HincII
Lane 15: Sau3AI
Lane 16: HpaI
Lane 17: SphI
Lane 18: HpaII

Lanes 5 and 14: λcI857 DNA cut with EcoRI+HindIII.

Note positively hybridising SstI band in lane 9 (arrowed).

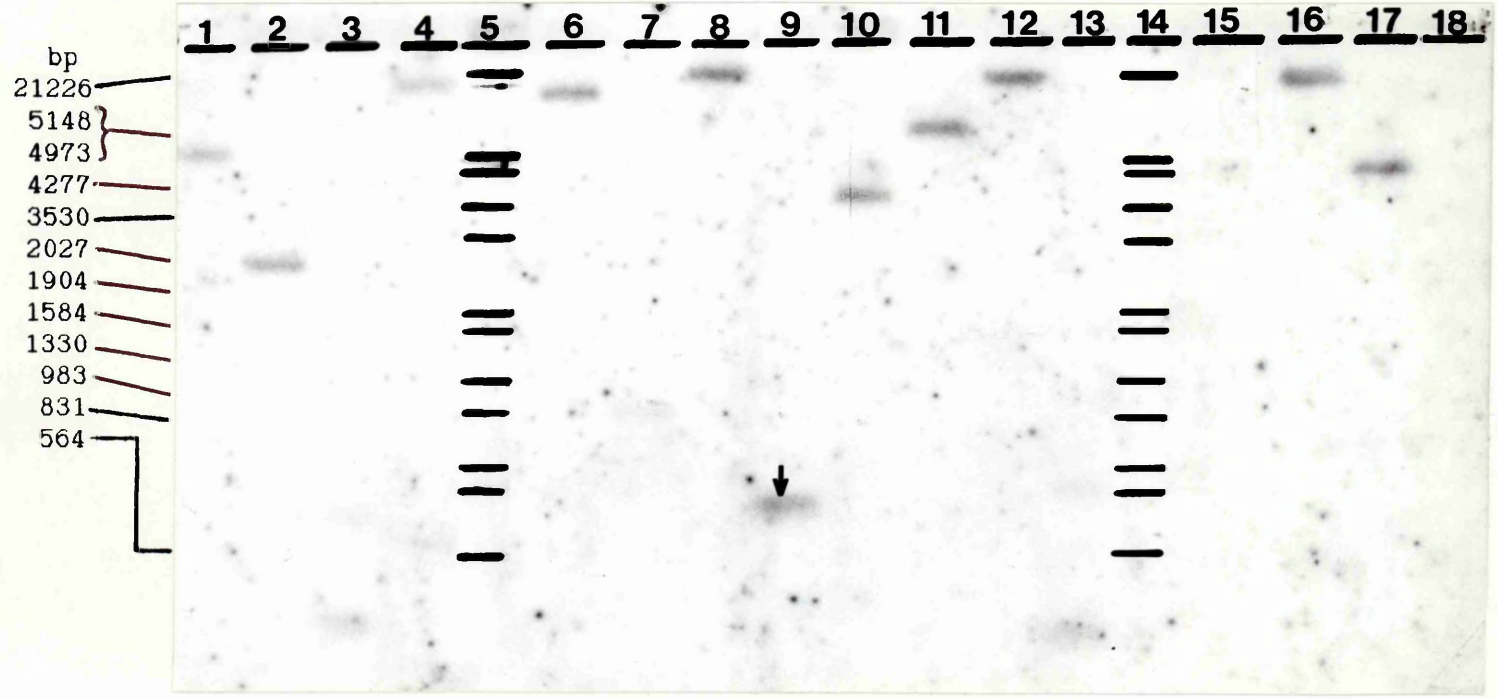


Figure 8.5

The gel shown in Figure 8.4 was blotted and the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown above.

case then it was thought likely that only one of the two 700bp bands would hybridise positively. The recombinants were transformed then screened as described previously.

Although not all the recombinant plaques hybridised positively (as was to be expected) those that did had differential signal strengths equivalent to those described for mp19-genomic clone 16 recombinants. In this case the ratio of weak to strong positives was ≈1:2.

Analysis of the SstI digest of clone 6 had shown that there were two SstI fragments of ≈700bp in size (figure 8.4). It did not occur at that time that positively hybridising regions may exist on both fragments. The "second" 700bp fragment was considered as part of the flanking DNA. Even now, there is no 'definitive' evidence which suggests that this is not the case. However, circumstantial evidence, which will be presented later, does suggest that both of these fragments may carry light harvesting genes.

The ratio of differences in signal strength from the positive plaques, combined with the high background of the autoradiograph prompted a repeated blot of clone 6 digested with SstI. The resulting gel and autoradiograph are shown in figure 8.6. This experiment showed clearly that there were at least two positively hybridising SstI fragments of ≈0.7 and ≈2.2kb. In addition a difference existed in the strength of their hybridisation signal. However, the difference of signal strength between the two bands could not fully account for the differences in signal strength between the different positively hybridising plaques.

It could be argued that the weaker plaques contained the larger SstI insert which may have slowed their growth and generated fewer phage for hybridisation. A process such as this, in addition to any differences in affinity of the inserts for the probe, may account for observed differences in plaque signal strength. However, evidence is presented in the next chapter which suggests that the

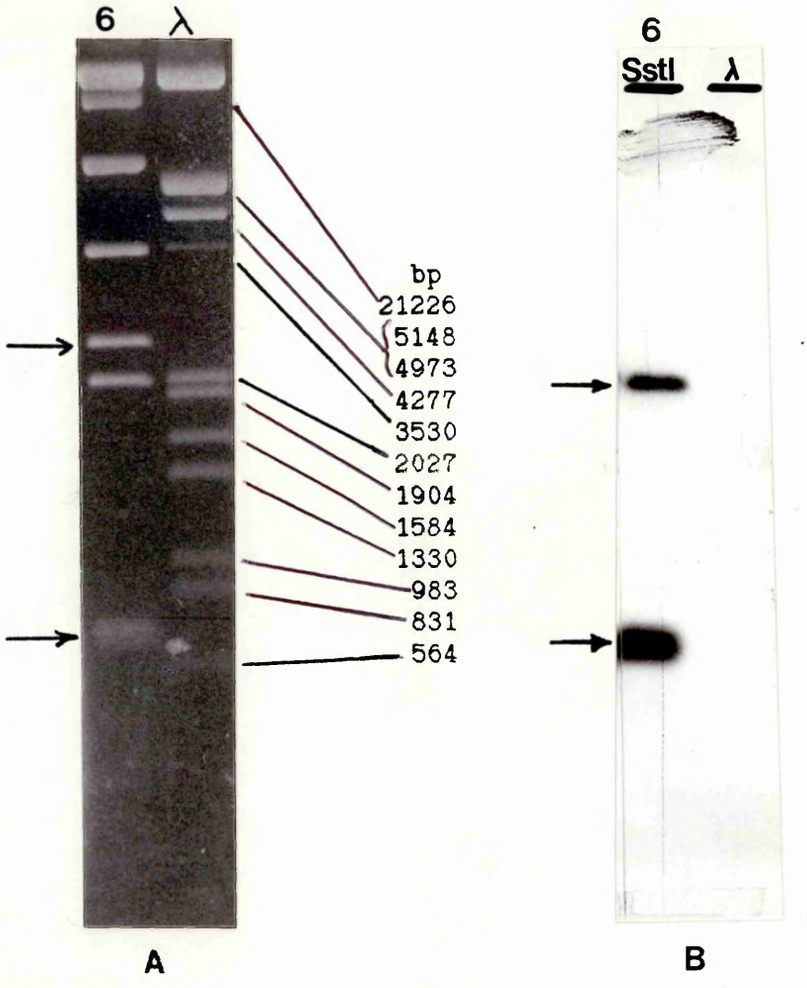


Figure 8.6

Genomic clone 6 was digested with SstI then the digest run on a 0.8% agarose gel (A). The gel was blotted and the filter probed with radiolabelled pLHIISB18 (B). λ= λ cI857 DNA digested with EcoRI+HindIII. Note that photographs A and B are not to the same scale.

The positively hybridising bands are arrowed.

differences in signal strength may not be due to the differential growth rates of 0.7 and 2.2kb fragment containing clones. Instead it is hypothesised that the ratio of 2:1 strong to weak plaques is due to a strong hybridisation signal from the 2.2kb and one of the 0.7kb inserts. The weak signal is thought to have arisen from the other 0.7kb insert.

Selection of fragments for sequencing from genomic clone 9

Genomic clone 9 was digested with a variety of restriction enzymes. The digests were run on a gel which was then blotted and the filter probed with radiolabelled pLHIISB18. The gel photograph and filter autoradiograph are shown in figures 8.7 and 8.8. From these it can be seen that positively hybridising fragments in lanes 1, 2, 6, 9, 10, 11, 12, 16 and 17 are >3.5kb in size, and therefore from the parameters described previously, unsuitable for subcloning into mp19. The DNA in these lanes was digested with ClaI, AccI, PstI, SstI, SstII, XhoI, KpnI, HpaI, and SphI.

Again XbaI (lane 8) has failed to cut the DNA presumably because the enzyme was inactive.

The fragments considered most suitable for sequencing were the positively hybridising ≈600bp AluI fragment (Lane 3) and the ≈500bp HpaII fragment (Lane 18). These were subcloned into the SmaI and AccI sites of the mp19 polylinker respectively. In addition the two HincII fragments of 0.7 and 1.6kb (Lane 13), and the RsaI fragments of 0.5 and 1.2kb (Lane 4) were also cloned. These were all inserted at the SmaI site of the mp19 polylinker.

The most surprising feature of the autoradiograph shown in fig. 8.8 was the large number of positively hybridising bands present in lane 7 where the DNA had been digested with TaqI. Initially this was thought to be the result of partial digestion of the DNA. However,

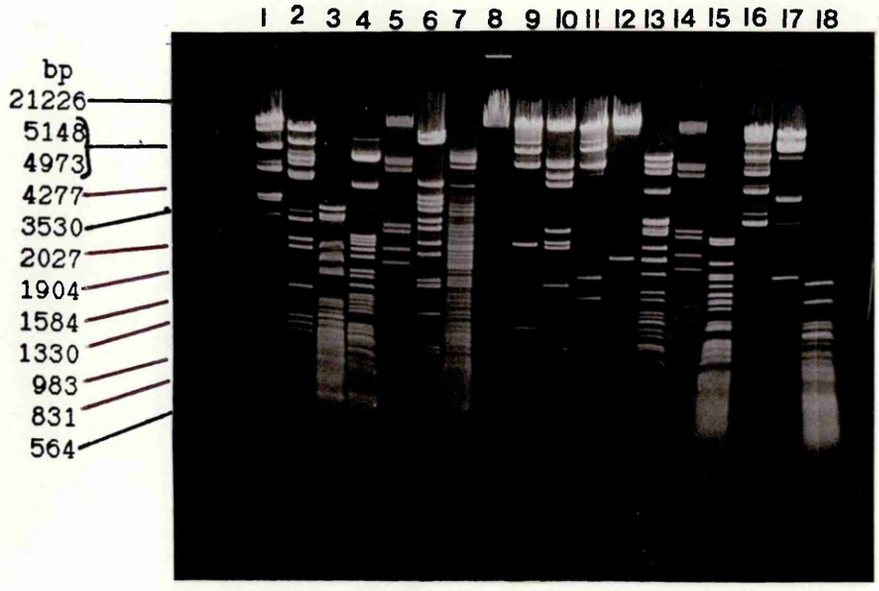


Figure 8.7

Genomic clone 9 DNA was digested with the following restriction enzymes.

Lane 1: ClaI	Lane 10: SstII
Lane 2: AccI	Lane 11: XhoI
Lane 3: AluI	Lane 12: KpnI
Lane 4: RsaI	Lane 13: HincII
Lane 6: PstI	Lane 15: Sau3AI
Lane 7: TaqI	Lane 16: HpaI
Lane 8: XbaI	Lane 17: SphI
Lane 9: SstI	Lane 18: HpaII

Lanes 5 and 14: λcI857 DNA cut with EcoRI+HindIII.

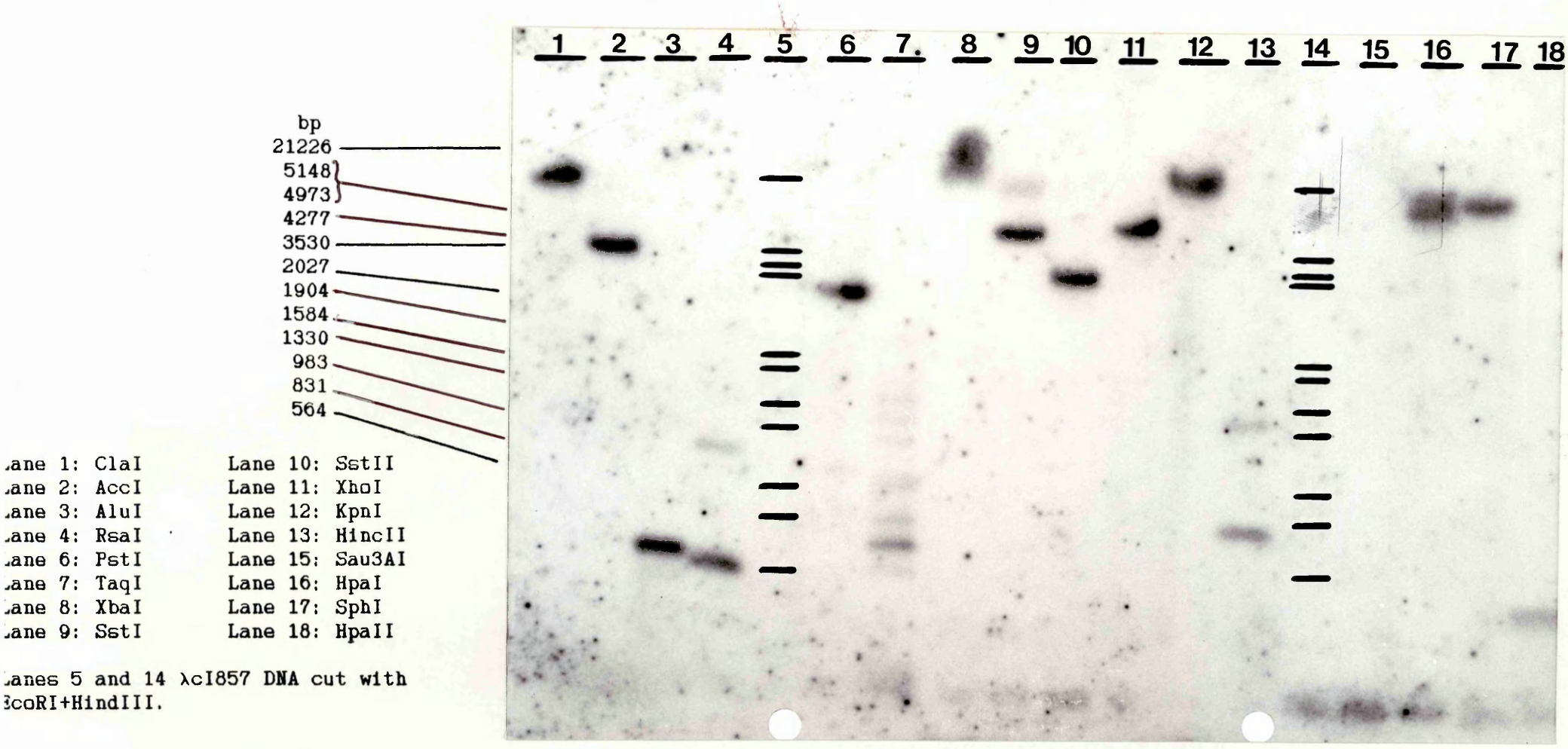


Figure 8.8

The gel shown in Figure 8.7 was blotted and the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown above.

there was no reason to assume that partial digestion had occurred. The batches of buffer and TaqI enzyme used to digest the DNA were the same as that used for the digestion of clone 6, described earlier. Both digestions were carried out at the same time and set up in parallel. Clone 6 however, did not show a similar multiple banding pattern. In addition, there was no reason to believe that the lambda DNA was impure and slowing enzyme digestion. Evidence for this came from the banding pattern produced by digestion of this DNA with the other enzymes. In these digests partial digestion did not appear to have occurred. It therefore appeared that this banding pattern was genuine and not a partial digestion artefact.

Restriction mapping had shown that the positively hybridising region of this clone was localised to two EcoRI fragments of 1.9 and 5.0kb in size and that these fragments were contiguous. That is, the positive region was localised on a 6.9kb partial EcoRI fragment. This suggested that the positive TaqI fragments were not scattered throughout the cloned insert but were localised within this ≃7kb region.

The assumption was then made that the the positively hybridising TaqI fragments were contiguous within the 7kb region. If this were the case then the size of the contig would be the sum of the positively hybridising fragments. The deduced sizes of these fragments were approximately: 2.0, 1.6, 1.4, 1.2, 2 x 0.9, 0.8, 2 x 0.7 and 0.5kb giving a total size of ≃10.2kb to the contig. This was clearly too large to fit within the ≃7kb region. However, if such a contig did exist it was possible that non-hybridising, flanking DNA was also being included in the total length of the hybridising region. It seemed reasonable to assume that TaqI fragments containing large amounts of flanking DNA and only a small part of a light harvesting gene, would hybridise more weakly than those containing only positively hybridising regions. The two weakest hybridising bands were 2.0 and 1.6kb in size. Although they clearly contained hybridising regions, it was considered too difficult to assign sizes to these regions and therefore these

fragments were removed entirely from the positively hybridising contig. This gave the positively hybridising contig a size of 6.6kb which fitted within the 6.9kb EcoRI partial fragment. Although this type of modelling was very speculative, what it did suggest, if correct, was one of the following.

a) The region contained multiple light harvesting genes or at least multiples of one of the components of such alpha-beta gene pairs.

b) Within the positively hybridising region lay a gene which was not a light harvesting gene. It was possible that such a region may have had sufficient homology to the pLHIISB18 probe to permit hybridisation. However, the bound probe may not have been removed under the conditions of low stringency used for washing the filters.

Evidence will be presented in the next chapter which suggests that the first model is correct and that at least six genes enoding three alpha and three beta light harvesting polypeptides lie within this region.

If such a structure did exist, how would it be possible to clone the entire hybridising region as a single 0.7kb AluI or 0.5kb HpaII fragment? One possibilty would be that if multiple copies of genes existed they had arisen by gene duplication. If this was the case then not only the structural regions but also the flanking restriction sites may have been conserved. It was therefore possible, that the single positively hybridising AluI and HpaII bands were composed of identical or at least similar sized fragments which possesed either the same or slightly differing nucleotide sequences.

The possibility of such an anomaly resulted in the re-examination of this clone digested with HpaII and AluI. The DNA was digested with these two enzymes, run on a gel, blotted and probed as described previously see figure 8.9.

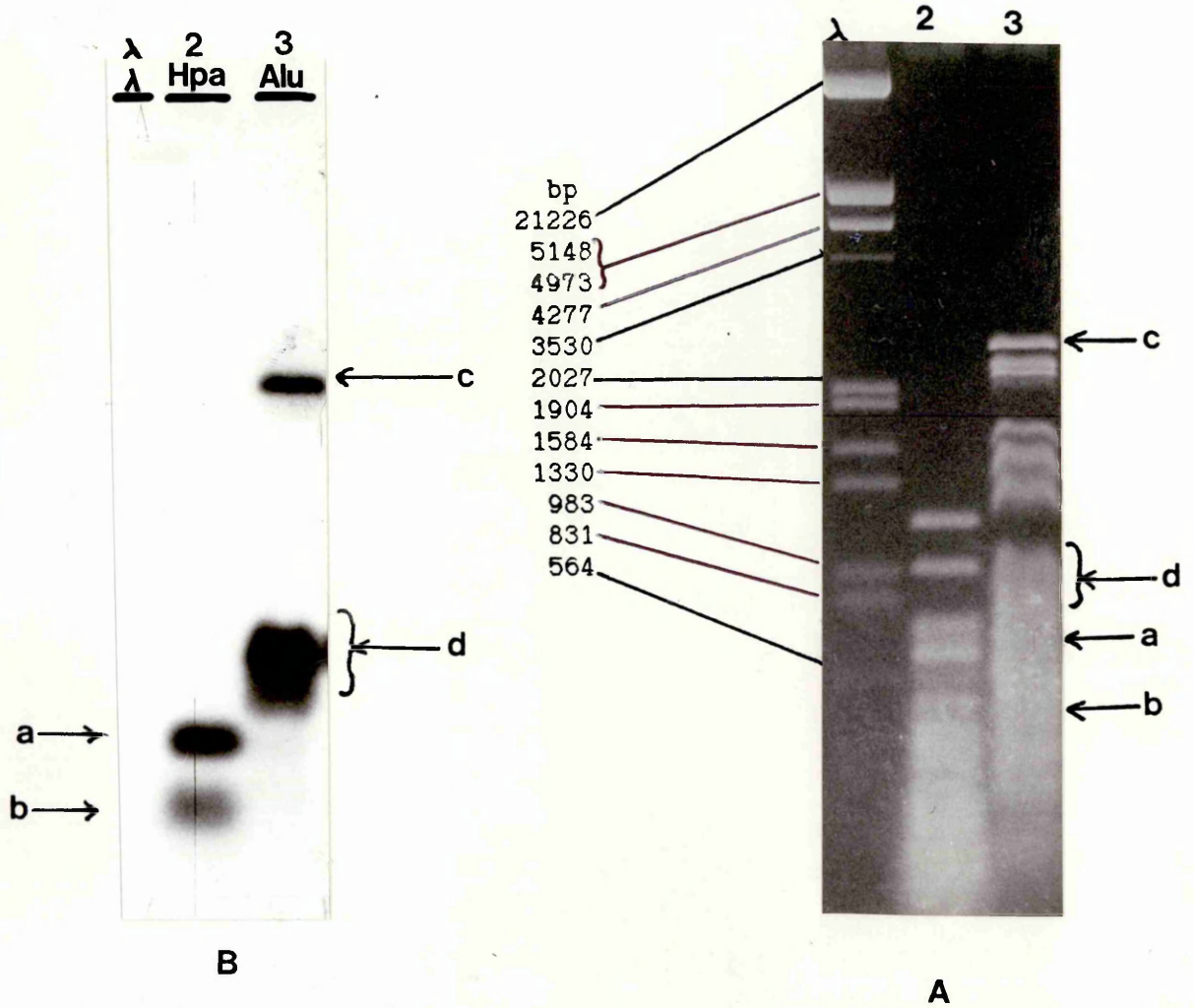


Figure 8.9

Genomic clone 9 was digested with HpaII (lane 2) and AluI (lane 3) and run on a 0.8% agarose gel (A). The gel was blotted and the filter probed with radiolabelled pLHIISB18 DNA (B). The positively hybridising bands in B are arrowed in A.

Note A and B are not the same scale.

λ = λcI857 DNA digested with EcoRI+HindIII.

The autoradiograph shows that both the AluI and HpaII digests resulted in more than a single hybridising band. The AluI digest generates a band of ≈2.2 kb and what appears to be a population of small fragments of between ≈500 and 700bp in size. HpaII, however, appears only to generate two positively hybridising bands both of less than 500bp in size. However, sequencing later showed that a number of HpaII fragments of ≈250bp were found within positively hybridising regions of this clone. Suggesting that the although the positively hybridising regions were considered as single bands each may have been composed of a population of similar sized DNA fragments.

CHAPTER 9 SEQUENCING THE SINGLE STRANDED TEMPLATES

The positively hybridising mp19 constructs, described in the previous chapter, were used to generate single stranded DNA templates which were then sequenced by the dideoxy chain termination method. The data presented in this chapter is the result of these sequencing reactions. All sequences shown were read twice from the sequencing autoradiographs then loaded into a VAX computer using Seqed, the sequence editing facility of the Genetics Computer Group (GCG) programme package. The sequences were then re-entered and checked using the "Check" facility within the Seqed programme.

The Bestfit programme was then used to compare the nucleotide sequences of the templates to that of pLHIISB18. The comparison was made with both the "Forward" and "Reverse" facilities within the programme. "Forward" was used to compare the entered template sequence to the sequence of pLHIISB18. "Reverse" was used to compare the computer generated reversed and complemented template sequence to the sequence of pLHIISB18.

The Bestfit programme, as its name suggests, finds the best match between two DNA sequences. However, this means it always finds some alignment between two sequences no matter how tenuous that relationship may be. It is then up to the system operator to use their judgement as to how likely the generated alignement is of being "genuine" or an "artefact" of the programme. In reality this was not difficult, genuine regions of alignment stretched over considerable distances with relatively few gaps or base mismatches between the compared sequences. Artefactual alignments, usually extended over ten to twenty bases, sometimes with mismatches and, or, gaps in the alignment. Any cases which were considered borderline, ie. about 30 bases of alignment between the template and the pLHIISB18 sequence, were treated in the following way. The entire template sequence was translated in the six possible reading

frames. The deduced amino acid sequences were then compared to each of the available *R. acidophila* light harvesting polypeptide amino acid sequences in a bid to find regions of similar primary sequence. However, none of the borderline cases ever generated an alignment with the *R. acidophila* sequences which extended over more than four amino acids.

The "genuine" template alignments were translated and compared in an identical way to the borderline cases. These alignments are presented in this chapter.

In the case of genomic clone 9, where a variety of enzymes had been used to generate fragments for sequencing, the Gelassemble programme was used in addition to the programmes described above. This programme was used to compare clone 9 template sequences (called contigs) to each other. Where overlapping contigs were found, the programme was then used to meld the contigs together to form a continuous sequence. This continuous piece of sequence is called the consensus sequence. However, I have also used this term to describe sequences where overlaps or joins have not been made.

In this chapter the M13 mp19 constructs 18SstI1-4 (see below) have been used as an example of the type of analysis used on all other template sequence data. Only in this case have poor Bestfit alignments been shown. Likewise, an analysis and discussion section for this particular set of constructs has been included, but the points presented apply equally well to the sequences obtained from other constructs. An in depth discussion of the sequences and their analyses are presented in Chapter 11.

Sequence data obtained from the 16SstI series of constructs

Digestion of genomic clone 16 with SstI appeared to generate a single SstI fragment of approximately 800bp. Southern blot analysis suggested that this fragment hybridised positively when probed with pLIISB18. For sequencing, the clone had been digested with SstI and

the fragments subcloned into the SstI site of mp19. *In situ* plaque hybridisation had shown that all the recombinant (white) plaques hybridised positively when probed with pLHIISB18. This was to be expected if only a single type of positively hybridising fragment was available for insertion into the mp19 vector. However, the intensity of the hybridisation signal from the mp19 recombinants was differential. There was a ratio of approximately 1:1 of strong : weak signals from the plaques. These differences were attributed to differential rates of growth of the host and/ or differences in growth rate of the phage possibly due to the effects of having oppositely orientated inserts.

Sixteen positively hybridising plaques were picked and purified. Of the sixteen, seven gave a strong, and nine a weak hybridisation signal. Single stranded sequencing templates were prepared for each of the recombinants. It was envisaged that because a single SstI fragment had been subcloned that sequencing would reveal two nucleotide sequences; a consequence of the two possible orientations of the insert within the vector.

Initially five templates were sequenced. However, instead of the expected two, three nucleotide sequences were found. This lead to the sequencing of the remaining eleven templates which revealed an additional piece of sequence. These four sequences have been named 16SstI1, 2, 3 and 4. They were found in a ratio of 3:4:4:5 respectively. Templates 16SstI1 and 2 were derived from weakly hybridising plaques, 16SstI3 and 16SstI4 from strongly hybridising plaques. The gel autoradiographs of the sequenced templates are shown in figure 9.1. The corresponding nucleotide sequences are presented in figure 9.2. Please note that the nucleotide sequences of 16SstI2 and 3 in these figures contain additional sequence data not presented in figure 9.1. This additional sequence was obtained by carrying out the gel electrophoresis for 1½ and 7 hours. The autoradiographs of these gels are presented in figures 9.3a and b. The nucleotide sequences were then compared to the pLHIISB18 sequence in both the Forward and Reverse orientation.

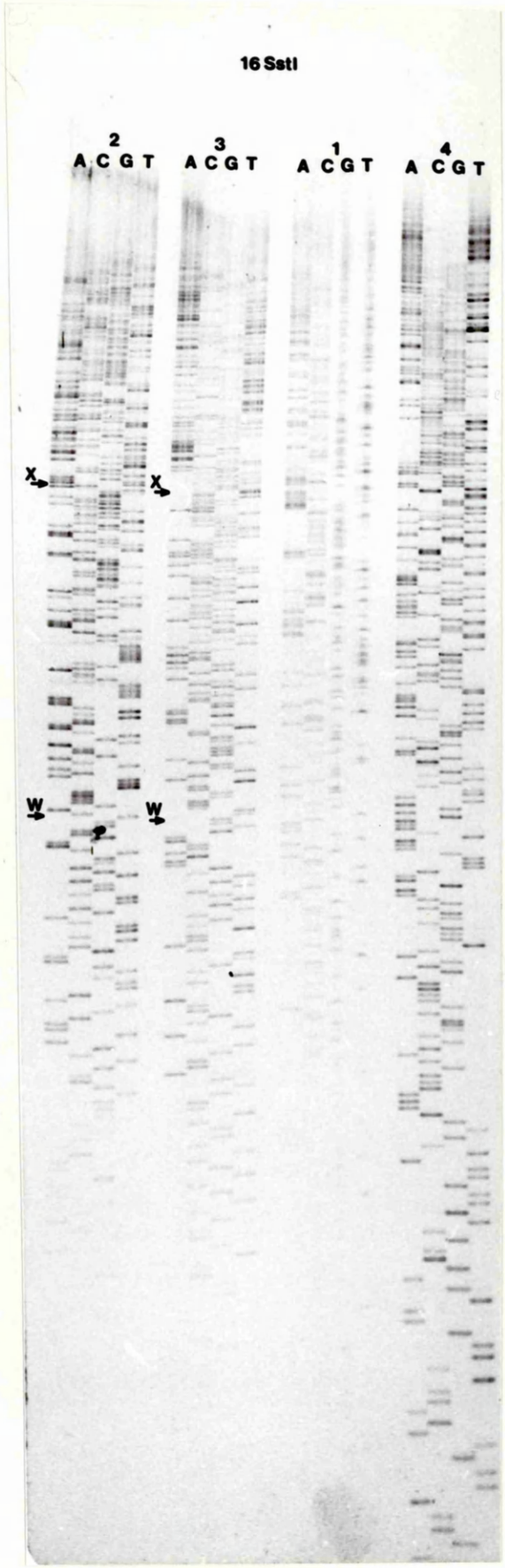


Figure 9.1

Shown opposite is the gel autoradiograph of sequences representative of sequences 16SstI1-4. These are shown in the order 16SstI2,3,1,4.

The gel has been run for ≈4 hours under the conditions described previously.

The nucleotide sequences taken from these autoradiographs are shown overleaf.

X and W are reference points, at least one of which can be seen on the autoradiographs of gels run for shorter and longer times, see later.

```
  1  GCTCTCGGTG GCGATCGTCC ATCAGCATGT CCTTGCGGCC GGGATCGACC
 51  TCCCCGAGCG CCTGGGTGAA ATAGCGCATG GCCGCGGCCA GCTCGCCCTG
101  CGGGCCGCCG AACTGCTCCA GCATGAAACT CGCGAGGACG GGATTCGGTC
151  GGGCGACCCT GACCGTGTAT TGCAGTTTCT TGTTGTGAAT GAACATGAAG
201  CCTCCGCATG CGTGCCTGGG TCGACAACGG GCGGCGCGGC GCGGCGTTCC
251  GAAATATAAA TTGAACAAAG CTTTAGGTTG CGTCGGCGCG AT
```

Nucleotide sequence 16SstI1, see Figure 9.1.

```
  1  CTCCTTAGTG CAGCCAGGGG GAGAAAGCGA ACGCGAGAAC GTGCGCGAAG
 51  ATCGCGATCA CGAAGAACAC GCGGGTGCCG TCGATAACGT ACTTGTGAAG
101  CTCTTCAGCT TCGGCTGCGG TCAAGCCGGT CAGCCCCTTT ACATCAGCCA
151  TGTCAACCTC TCAAATTTTT CGCCACGCTT TTTTTCGTGC TAAACCGCAC
201  GTTGATCACC GGTGGCGCGG TGGCCGAACT TGTGCGAAAC TGAGCTTCGC
251  CGTCGCAGCG CAGGTGTCAC TATAGATTGC CGCTTTGTAA TGTCAAATTC
301  ACTTGACATG GAAACGCATT TCTGAACGAC ACGTGA
```

Nucleotide sequence 16SstI2, see Figure 9.1.

```
  1  CTCCTGACAT GAACCAGGGC AAGATCTGGA CCGTCGTCAA CCCGAGCGTT
 51  GGCCTGCCGC TGCTGCTCGG CTCCGTGACC GTCATCGCCA TTCTCGTTCA
101  CCTCGCTGTG CTGTCGAACA CCAAGTGGTT CCCGGCCTAC TGGCAGGGTG
151  GCCTGAAGAA GGCCGCTGCG ATCGAGACCA CGATCGTCGG CTGACGCCGC
201  CTGACGGCTG CCGCATGTGA CGGCACACCG CCGATGGTGT GCGGCGC
```

Nucleotide sequence 16SstI3, see Figure 9.1.

```
  1  TNTTCACATT GCTCCGGAGT CAGAACAGCC ATTGTACACC TCTTGAATGA
 51  GCCGCTGTTG TTATCGTGCA AACCCGCACG CGGGCGCCCC AGCGACTGGG
101  CGGCGCAAGA ACTTTGAAAT CAAAGAGATA TTCGTCGTCC GAACGCAGGT
151  GTCATCATAA ATTGACGCCG ACGGAGATCA CTTGGGTTGA CATAAGAGTA
201  CGTAGAATTC GG
```

Nucleotide sequence 16SstI4, see Figure 9.1.

Figure 9.2

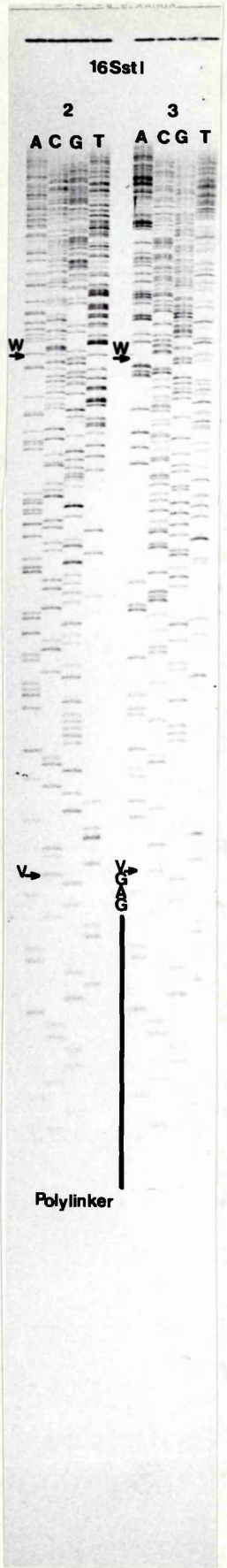


Figure 9.3a

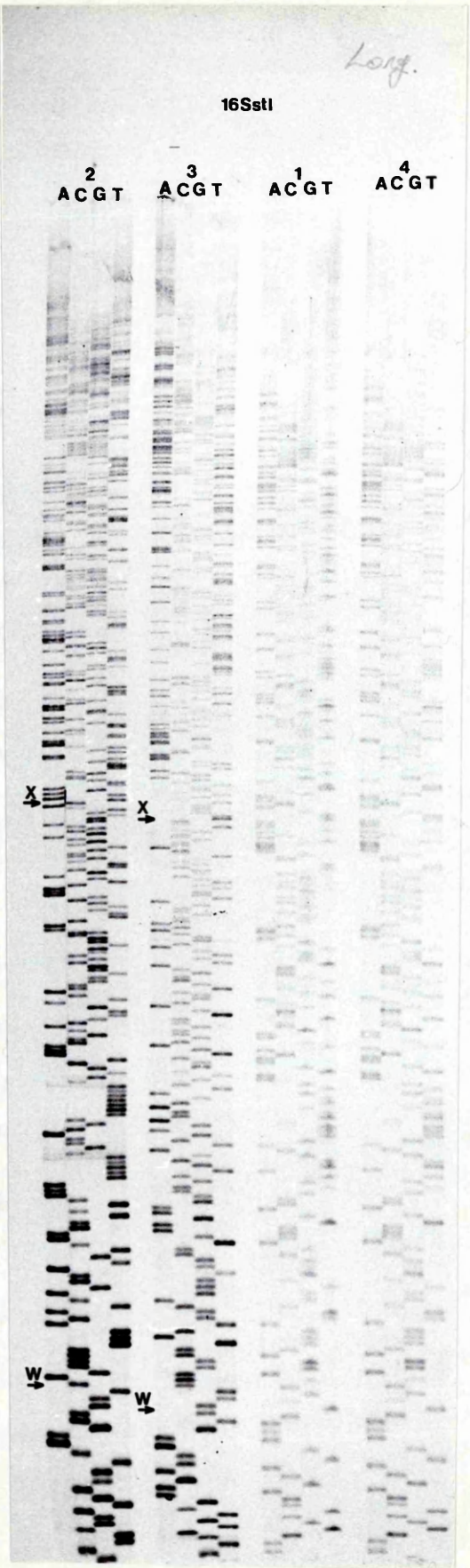


Figure 9.3b

Figure 9.3a

Sequencing autoradiograph showing the 16SstI2, and 3 sequences at the junction of the mp19 polylinker and SstI insert. Re-creation of the SstI site at this position (V) is signified by the sequence GAG (marked) followed by CTC. This gel was run for 2 hours. Reference point W from Figure 7.1 is shown.

Figure 9.3b

Sequencing autoradiograph showing 16SstI1-4 sequences after 7 hours electrophoresis. W and X are reference points shown also in Figure 9.1 and 9.3b.

Sequences 16SstI1 and 4 showed insignificant alignment in both Forward and Reverse Bestfit analyses, see figure 9.4 for example.

16SstI3 showed significant alignment with the pLHIISB18 sequence. The alignment began at the 4th nucleotide of 16SstI3 and extended through to nucleotide 148. This sequence showed 75% identity with the region from nucleotide 506 to 650 of pLHIISB18 which encodes the *Rb. sphaeroides* B800-850 alpha polypeptide, see figure 9.5a.

16SstI2 showed insignificant alignment with the pLHIISB18 sequence in the Forward direction 9.5b. However, in the Reverse (ie. reverse and complement) orientation there was 72% identity between the region from bases 12 to 135 of the 16SstI2 sequence and the region from base 365 to 488 of the pLHIISB18 sequence see figure 9.5c. This region of pLHIISB18 encodes the *Rb. sphaeroides* B800-850 beta polypeptide.

Before carrying out further analysis of the data, I would like to return to the hypothesis that complete digestion of lambda clone 16 with SstI resulted in the generation of a single, positively hybridising, 800bp fragment. The following pieces of evidence suggest that this is hypothesis is incorrect. Firstly, it was expected that sequencing of the templates would have resulted in two different nucleotide sequences. It was envisaged that these sequences would correspond to the 5' regions of the two complementary strands of the 800bp fragment. However, instead four different nucleotide sequences were found, suggesting the presence of at least two positively hybridising fragments.

If there had been three or more positively hybridising fragments it would have been expected that fifth and sixth nucleotide sequences would have been uncovered; such sequences were not found. If a third non-hybridising fragment was present it would then have been expected that at least a proportion of the mp19 recombinants would have not hybridised when probed with pLHIISB18. However, all the

```
          Quality:   12.5              Length:      41
            Ratio:  0.305               Gaps:       0
Percent Similarity: 63.415   Percent Identity: 63.415

16sstil.Dat x Sphaerlh2.Dat April 7, 1990  20:21  ..        A

                .         .         .         .
     60 GCGCCTGGGTGAAATAGCGCATGGCCGCGGCCAGCTCGCCC 100
        ||| | | || |    ||||| |||   |||| ||  ||||
    664 GCGGCCGAGTAATGCTGCGCAAGGCGCGGGCCTGCGGGCCC 704


          Quality:   13.5              Length:      42
            Ratio:  0.321               Gaps:       0
Percent Similarity: 64.286   Percent Identity: 64.286

16sstil.Dat x Sphaerlh2.Dat April 7, 1990  20:24  ..        B

                .         .         .         .
    243 CGCCGCGCCGCCCGTTGTCGACCCAGGCACGCATGCGGAGGC 202
        | ||  || |||||    | ||| |||| ||  |||||  ||
    624 CACCTGGCTGCCCGCCTACTACCAAGGCTCGGCTGCGGTCGC 665


          Quality:   11.1              Length:      32
            Ratio:  0.347               Gaps:       0
Percent Similarity: 65.625   Percent Identity: 65.625

16ssti4.Dat x Sphaerlh2.Dat April 7, 1990  20:22  ..        C

                .         .         .
     60 GTTATCGTGCAAACCCGCACGCGGGCGCCCCA 91
        | | ||||  || |||||  || ||| |  ||
    272 GGTCTCGTCGAAGCCCGCGTGCAGGCCCTACA 303


          Quality:   11.8              Length:      27
            Ratio:  0.437               Gaps:       0
Percent Similarity: 70.370   Percent Identity: 70.370

16ssti4.Dat x Sphaerlh2.Dat April 7, 1990  20:25  ..        D

                .         .
     99 CCAGTCGCTGGGGCGCCCGCGTGCGGG 73
        || ||| |   |  ||||||||||| ||
    270 CCGGTCTCGTCGAAGCCCGCGTGCAGG 296


    Gap Weight:  5.000      Average Match:  1.000
 Length Weight:  0.300   Average Mismatch: -0.900        E
```

Figure 9.4

Bestfit analyses of sequences SstI1 and 4 were made to the sequence of pLHIISB18, see opposite. The analyses were carried out in both the Forward and Reverse orientation. The Bestfit found between the sequences were:

A = Bestfit of Forward 16SstI1 sequence (60-100) to Forward pLHIISB18 sequence (664-704).

B = Bestfit of Reverse 16SstI1 sequence (243-202) to Forward pLHIISB18 sequence (624-665).

C = Bestfit of Forward 16SstI4 sequence (60-91) to Forward pLHIISB18 sequence (272-303).

D = Bestfit of Reverse 16SstI4 sequence (99-73) to Forward pLHIISB18 sequence (270-296).

E = The table of score values used for the Bestfit comparisons.

```
          Quality:   63.7              Length:    148
            Ratio:  0.439               Gaps:      2
Percent Similarity: 75.352   Percent Identity: 75.352

16ssti3.Dat x Sphaerlh2.Dat April 7, 1990  20:26  ..

          .         .         .         .         .
  4 CTGACATGAACCAGGGCAAGATCTGGACCGTCGTCAACCCGAGCGTTGGC 53
    ||||||||| | | ||||| ||||||  ||| || || |||| ||| |||
506 CTGACATGACCAACGGCAAAATCTGGCTCGTGGTGAAACCGACCGTCGGC 555
          .         .         .         .         .
 54 CTGCCGCTGCTGCTCGGCTCCGTGACCGTCATCGCCATTCTCGTT...CA 100
    | |||||| | ||| || | |  ||||||||||    |||||  ||
556 GTTCCGCTGTTCCTCAGCGCTG...CCGTCATCGCCTCCGTCGTTATCCA 602
          .         .         .         .
101 CCTCGCTGTGCTGTCGAACACCAAGTGGTTCCCGGCCTACTGGCAGGG 148
    |  |||||||| ||| |||||  ||| | || |||||||  || ||
603 CGCTGCTGTGCTGACGACCACCACCTGGCTGCCCGCCTACTACCAAGG 650
```

A

```
          Quality:   12.2              Length:     16
            Ratio:  0.763               Gaps:      0
Percent Similarity: 87.500   Percent Identity: 87.500

16ssti2.Dat x Sphaerlh2.Dat April 7, 1990  20:27  ..

              .
107 AGCTTCGGCTGCGGTC 122
    ||  ||||||||||||
648 AGGCTCGGCTGCGGTC 663
```

B

```
          Quality:   59.4              Length:    124
            Ratio:  0.479               Gaps:      0
Percent Similarity: 72.581   Percent Identity: 72.581

16ssti2.Dat x Sphaerlh2.Dat April 7, 1990  20:29  ..

          .         .         .         .         .
135 GGCTGACCGGCTTGACCGCAGCCGAAGCTGAAGAGCTTCACAAGTACGTT 86
    ||| || |||| |||||| |||||||| |||||  |||| ||| |  |
365 GGCCGAGCGGCCTGACCGTTGCCGAAGCCGAAGAAGTTCATAAGCAACTC 414
          .         .         .         .         .
 85 ATCGACGGCACCCGCGTGTTCTTCGTGATCGCGATCTTCGCGCACGTTCT 36
    |||  |||||||||||| |||   |  || ||| || |||||||| | ||
415 ATCCTCGGCACCCGCGTCTTCGGTGGCATGGCGCTCATCGCGCACTTCCT 464
              .         .
 35 CGCGTTCGCTTTCTCCCCCTGGCT 12
    |||   ||||    |||| |||||
465 CGCCGCCGCTGCGACCCCGTGGCT 488
```

C

```
     Gap Weight:  5.000      Average Match:  1.000
  Length Weight:  0.300   Average Mismatch: -0.900
```

D

Figure 9.5

Bestfit analyses of sequences SstI2 and 3 were made to the sequence of pLHIISB18, see opposite. The Bestfit found between the sequences were:

A = Bestfit of Forward 16SstI3 sequence (4-148) to Forward pLHIISB18 sequence (506-650).

B = Bestfit of Forward 16SstI2 sequence (107-122) to Forward pLHIISB18 sequence (648-663).

C = Bestfit of Reverse 16SstI2 sequence (135-12) to Forward pLHIISB18 sequence (365-488).

D = The table of score values used for the Bestfit comparisons.

mp19 recombinants were positive, suggesting that there were only two 800bp fragments and that they both hybridised positively.

This evidence lead to the re-examination of lambda clone 16 after digestion with SstI. The digest was loaded onto agarose gels of various percentages (0.4, 0.8, 1.0, 1.5, 2.0 and 2.5%) in an attempt to resolve any differences in size which may have existed between the two '800bp' fragments. Figure 9.6 shows the result of running this digest on a 1.5% agarose gel. Two bands of ≈800bp have just been resolved, though I would estimate the size difference can only be tens of bases. Ideally these fragments could have been run on an 8% acrylamide gel and the difference in size determined more accurately.

Southern blot analysis of the 1.5% agarose gel was attempted. Unfortunately the resolution of the blot was, as expected, insufficient to determine if one or both of the bands hybridised positively with pLHIISB18. Though, again evidence from *in situ* plaque hybridisation suggest that both fragments do hybridise positively. Indeed it may have been the difference in the degree of hybridisation of pLHIISB18 to two different inserts which resulted in the difference in signal strength found by plaque hybridisation. However, although the computer generated hybridisation, Bestfit, did find a difference in percentage identity of 3% between the 16SstI2 and 16SstI3 sequences and pLHIISB18, it seems unlikely that this alone could account for the relatively large difference in hybridisation signal. However it is also difficult to judge how closely the computer generated hybridisation mimics that of the *in vitro* reaction. It seems more likely that the difference in signal strength were due to different insert types and/or different insert orientations within the mp19 vector.

If we now carry on with the assumption that there are two, positively hybridising, ≈800bp fragments then it now has to be decided how the fragments are organised with respect to one another. Recall, that the sequence of 16SstI3 and the reverse complement of

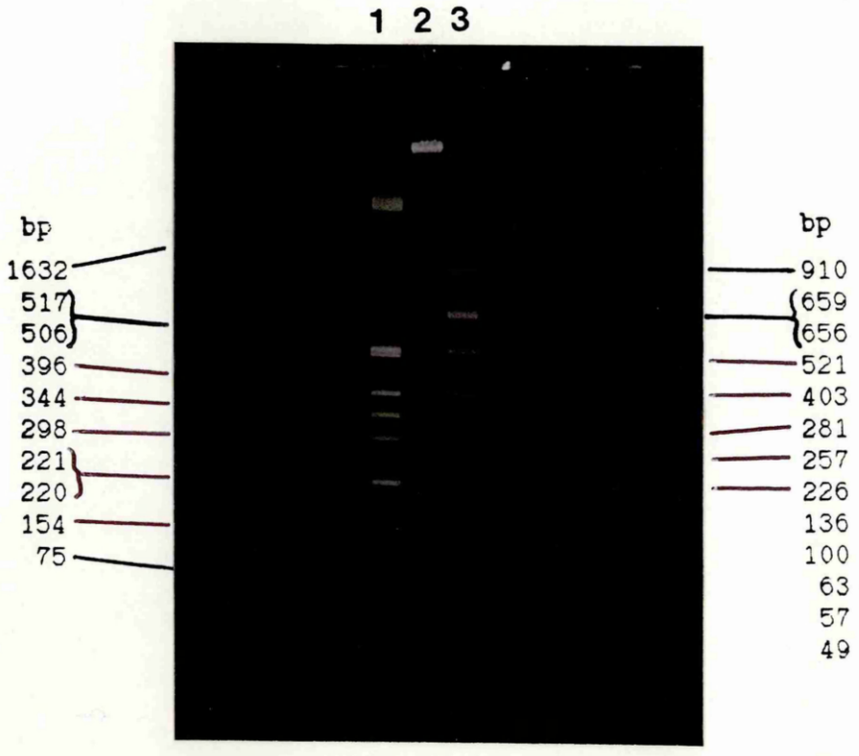


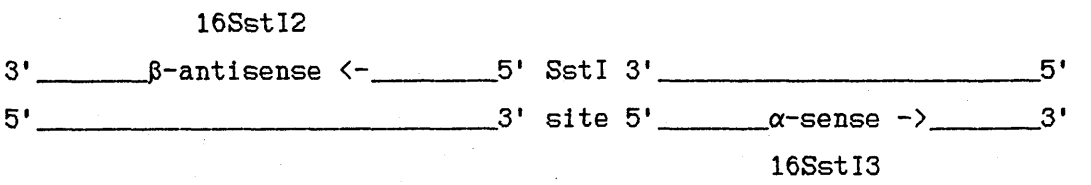
Figure 9.6

To determine if genomic clone 16 contained two 800bp SstI fragments, the DNA (Lane 2) was digested with SstI and run on a 1.5% agarose gel.

Two fragments of ≈ 800bp have just been resolved.

Lanes 1 and 3 contain pBR322 DNA digested with HinfI, and AluI.

the sequence of 16SstI2 showed alignment with pLHIISB18. This suggested that the "sense" strand of the alpha coding region and "antisense" strand of the beta coding region had been sequenced. In addition, because the sequences were "sense" and "antisense" it suggested that the sequences sat on either side of a central SstI site which had facilitated their subcloning, see below.



The arrows indicate the direction of sequencing. β and α are the encoded polypeptides.

As all other light harvesting genes examined to date have the alpha and beta genes encoded on the one strand, it seemed reasonable to reproduce this organisation with the 16SstI2 and 3 sequences. The 16SstI2 sequence was reversed and complemented to give the sequence presented in figure 9.7. The 16SstI3 sequence, and the reversed complement 16SstI2 sequence, were then joined together to give consensus sequence 16 presented in figure 9.8. This sequence was then translated in the six possible reading frames, one of which gave the deduced amino acid sequences of the presumptive alpha and beta light harvesting polypeptides also presented in figure 9.9.

The genes encoding these polypeptides are ordered beta-alpha, the same order as that found for other LHII genes in all other photosynthetic bacteria examined previously. The region encoding these two polypeptides is 354bp in length inclusive of the start and termination codons. A 12bp non-coding region lies between regions encoding the two polypeptides. Both polypeptides have the initiation codon ATG which codes for methionine. The alpha and beta polypeptides are terminated by TGA and TAA respectively.

16ssti2.Rev Length: 336 April 7, 1990 18:47 Check: 2191

```
   1  TCACGTGTCG TTCAGAAATG CGTTTCCATG TCAAGTGAAT TTGACATTAC
  51  AAAGCGGCAA TCTATAGTGA CACCTGCGCT GCGACGGCGA AGCTCAGTTT
 101  CGCACAAGTT CGGCCACCGC GCCACCGGTG ATCAACGTGC GGTTTAGCAC
 151  GAAAAAAAGC GTGGCGAAAA ATTTGAGAGG TTGACATGGC TGATGTAAAG
 201  GGGCTGACCG GCTTGACCGC AGCCGAAGCT GAAGAGCTTC ACAAGTACGT
 251  TATCGACGGC ACCCGCGTGT TCTTCGTGAT CGCGATCTTC GCGCACGTTC
 301  TCGCGTTCGC TTTCTCCCCC TGGCTGCACT AAGGAGCTCC TGACATGAAC
 351  CAGGGCAAGA TCTGGACCGT CGTCAACCCG AGCGTTGGCC TGCCGCTGCT
 401  GCTCGGCTCC GTGACCGTCA TCGCCATTCT CGTTCACCTC GCTGTGCTGT
 451  CGAACACCAA GTGGTTCCCG GCCTACTGGC AGGGTGGCCT GAAGAAGGCC
 501  GCTGCGATCG AGACCACGAT CGTCGGCTGA CGCCGCCTGA CGGCTGCCGC
 551  ATGTGACGGC ACACCGCCGA TGGTGTGCGG CGC
```

Figure 9.7

The sequence shown above is the reversed complement of sequence 16SstI2.

```
   1  TCACGTGTCG TTCAGAAATG CGTTTCCATG TCAAGTGAAT TTGACATTAC
  51  AAAGCGGCAA TCTATAGTGA CACCTGCGCT GCGACGGCGA AGCTCAGTTT
 101  CGCACAAGTT CGGCCACCGC GCCACCGGTG ATCAACGTGC GGTTTAGCAC
 151  GAAAAAAAGC GTGGCGAAAA ATTTGAGAGG TTGACATGGC TGATGTAAAG
 201  GGGCTGACCG GCTTGACCGC AGCCGAAGCT GAAGAGCTTC ACAAGTACGT
 251  TATCGACGGC ACCCGCGTGT TCTTCGTGAT CGCGATCTTC GCGCACGTTC
                                          SstI
 301  TCGCGTTCGC TTTCTCCCCC TGGCTGCACT AAGGAGCTCC TGACATGAAC
                                       2 <--+--> 3
 351  CAGGGCAAGA TCTGGACCGT CGTCAACCCG AGCGTTGGCC TGCCGCTGCT
 401  GCTCGGCTCC GTGACCGTCA TCGCCATTCT CGTTCACCTC GCTGTGCTGT
 451  CGAACACCAA GTGGTTCCCG GCCTACTGGC AGGGTGGCCT GAAGAAGGCC
 501  GCTGCGATCG AGACCACGAT CGTCGGCTGA CGCCGCCTGA CGGCTGCCGC
 551  ATGTGACGGC ACACCGCCGA TGGTGTGCGG CGC
```

Figure 9.8

Sequence 16SstI3 and the reverse complement of 16SstI2 (Fig.9.7) have been joined together. The resulting sequence, consensus 16SstI, is shown above. This contains an SstI site (boxed) which is believed to have facilitated the cloning of the 800bp fragments. Arrows 2 and 3 show the direction in which the templates were sequenced to give 16SstI2 and 3.

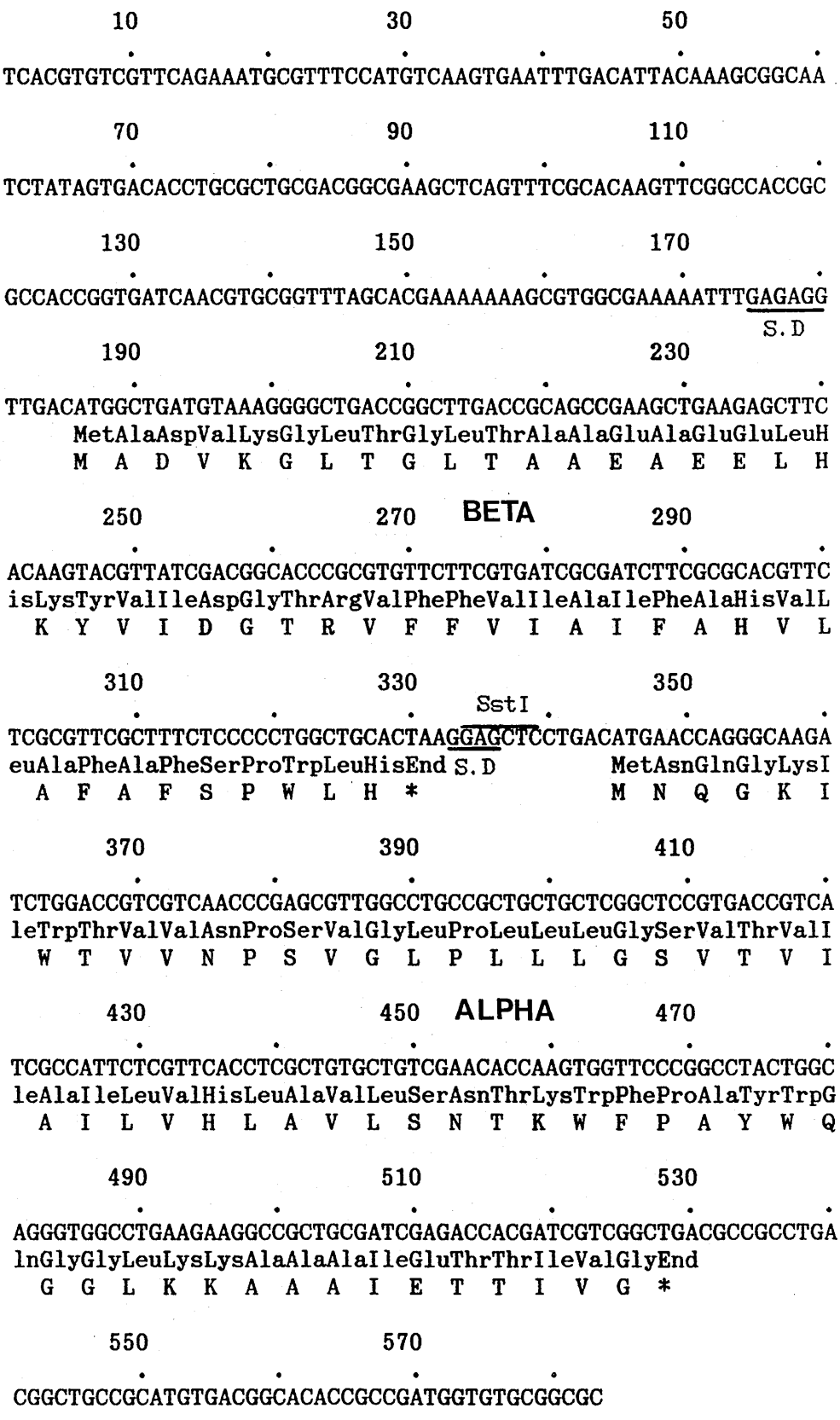


Figure 9.9

Translation of consensus sequence 16SstI generated the presumptive light harvesting alpha and beta polypeptides shown above. Possible Shine-Dalgarno sequences are marked S.D. The SstI site shown in the previous figure is also marked.

Figure 9.10

The presumptive 16SstI beta (positions 1-50) and alpha (positions 51-111) polypeptides have been aligned with native B800-850 and B800-820 beta and alpha polypeptides from *R. acidophila* strains 7050 and 7750.

The reference positions of alignment are the conserved histidine residues (*) at positions 19, 37 and 81.

The presumptive polypeptides are presented in bold typeface.

Positions where there is absolute residue conservation have been blocked.

At other positions, where the native polypeptides differ from the presumptive polypeptides the differences have been signified by underlining the relevant residue eg. A.

Where the type of amino acid residue at a particular position is unknown, this is signified by (.).

The differences between the presumptive and native polypeptides have been quantified and are presented in the difference score table shown in the text.

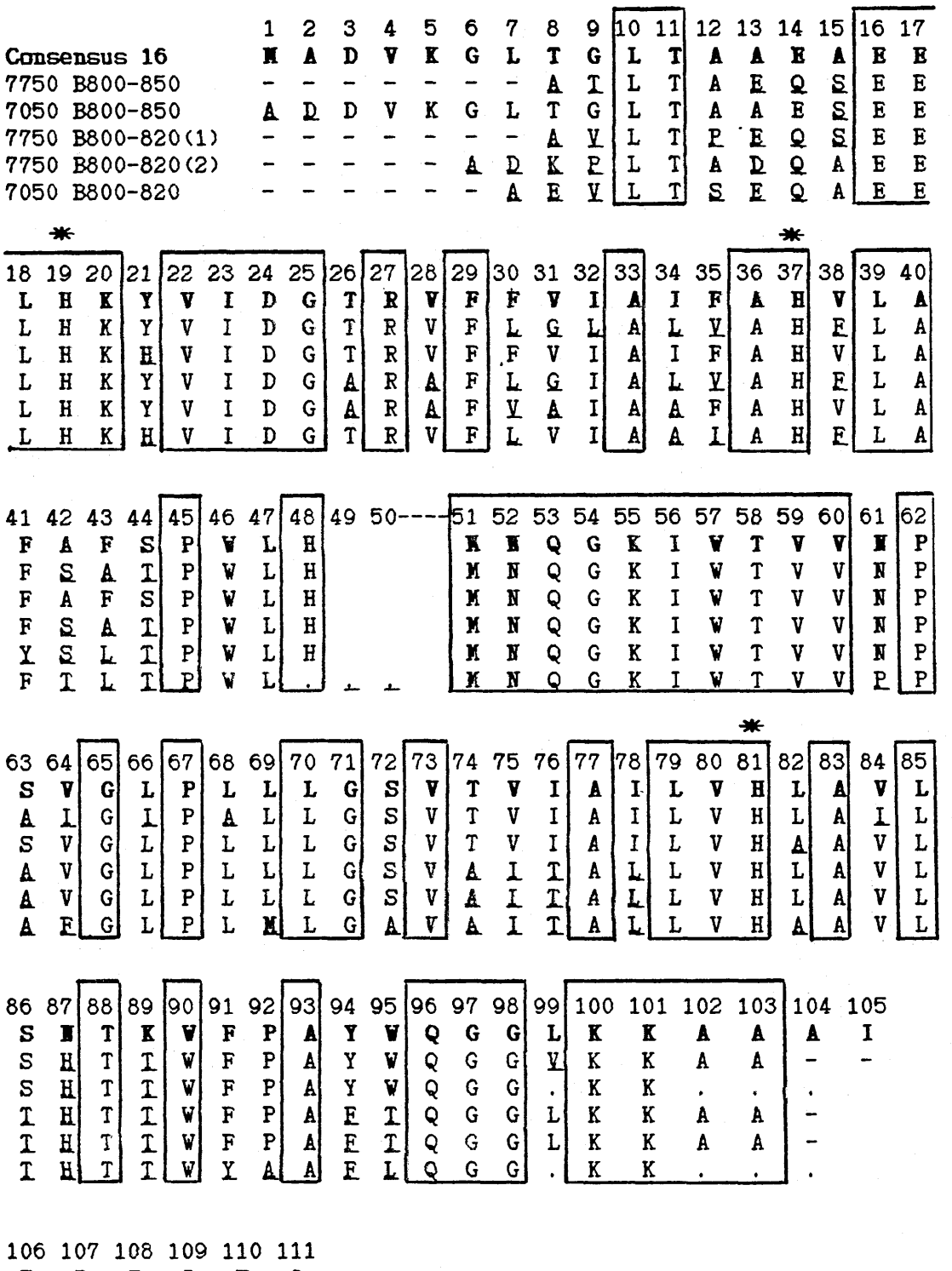

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Consensus 16 M A D V K G L T G L T A A E A E E
7750 B800-850 - - - - - - - A I L T A E Q S E E
7050 B800-850 A D D V K G L T G L T A A E S E E
7750 B800-820(1) - - - - - - - A V L T P E Q S E E
7750 B800-820(2) - - - - - A D K P L T A D Q A E E
7050 B800-820 - - - - - - A E V L T S E Q A E E
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
L H K Y V I D G T R V F F V I A I F A H V L A
L H K Y V I D G T R V F L G L A L V A H F L A
L H K H V I D G T R V F F V I A I F A H V L A
L H K Y V I D G A R A F L G I A L V A H F L A
L H K Y V I D G A R A F V A I A A F A H V L A
L H K H V I D G T R V F L V I A A I A H F L A
41 42 43 44 45 46 47 48 49 50---51 52 53 54 55 56 57 58 59 60 61 62
F A F S P W L H M N Q G K I W T V V N P
F S A T P W L H M N Q G K I W T V V N P
F A F S P W L H M N Q G K I W T V V N P
F S A T P W L H M N Q G K I W T V V N P
Y S L T P W L H M N Q G K I W T V V N P
F T L T P W L . . . M N Q G K I W T V V P P
63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85
S V G L P L L L G S V T V I A I L V H L A V L
A I G I P A L L G S V T V I A I L V H L A I L
S V G L P L L L G S V T V I A I L V H A A V L
A V G L P L L L G S V A I T A L L V H L A V L
A V G L P L L L G S V A I T A L L V H L A V L
A F G L P L M L G A V A I T A L L V H A A V L
86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
S N T K W F P A Y W Q G G L K K A A A I
S H T T W F P A Y W Q G G V K K A A - -
S H T T W F P A Y W Q G G . K K . . .
T H T T W F P A F T Q G G L K K A A -
T H T T W F P A F T Q G G L K K A A -
T H T T W Y A A F L Q G G . K K . . .
106 107 108 109 110 111
E T T I V G

Figure 9.10

Upstream of both coding regions lies a possible ribosome binding site (Shine-Delgarno sequence). The sequence GAGAGG is found 5bp upstream of the beta polypeptide and GGAG is found 8bp upstream of the alpha polypeptide. The GAG part of the GGAG sequence is half of the SstI site (GAGCTC) which joins the 16SstI2 and 3 sequences. The presumptive Shine-Delgarno sequences are discussed in greater depth in chapter 11.

The deduced presumptive amino acid sequences were then compared to those of the native B800-850 and B800-820 polypeptides from *R. acidophila* strains 7050 and 7750. Note that the amino acid sequences of four alpha and five beta polypeptides were available for comparison:

Strain 7050 B800-850 alpha and beta polypeptides
Strain 7050 B800-820 alpha and beta polypeptides

Strain 7750 B800-850 alpha and beta polypeptides
Strain 7750 B800-820 alpha and two beta polypeptides.

For convenience, I have joined together the appropriate alpha and beta polypeptides eg. B800-850 beta with B800-850 alpha. However, as the B800-820 complex from strain 7750 only has one alpha sequence and two beta sequences I have joined the one alpha sequence to both of the beta sequences, ie.

B800-820 beta [1] with B800-820 alpha
B800-820 beta [2] with B800-820 alpha

This joining has been done for convenience, it may not represent the *in vivo* state.

The comparison (figure 9.10) compares the presumptive polypeptide amino acid sequences to those determined by amino acid sequencing of the native *R. acidophila* LHII polypeptides of strains 7050 and 7750. Differences between the presumptive polypeptides and the native

polypeptides are signified by underlining the relevant position on the native polypeptide. These differences have been scored for and the resulting scores are presented in table 9.1.

R. acidophila strain	Antenna Complex Type	Amino Acid Differences		
		Polypeptide Type		
		BETA	ALPHA	TOTAL
7750	B800-850	14	8	22
7050	B800-850	4	3	7
7750	B800-820 [1]	16	10	26
7750	B800-820 [2]	15	10	25
7050	B800-820	16	17	33

Table 9.1

The scores shown represent how the native polypeptides differ with respect to the presumptive polypeptides. The scoring system used is explained in the following examples which refer to the comparison in figure 9.10:

1) At position 13 of the presumptive polypeptide there is an alanine residue (A). Alanine is also found at this position in the B800-850 sequence from strain 7050. In all the other beta polypeptides glutamine (Q) is found at this position. The single letter code for glutamine is underlined, Q , and this underlining signifies a difference scoring of 1.

(2) At position 5 there is a lysine residue (K) in the presumptive polypeptide sequence. Lysine is also found at this position in the B800-850 sequence from strain 7050. All other sequences lack a residue at this position. This is signified by, -, and signifies a difference score of 1.

(3) With the exception of the B800-850 beta polypeptide from strain 7050, both the presumptive and the native polypeptides have histidine (H) at position 48. The B800-850 polypeptide from strain 7050 has a residue of unknown type (.) at this position. As the

unknown residue may or may not be histidine It has been given a difference score of 0.

(4) Positions 49 and 50 of the B800-820 beta polypeptide from strain 7050 have residues of an unknown type. However, as there are no residues at these positions in the presumptive polypeptide the unknown residues have been underlined, _, and given a difference score of 1.

The scoring system for differences continues to position 104, the last residue determined by amino acid sequencing of the native polypeptides. Differences which occur after this position have not been scored for as it is believed that residues found in the presumptive polypeptide beyond this position are absent in the native polypeptide, probably as a result of C-terminal processing. This is explained more fully in chapter 11.

The number of differences between the native polypeptides and the presumptive polypeptides was determined by summing the difference scores. These differences are presented for the alpha and beta polypeptides individually and as a total in table 9.1.

Discussion

Two main lines of evidence suggest that a nucleotide sequence encoding light harvesting alpha and beta polypeptides has been isolated. Firstly, recombinant mp19 plaques hybridise positively when probed with pLHIISB18. These recombinants subsequently yielded nucleotide sequences which showed over 70% identity with nucleotide sequences encoding the B800-850 alpha and beta polypeptides of *Rb. sphaereoides* strain 2.4.1. The genes encoding these polypeptides are ordered beta-alpha, and have upstream of their coding regions sequences which may act as ribosome binding sites. These features have also been found in all other photosynthetic bacteria examined so far.

Secondly, translation of the recombinant sequences generated amino acid sequences which have considerable similarity to those of the LHII polypeptides of *R. acidophila* strains 7050 and 7750. This can be seen from figure 9.10 where there are large blocks of amino acid identity between the various polypeptide types. Regions where the sequence identity breaks down is often not the result of the presumptive amino acid sequence. At position 61, for example, the identity block between positions 60 and 62 would have continued were it not for the proline (P) present in the B800-820 alpha polypeptide of strain 7050. Asparagine (N) was present at this position in all the other sequences.

At some positions in the native polypeptides a number of amino acids are found, for example, phenylalanine (F), leucine (L) and valine (V) are found at position 30. The presumptive polypeptide has phenylalanine at this position. Therefore, although there may be differences at a particular position in the various sequences, the presumptive polypeptide usually has a residue which is found in that position in one or more of the native polypeptides. This gives the presumptive polypeptides the "feel" of alpha and beta light harvesting polypeptides without them being identical to any particular native polypeptide. However, such a situation is not always the case.

At positions 87 and 89 asparagine (N) and lysine (K) are found respectively in the presumptive alpha polypeptide. In all other alpha polypeptides histidine (H) and threonine (T) are found at these positions. These differences however do not necessarily exclude the presumptive sequence from being that of a light harvesting alpha polypeptide. This hypothesis is based on the examination of positions 91 and 92. The B800-820 alpha polypeptide of strain 7050 has tyrosine (Y) and alanine (A) at these positions, whereas in all the other alpha polypeptides phenylalanine (F) and proline (P) are found. Surely if these unilateral differences do not exclude the B800-820 sequence from being that of a light harvesting polypeptide then they cannot be used to exclude the

presumptive alpha polypeptide from that same role. This argument appears even stronger when the hydropathy indices are examined, see table below.

	Position 87	Position 89
Presumptive residue	N= 31, (-3.5)	K= 17.6, (-3.9)
Native residues	H= 14.3, (-3.2)	T= 3.8, (0.7)
	Position 91	**Position 92**
B800-820 residue	Y= 4.7, (-1.3)	A= -4.2, (1.8)
Other native and presumptive residues	F= -14.2, (2.8)	P= 13.9, (-1.6)

Table 9.2 Comparison of the hydropathy indices of residues at particular positions within the native and presumptive polypeptide sequences.
For each residue there are two numbers eg. N= 31, (-3.5). The first number is the hydropathy index according to the scale of Von Heijne [Von Heijne, 1981], the second is the hydropathy index according to the scale of Kyte and Doolittle [Kyte and Doolittle, 1982]. Note the two scales have been determined with opposite signs. The hydropathy values are based on the change in free energy associated with the transfer of an amino acid from a hydrophobic to a hydrophilic environment and have been corrected theoretically and experimentally.

It can be seen from table 9.2 that changing histidine to asparagine at position 87 and threonine to lysine at position 89 results in changes in the hydrophibicity of the residues at these position by 16.7, (-0.2) and 13.8, (4.6) units respectively. Changing phenylalanine to tyrosine at position 91 and proline to alanine at position 92 involves equally large changes in the hydrophobicity of

the residues at these positions. This suggests that changing one residue for another of considerably different hydrophobicity does not necessarily exclude the resulting polypeptide from being a functional. However, what is unknown is how critical these changes would be at a particular position. It may well be that changing phenylalanine to tyrosine at position 91 has very little structural or functional effect, whereas it is possible that changing histidine to asaparagine at position 87 results in a completely non-functional polypeptide. Likewise it may be that the hydrophobicity of the residue is irrelevant and that the important factors in functional terms of the polypeptide are the sizes, charge, helix forming ability, and interactions the residue makes with other residues and/or pigments in the same or other polypeptides. However, in order to determine if this were the case further experimental work would need to be carried out.

The result of the scoring table (9.1) suggest that the presumptive polypeptide is most like the B800-850 polypeptide of *R. acidophila* strain 7050. There are 4 differences between the beta sequences and 3 differences between the alpha sequences of the presumptive and B800-850 polypeptides of this strain. This gives a total of 7 amino acid differences between the two alpha-beta sequences.

The differences between the presumptive sequence and the other native polypeptides ranges from 14-16 and 8-17 changes for the beta and alpha polypeptides respectively. This gives difference totals ranging from 22-33 amino acid changes.

It would therefore appear from the data presented above that a nucleotide sequence encoding an alpha and beta light harvesting polypeptide has been successfully isolated and sequenced. Of the polypeptides with which it has been compared it is most like the B800-850 alpha and beta polypeptides of *R. acidophila* strain 7050.

Sequence data obtained from the 6SstI series of constructs

Digestion of lambda clone 6 with SstI appeared to generate two SstI fragments of 0.7 and 2.2kb which hybridised positively with pLHIISB18. For sequencing, the clone was digested with SstI and the fragments shotgun cloned into the SstI site of mp19. *In situ* hybridisation with radiolabelled pLHIISB18 had permitted the isolation of 15 positively hybridising plaques. Templates were prepared from each of these, subsequently revealing five different nucleotide sequences.

Bestfit analysis of these sequences, rather disappointingly generated only a single significant alignment with the pLHIISB18 sequence. The sequence which gave this alignment, consensus 6SstI1 is presented as a gel autoradiograph in figure 9.11 and as the written sequence in figure 9.12. The alignment (presented in figure 9.13) which was made in the Forward orientation, extends from nucleotide 11 to 182 of the 6SstI1 sequence and shows 70% identity with the region from nucleotide 511 to 670 of the pLHIISB18 sequence. The region from nucleotide 501-675 of the pLHIISB18 sequences encodes the B800-850 alpha polypeptide of *Rb. sphaeroides*.

The 6SstI1 sequence (renamed as consensus 6) was translated in six reading frames, one of which revealed the deduced alpha polypeptide amino acid sequence presented in figure 9.14.

The region encoding the presumptive polypeptide is 182 nucletides long including the start and termination codons. There is a 10bp region upstream of the start codon, which may be non-coding. The initiation codon is ATG, the termination codon TGA.

Comparison of the presumptive alpha polypeptide sequence with the LHII polypeptide sequences from *R. acidophila* strains 7050 and 7750 is shown in figure 9.15. A difference score table of this comparison is shown in table 9.3.

Figure 9.11

Shown opposite are three gel autoradiographs of a template which gave consensus sequence 6SstI1.

The gels have been run for ≃2, ≃4 and ≃7 hours under the conditions described previously.

The nucleotide sequence deduced from these autoradiographs are shown overleaf.

X and Y are reference points in the sequence.

The junction between the subcloned SstI fragment and the mp19 polylinker is shown by the recreation of the SstI site GAGCTC.

```
   1  AGCTCCGGTT ATGAACCAAG GCAAAATCTG GACCGTCGTC AATCCGGCCG

  51  TCGGCCTTCC GCTGCTGCTC GGCTCCGTGG CCATCACCGC GCTGCTGGTG

 101  CACCTCGCCG TTCTCACGCA CACCACCTGG TTCCCCGCGT TCATGCAGGG

 151  CGGCCTGAAG AAGGCGGCTG CGATCGAGCA CGTCGTCGGC TGATGGCGTA

 201  AAGCCGCTCC GGAGCGGTGA GGCAGCGCCC GGCTTGCGAC GCGGCAGGCG

 251  TCGTTTGGTC CGCCGCCGCG
```

Figure 9.12

Sequence 6Sst1 deduced from the autoradiographs shown in figure 9.11.

```
          Gap Weight:  5.000      Average Match:  1.000
       Length Weight:  0.300   Average Mismatch: -0.900

             Quality:   60.2             Length:    172
               Ratio:  0.376              Gaps:      1
  Percent Similarity: 70.000   Percent Identity: 70.000

  6ssti1.Dat x Sphaerlh2.Dat April 7, 1990  16:03  ..

                .         .         .         .         .
   11 ATGAACCAAGGCAAAATCTGGACCGTCGTCAATCCGGCCGTCGGCCTTCC 60
      |||| | | ||||||||||||  ||| || || ||| |||||||| ||||
  511 ATGACCAACGGCAAAATCTGGCTCGTGGTGAAACCGACCGTCGGCGTTCC 560
                .         .         .         .         .
   61 GCTGCTGCTCGGCTCCGTGGCCATCACCGCGCTGCTGGTGCACCTCGCCG 110
      |||| | ||| || | |  | |||| || |  |  |  | |||   || |
  561 GCTGTTCCTCAGCGCTGCCGTCATCGCCTCCGTCGTTATCCACGCTGCTG 610
                .         .         .         .         .
  111 TTCTCACGCACACCACCTGGTTCCCCGCGTTCATGCAGGGCGGCCTGAAG 160
      | || |||  |||||||||| | |||| | |   || |||
  611 TGCTGACGACCACCACCTGGCTGCCCGCCTACTACCAAGGC..... 651
                .         .
  161 AAGGCGGCTGCGATCGAGCACG 182
          |||||||| ||| |  ||
  652 ...TCGGCTGCGGTCGCGGCCG 670
```

Figure 9.13

Shows the result of Bestfit analysis between sequence 6SstI1 (11-182) and the nucleotide sequence of pLHIISB18 (511-670). The analysis was carried out with both sequences in the Forward orientation.

```
        10                  30                  50
         .         .         .         .         .         .
AGCTCCGGTTATGAACCAAGGCAAAATCTGGACCGTCGTCAATCCGGCCGTCGGCCTTCC
          MetAsnGlnGlyLysIleTrpThrValValAsnProAlaValGlyLeuPr
           M  N  Q  G  K  I  W  T  V  V  N  P  A  V  G  L  P

        70                  90                  110
         .         .         .         .         .         .
GCTGCTGCTCGGCTCCGTGGCCATCACCGCGCTGCTGGTGCACCTCGCCGTTCTCACGCA
oLeuLeuLeuGlySerValAlaIleThrAlaLeuLeuValHisLeuAlaValLeuThrHi
 L  L  L  G  S  V  A  I  T  A  L  L  V  H  L  A  V  L  T  H

       130                 150                 170
         .         .         .         .         .         .
CACCACCTGGTTCCCCGCGTTCATGCAGGGCGGCCTGAAGAAGGCGGCTGCGATCGAGCA
sThrThrTrpPheProAlaPheMetGlnGlyGlyLeuLysLysAlaAlaAlaIleGluHi
 T  T  W  F  P  A  F  M  Q  G  G  L  K  K  A  A  A  I  E  H

       190                 210                 230
         .         .         .         .         .         .
CGTCGTCGGCTGATGGCGTAAAGCCGCTCCGGAGCGGTGAGGCAGCGCCCGGCTTGCGAC
sValValGlyEnd
 V  V  G  *

       250                 270
         .         .         .
GCGGCAGGCGTCGTTTGGTCCGCCGCCGCG
```

Figure 9.14

Translation of the 6SstI1 nucleotide sequence revealed the deduced amino acid sequence shown above. This appears to encode an alpha type light harvesting polypeptide.

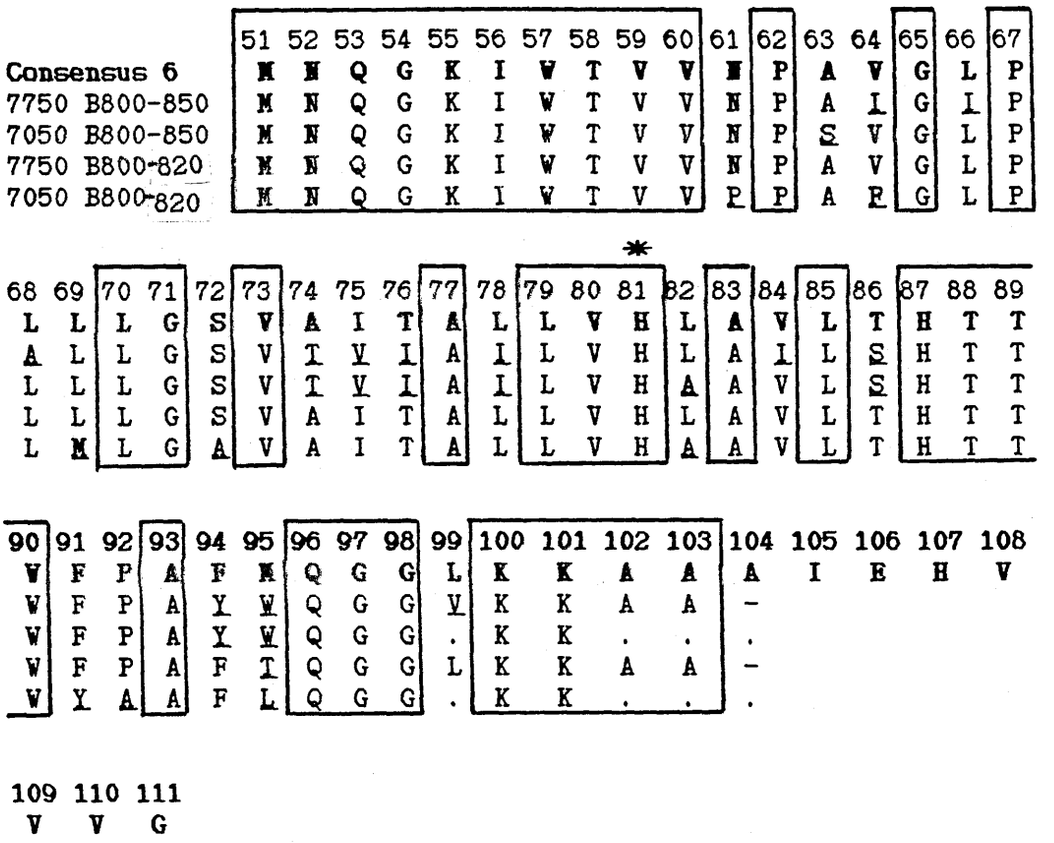


Figure 9.15

The presumptive 6 alpha polypeptide (positions 51-111) have been aligned with native B800-850 and B800-820 beta polypeptides from *R. acidophila* strains 7050 and 7750.

The reference position of alignment is the conserved histidine residue (*) at position 81.
The presumptive polypeptide is presented in bold typeface.
Positions where there is absolute residue conservation have been blocked.
At other positions, where the native polypeptides differ from the presumptive polypeptides the differences have been signified by underlining the relevant residue eg. A.
Where the type of amino acid residue at a particular position is unknown, this is signified by (.).

R. acidophila strain	Antenna complex type	Amino Acid Differences
7750	B800-850	11
7050	B800-850	9
7750	B800-820	2
7050	B800-820	8

Table 9.3

The difference score shows how many differences there are between the native polypeptides relative to the presumptive polypeptide.

The difference scores suggest that the presumptive alpha polypeptide is most like the B800-820 complex from *R. acidophila* strain 7750. There are two amino acid differences between these polypeptides, the other polypeptides having between 8 and 11 differences. The two differences occur at positions 95 and 104. At position 95, threonine (T) and methionine are found in the native and presumptive polypeptides respectively. At position 104 alanine (A) is found in the presumptive polypeptide, but there is no residue at this position in the native polypeptide. It is thought that the presumptive polypeptide may be postranslationally modified. If this is the case the alanine at position 104 may be absent from the functional polypeptide. Details of these possible modifications will be presented in the next chapter.

Before moving on to the next set of sequence data, I would like to indulge in some speculation regarding the apparent absence of any upstream candidates for Shine-Delgarno sequences and the lack of any presumptive beta polypeptide. The first five nucleotides in the 6SstI1 nucleotide sequence are AGCTC. As this fragment was generated using SstI, and as its recognition sequence is GAGCTC, it is likely that the base upstream of AGCTC is G. The GAG of the GAGCTC sequence could then form part of a Shine-Delgarno sequence such as GGAG. If this were the case the Shine-Delgarno sequence

would be 8bp upstream of the start codon of the presumptive polypeptide. A situation identical to this was found in the 16SstI consensus sequences described earlier. The scenario can now be extended to account for the absence of the beta polypeptide. I stated earlier that in genomic clone 6 there were two ≃700bp SstI fragments, at least one of which hybridised positively when probed with pLHIISB18. In addition there was also a 2.2kb SstI fragment which also hybridised positively when probed in this way. This suggests that the 2.2kb fragment had been cloned in both orientations, and the 5' ends of both complementary strands sequenced. However, the light harvesting genes are buried within this fragment and have not been reached by sequencing. It also suggests that although one 0.7kb fragment has also been subcloned and sequenced in both orientations, the other one has only been subcloned in one orientation. Unfortunately, the orientation which I believe is missing probably would have revealed the nucleotide sequence encoding the presumptive beta polypeptide and the other part of the possible Shine-Delgarno sequence!

If however an adjacent upstream beta polypeptide does not exist, then that in itself would be extremely interesting. This would be the first example (that I know of) where a light harvesting polypeptide gene has been found without its usual complementary neighbour. Even if the beta polypetide gene were found, but buried in the 2.5kb fragment, the distance between the two genes would be at least 250bp, ie. the distance already sequenced into this fragment. This would certainly be the maximum distance found separating two of such genes. I think more likely however, is that the beta encoding gene has not yet been subcloned, but *in vivo* is contiguous with the alpha polypeptide encoding gene. The 2.5kb fragment probably encodes a separate alpha-beta polypeptide gene pair. Such a situation has already been found in *R. palustris* where light harvesting genes have been found clustered [Tadros and Waterkamp, 1989] and I will now present data which suggests that a comparable gene organisation also occurs in *R. acidophila*.

Sequence data from lambda clone 9

In order to generate fragments for sequencing, genomic clone 9 was digested with HpaII, RsaI, HincII, and AluI. The HpaII fragments were cloned into the AccI site, all other fragments were cloned into the SmaI site of mp19. Recombinants which hybridised positively to pLHIISB18 were isolated by *in situ* plaque hybridisation. For each of the positively hybridising plaques, single stranded templates were prepared then sequenced. The sequences obtained were aligned with the nucleotide sequence of pLHIISB18 using Bestfit. This analysis suggested that at least six presumptive light harvesting polypeptides were encoded within this genomic clone. In addition the genes were arranged in three gene pairs (9A, 9B and 9C) each encoding an alpha and beta polypeptide.

This hypothesis was tested by using Gelassemble, a programme which compared each of the template sequences against all of the other template sequence. It then clustered sequences which shared regions of identical sequence. This had two main effects, a) it ensured that the nucleotide sequences which encoded the presumptive polypeptide sequences were different to one another, and b) it permitted regions which lacked similarity to the pLHIISB18 to be added to the the presumptive polypeptide sequences on the basis of identity between sequenced overlaps.

In this section the nomenclature of the Gelassemble programme has been used to describe the nucleotide sequences. Each individual template sequence is called a contig, and joining contigs together is called melding. The result of melding contigs is the generation of a single piece of sequence, the so called consensus sequence. To avoid confusion, I have also split the sequence data into three parts, each part pertaining to a single alpha-beta polpeptide encoding consensus sequence.

Consensus sequence 9A

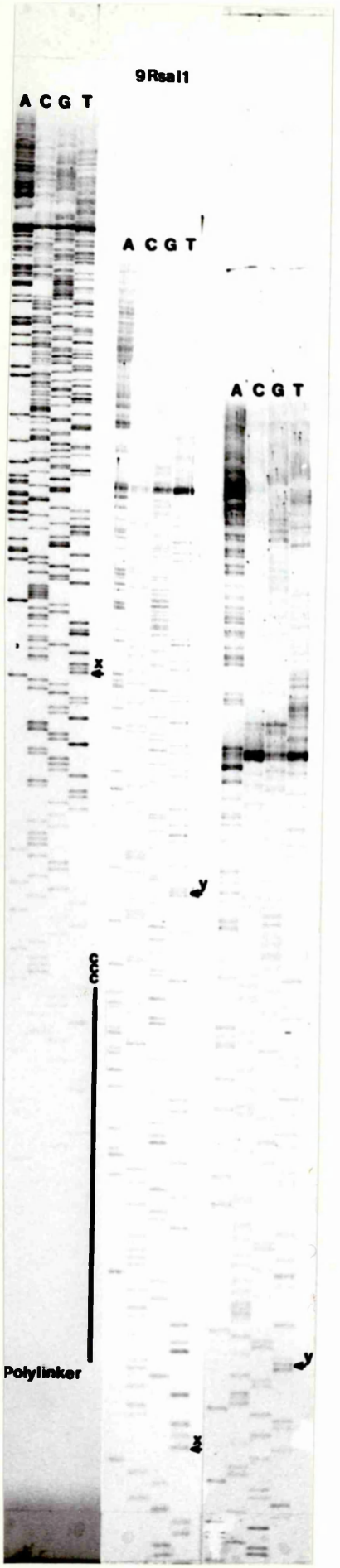
Consensus sequence 9A was derived from two mp19 recombinants, 9RsaI1 and 9HpaII4. The sequencing gel autoradiographs of 9RsaI1 are presented in figure 9.16, and the nucleotide sequences derived from both recombinants presented in figure 9.17.

The Gelassemble programme found complete identity between 9HpaII4 and the region between nucleotides 138 and 227 of 9RsaI1, this alignment is shown schematically in figure 9.17c. From this figure it can be seen that 9HpaII4 did not give any additional sequence information 3' or 5' to the ends of 9RsaI1. However, I have included the 9HpaII4 sequence as the direction of the alignment arrows indicate that these two contigs have been obtained by sequencing the opposite strands of the same region. It is thus a useful check as to the accuracy of the sequence data.

The result of the Bestfit analysis between the 9RsaI1 sequence and the pLHIISB18 nucleotide sequence is presented in figure 9.19. This shows that the region from nucleotide 3 to 241 of the RsaI1 sequence has 68.9% identity with the region from nucleotide 409 to 650 of pLHIISB18. This region of pLHIISB18 encodes part of the alpha polypeptide, the intergenic non coding region, and part of the beta polypeptide of the B800-850 antenna complex of *Rb. sphaeroides*.

Consensus sequence 9A, which is effectively the same as contig 9RsaI, was translated in the six possible reading frames, one of which generated the amino acid sequence presented in figure 9.20.

The 9A consensus sequence translates into a presumptive alpha, and about 60% of a presumptive beta, light harvesting polypeptide. The alpha polypeptide is encoded by 186 nucleotides, initiated by the codon ATG, and terminated by TGA. The beta polypeptide is terminated by TAA. Between the two coding regions is an intergenic region of 12bp which contains a possible Shine-Delgarno sequence GGAG, 8bp upstream of the alpha polypeptide initiation codon. This



≃2hrs ≃4hrs ≃7hrs

Figure 9.16

Shown opposite are three gel autoradiographs of a template which gave contig sequence 9RsaI1. This contig encodes all of a presumptive beta and most of an alpha light harvesting polypeptide.

The gels have been run for ≃2, ≃4 and ≃7 hours under the conditions described previously.

The nucleotide sequence taken from these autoradiographs is shown overleaf.

X and Y are reference points in the sequence.

The junction between the subcloned RsaI1 fragment and the mp19 polylinker is shown by CCC.

A

```
  1  CCCACGTGAT CGACGGCACC CGCGTGTTCC TGGGCCTGGC CCTGGTTGCG

 51  CATTTCCTCG CCTTCTCCGC GACCCCCTGG CTGCACTAAG GAGTTCTGAT
                                              HpaII
101  CATGAACCAA GGCAAGATCT GGACCGTCGT TGACCCGGCT TTCGGCATCC

151  CCGCTCTGCT CGGCTCCGTC ACCGTTATCG CGCTGCTCGT GCACCTCGCG
                           HpaII
201  ATCCTGTCGA ACACCACCTG GTTCCCGGCC TACTGGCAGG GTGGTGTGAA

251  GAAGGCTGCA GCCATCGAGA CCACCGTTCT CGGCTGAGTG CTTCAGCGAC

301  ATCTCGATGT
```

B

```
  1  GAACCAGGTG GTGTTCGACA GGATCGCGAG GTGCACGAGC AGCGCGATAA

 51  CGGTGACGGA GCCGAGCAGA GCGGGGATGC CGAAAGCCG
```

C

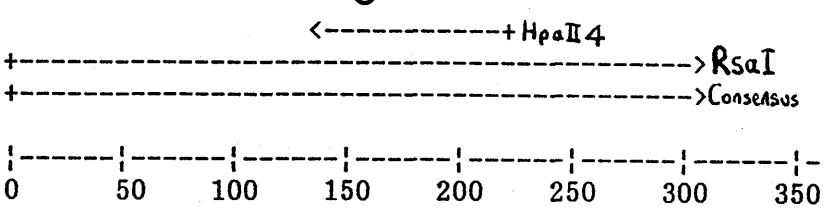


Figure 9.17

Contig sequences 9RsaI1 and 9HpaII4 are shown above as A and B respectively. The sequences overlap as shown in C. The HpaII sites within 9RsaI1 are shown boxed with the direction in which the HpaII4 sequence was determined being arrowed.

```
        Gap Weight:  5.000      Average Match:  1.000
     Length Weight:  0.300   Average Mismatch: -0.900

           Quality:   73.0             Length:    246
             Ratio:  0.305              Gaps:      4
Percent Similarity: 68.936   Percent Identity: 68.936

9rsail.Dat x Sphaerlh2.Dat April 7, 1990  15:41  ..

                .         .         .         .         .
   3 CACGTGATCGACGGCACCCGCGTGTTCCTGGGCCTGGCCCTGGTTGCGCA 52
     ||  | |||  |||||||||||| |||   ||| |||| ||  | |||||
 409 CAACTCATCCTCGGCACCCGCGTCTTCGGTGGCATGGCGCTCATCGCGCA 458
                .         .         .         .         .
  53 TTTCCTCGCCTTCTCCGCGACCCCCTGGCTGCACTAA....GGAGTTCTG 98
      |||||||||  | | ||||||||| |||||   || |     ||   |||
 459 CTTCCTCGCCGCCGCTGCGACCCCGTGGCTCGGCTGATAGGAGAAGACTG 508
                .         .         .         .         .
  99 ATCATGAACCAAGGCAAGATCTGGACCGTCGTTGACCCGGCTTTCGGCAT 148
     | ||||| | | ||||| |||||| ||| ||  | ||| |  |||||
 509 A.CATGACCAACGGCAAAATCTGGCTCGTGGTGAAACCGACCGTCGGCGT 557
                .         .         .         .         .
 149 CCCCGCTCTGCTCGGCTCCGTCACCGTTATCGCGCTGCTCG...TGCACC 195
      |   | ||| ||  |  ||   |||| |||||     |||   | |||
 558 TC...CGCTGTTCCTCAGCGCTGCCGTCATCGCCTCCGTCGTTATCCACG 604
                .         .         .         .
 196 TCGCGATCCTGTCGAACACCACCTGGTTCCCGGCCTACTGGCAGGG 241
       ||  | ||| ||| ||||||||||  | || |||||||  || ||
 605 CTGCTGTGCTGACGACCACCACCTGGCTGCCCGCCTACTACCAAGG 650
```

Figure 9.19

Bestfit analysis was carried out between sequences 9RsaI1 (3-241) and pLHIISB18 (409-650). The comparison was made in the Forward direction of both sequences. The result is shown above.

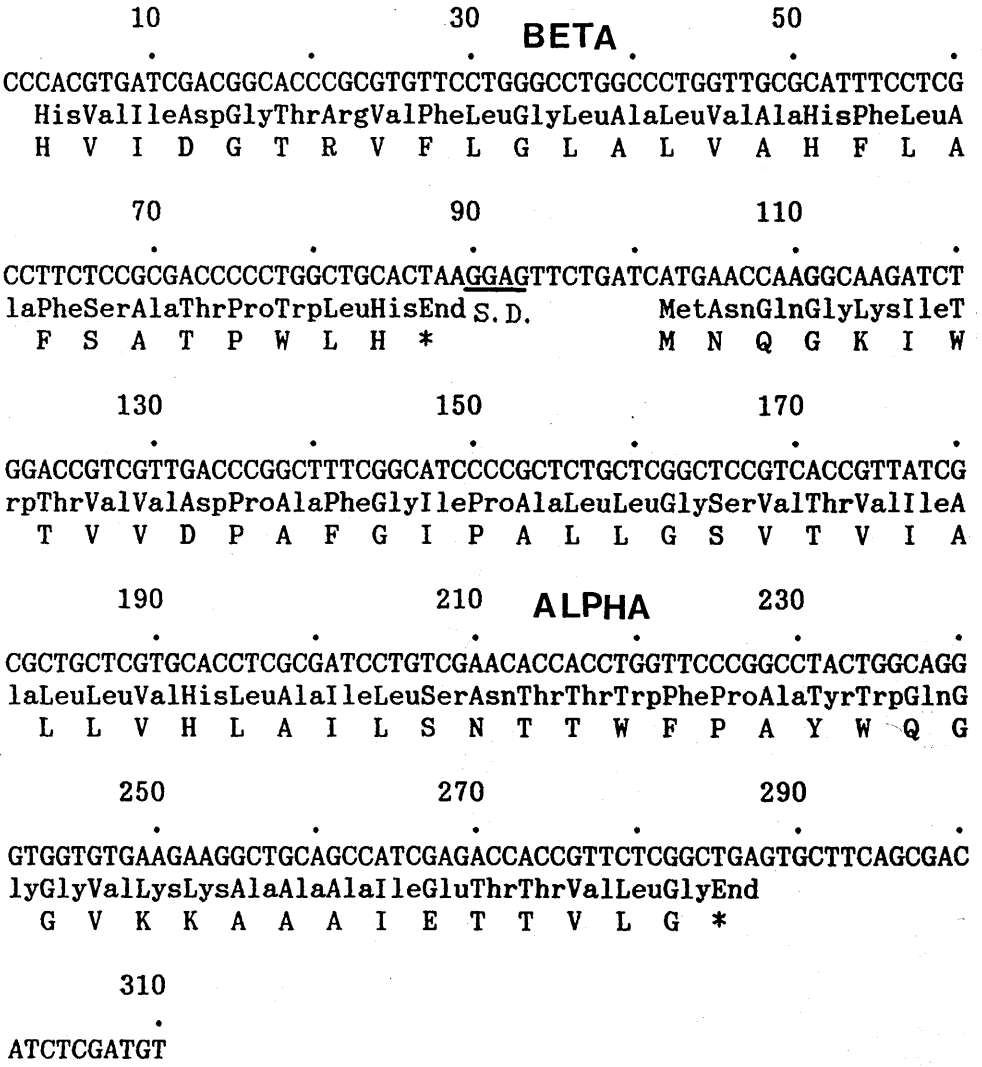


Figure 9.20

Consensus sequence 9A was translated in the six possible reading frames, one of which is shown above. The deduced amino acid sequence encodes a presumptive alpha and part of a beta light harvesting polypeptide. A possible Shine-Delgarno sequence is marked S.D.

Figure 9.21

The presumptive 9A beta (positions 21-50) and alpha (positions 51-111) polypeptides have been aligned with native B800-850 and B800-820 beta and alpha polypeptides from *R. acidophila* strains 7050 and 7750.

The reference positions of alignment are the conserved histidine residues (*) at positions 37 and 81.

The presumptive polypeptides are presented in bold typeface.

Positions where there is absolute residue conservation have been blocked.

At other positions, where the native polypeptides differ from the presumptive polypeptides the differences have been signified by underlining the relevant residue eg. A.

Where the type of amino acid residue at a particular position is unknown, this is signified by (.).

The differences between the presumptive and native polypeptides have been quantified and are presented in the difference score table shown in the text.

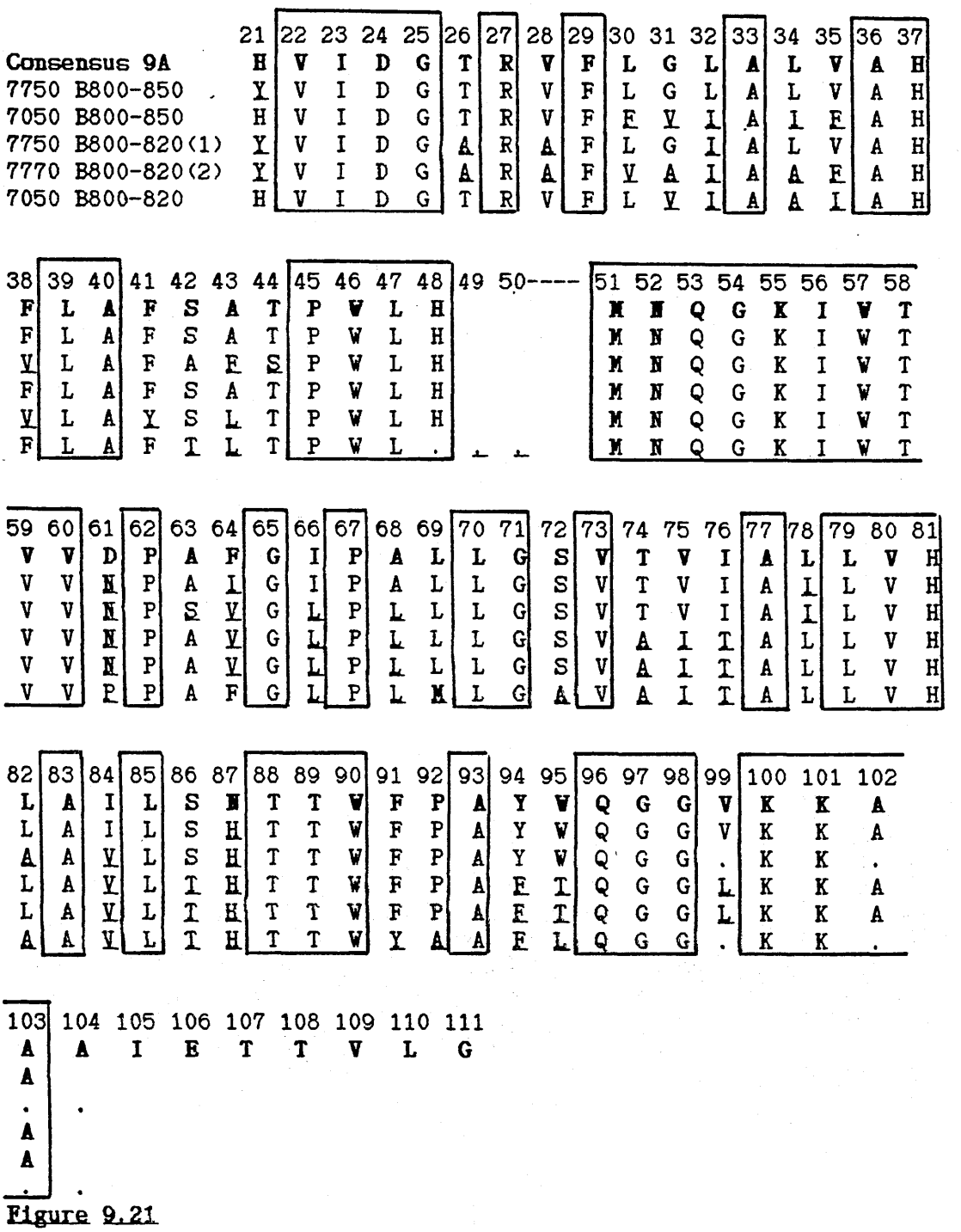

21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
Consensus 9A H V I D G T R V F L G L A L V A H
7750 B800-850 Y V I D G T R V F L G L A L V A H
7050 B800-850 H V I D G T R V F F V I A I F A H
7750 B800-820(1) Y V I D G A R A F L G I A L V A H
7770 B800-820(2) Y V I D G A R A F V A I A A F A H
7050 B800-820 H V I D G T R V F L V I A A I A H
38 39 40 41 42 43 44 45 46 47 48 49 50---- 51 52 53 54 55 56 57 58
F L A F S A T P W L H M N Q G K I W T
F L A F S A T P W L H M N Q G K I W T
V L A F A F S P W L H M N Q G K I W T
F L A F S A T P W L H M N Q G K I W T
V L A Y S L T P W L H M N Q G K I W T
F L A F T L T P W L . M N Q G K I W T
59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81
V V D P A F G I P A L L G S V T V I A L L V H
V V N P A I G I P A L L G S V T V I A I L V H
V V N P S V G L P L L L G S V T V I T A L V H
V V N P A V G L P L L L G S V A I T A L V H
V V N P A V G L P L L L G S V A I T A L V H
V V P P A F G L P L M L G A V A I T A L V H
82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102
L A I L S N T T W F P A Y W Q G G V K K A
L A I L S H T T W F P A Y W Q G G V K K A
A A V L S H T T W F P A Y W Q G G . K K .
L A V L T H T T W F P A F T Q G G L K K A
L A V L T H T T W F P A F T Q G G L K K A
A A V L T H T T W Y A A F L Q G G . K K .
103 104 105 106 107 108 109 110 111
A A I E T T V L G
A
.
A
A
.

Figure 9.21

sequence occurs immediately after the termination codon of the beta polypeptide giving the sequence TAAGGAG.

The deduced amino acid sequences of these two presumtive polypeptides have been compared to the native *R. acidophila* LHII polypetides of strains 7050 and 7750, see figure 9.21. The result of this comparison in the form of a difference table is shown below.

		Amino Acid Differences		
R. acidophila Strain	Antenna Complex Type	Polypeptide Type BETA	ALPHA	TOTAL
7750	B800-850	1	4	5
7050	B800-850	8	9	16
7750	B800-820[1]	4	13	17
7750	B800-820[2]	11	13	24
7050	B800-820	8	8	16

Table 9.4

The scores shown represent how the native polypeptides differed with respect to the presumptive polypeptides. The differences shown for the native beta polypeptides were determined only for that piece of sequence which was available from the presumptive polypeptide. I have made the assumption that the N-terminal sequence is missing through not being subcloned and therefore have not scored against the missing residues.

From the above data the presumptive polypeptides appear most like those of the B800-850 light harvesting polypeptides from *R. acidophila* strain 7750. There are only 5 differences in total between these sequences whereas all other sequences have between 16 and 24 differences. If the positions where the differences occur between strain 7750 B800-850 polypeptides and the presumptive polypeptides are examined, it will be seen that with one exception, the residue present in the presumptive polypeptide can also be found at that same position in at least one of the other

native polypeptides. The exception is found at position 61 where aspartic acid, proline and asparagine are found in the presumptive, B800-820 strain 7050, and all the other alpha polypeptides respectively. Like the other presumptive alpha polypeptides presented, this alpha polypetide also has a carboxy terminal tail of, in this case 7 residues (I-E-T-T-V-L-G). However, even with these differences, these two presumptive polypeptides again have the overall appearance or feel of being genuine light harvesting polypeptides.

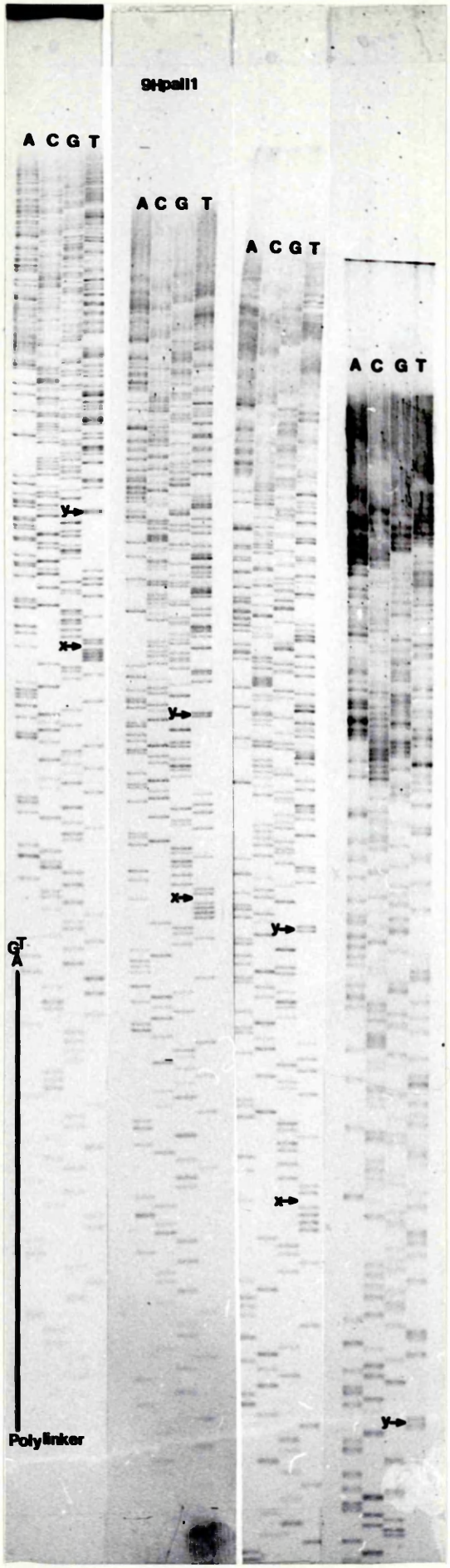
Consensus sequence 9B

This sequence was obtained entirely from the mp19 recombinant 9HpaII1. Two other mp19 recombinants 9HpaII7 and 9HpaII9 had an identical sequence, so although the sequence of both strands has not be determined, the sequence has at least been checked from independent clones. The sequencing gel autoradiographs of 9HpaII1 are presented in fig 9.22 and the sequence itself in figure 9.23a.

The result of the Bestfit analysis between the 9HpaII1 and pLHIISB18 sequences is shown in figure 9.23b. From this it can be seen that the region between nucleotides 86 and 253 of the 9HpaII1 sequence 69.4% identity with the region between nucleotides 376 and 546 of pLHIISB18. This region of pLHIISB18 encodes most of the beta and part of the alpha B800-850 light harvesting polypeptide of *Rb. sphaeroides*.

Translation of the 9HpaII1 sequence in six reading frames revealed the presumptive light harvesting polypeptides shown in figure 9.24. The amino acid sequences comprise a complete beta and about 20% of an alpha polypeptide.

The region encoding the beta polypeptide is 130 nucleotides in length and is separated from the alpha polypeptide coding region by a 12bp sequence. Both presumptive polypeptides have the initiation codon ATG. The termination codon of the beta polypeptide is TAA.



≃2hrs ≃4hrs ≃7hrs ≃9hrs

Figure 9.22

Shown opposite are four gel autoradiographs of a template which gave consensus sequence 9HpaII1.

The gels have been run for ≃2, ≃4, ≃7 and ≃9 hours under the conditions described previously.

The nucleotide sequence deduced from these autoradiographs is shown overleaf as the reversed and complemented sequence.

X and Y are reference points in the sequence.

The junction between the subcloned HpaII fragment and the mp19 polylinker is shown by the sequence AGT.

```
   1  CGGTCGCCGA GCCACAGGTG CAGAATCGCG TGCGGTCAAC GCACGATAAA

  51  CAGTGGCTTT TATTGAGAGG TGGACAATGG CTGTACTCAA CGAGGCGCAA

 101  GCCGAAGAAC TTCACAAGCA CGTCATTGAC GGCGCTCGCG TGTTTGGCGT

 151  CATCGCGCTC TTTGCGCATG TTCTCGCCTT ATCGCTGACC CCCTGGCTGC

 201  ACTAAGGAGT TCTCATCATG AACCAAGGCA AGATCTGGAC CGTCGTTAAC

 251  CCG
```

A

```
      Gap Weight:  5.000      Average Match:  1.000
  Length Weight:  0.300   Average Mismatch: -0.900

           Quality:   58.6              Length:    172
             Ratio:  0.349               Gaps:      2
 Percent Similarity: 69.461   Percent Identity: 69.461

 9HpaII1.dat x Sphaerlh2.dat April 5, 1990  18:28  ..

                .         .         .         .         .
  86 CTCAACGAGGCGCAAGCCGAAGAACTTCACAAGCACGTCATTGACGGCGC 135
     || | ||  ||  |||||||||||| |||| ||||| ||||  |||| |
 376 CTGACCGTTGCCGAAGCCGAAGAAGTTCATAAGCAACTCATCCTCGGCAC 425
                .         .         .         .         .
 136 TCGCGTGTTTGGCGTCATCGCGCTCTTTGCGCATGTTCTCGCCTTATCGC 185
      ||||| || || | ||| |||||| | |||||  | ||||||    |
 426 CCGCGTCTTCGGTGGCATGGCGCTCATCGCGCACTTCCTCGCCGCCGCTG 475
                .         .         .         .         .
 186 TGACCCCCTGGCTGCACTAA....GGAGTTCTCATCATGAACCAAGGCAA 231
      |||||| |||||   || |     ||   || | ||||| | | |||||
 476 CGACCCCGTGGCTCGGCTGATAGGAGAAGACTGA.CATGACCAACGGCAA 524
                .         .
 232 GATCTGGACCGTCGTTAACCCG 253
      ||||||  ||| || || |||
 525 AATCTGGCTCGTGGTGAAACCG 546
```

B

Figure 9.23

The reverse complement of the 9HpaII1 sequence shown in Figure 9.22 is shown in A above. Bestfit analysis of this sequence (86-253) to the sequence of pLHIISB18 (376-546) was then made. The result of this analysis is shown in B.

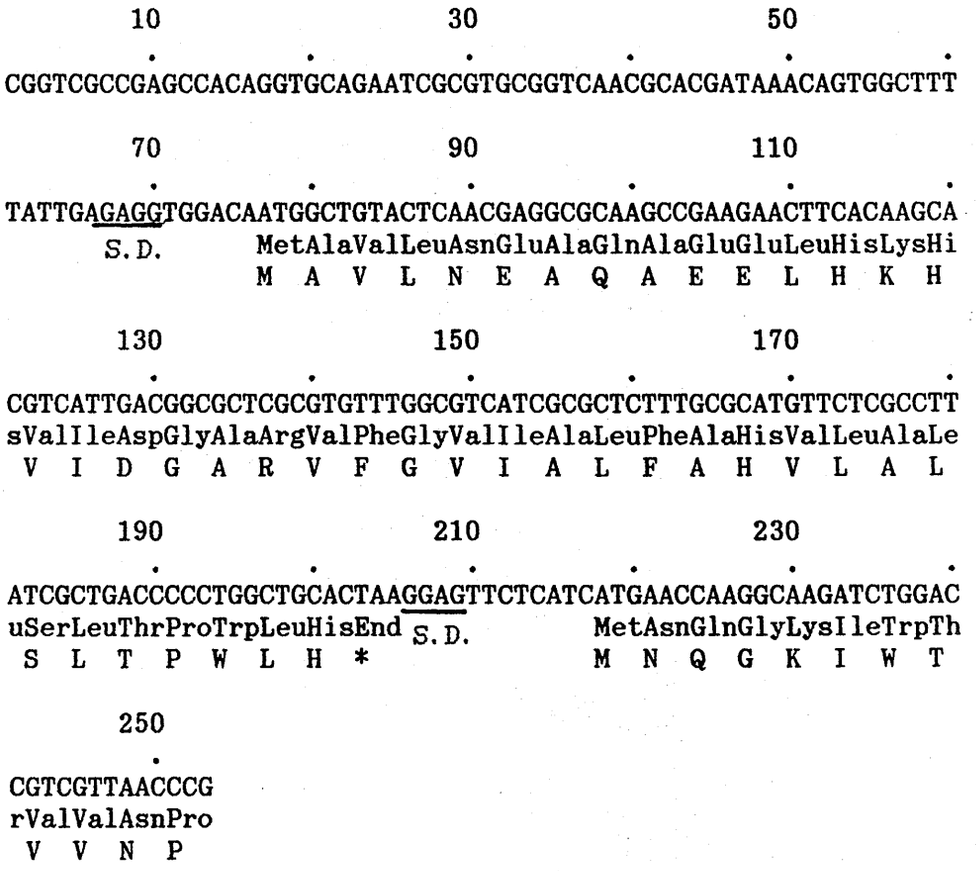


Figure 9.24

The reverse complement of the 9HpaII1 sequence was translated in the six possible reading frames, one of which is shown above. The sequence translates into a presumptive beta and part of an alpha light harvesting polypeptide. Possible Shine-Delgarno sequences are indicated by S.D.

Figure 9.25

The presumptive 9B beta (positions 7-50) and alpha (positions 51-62) polypeptides have been aligned with native B800-850 and B800-820 beta (positions 1-50) and alpha (positions 51-62) polypeptides from *R. acidophila* strains 7050 and 7750.

The reference positions of alignment are the conserved histidine residues (*) at positions 19 and 37, and the initiating methionine at position 51.

The presumptive polypeptides are presented in bold typeface.

Positions where there is absolute residue conservation have been blocked.

At other positions, where the native polypeptides differ from the presumptive polypeptides the differences have been signified by underlining the relevant residue eg. A.

Where the type of amino acid residue at a particular position is unknown, this is signified by (.).

The differences between the presumptive and native polypeptides have been quantified and are presented in the difference score table shown in the text.

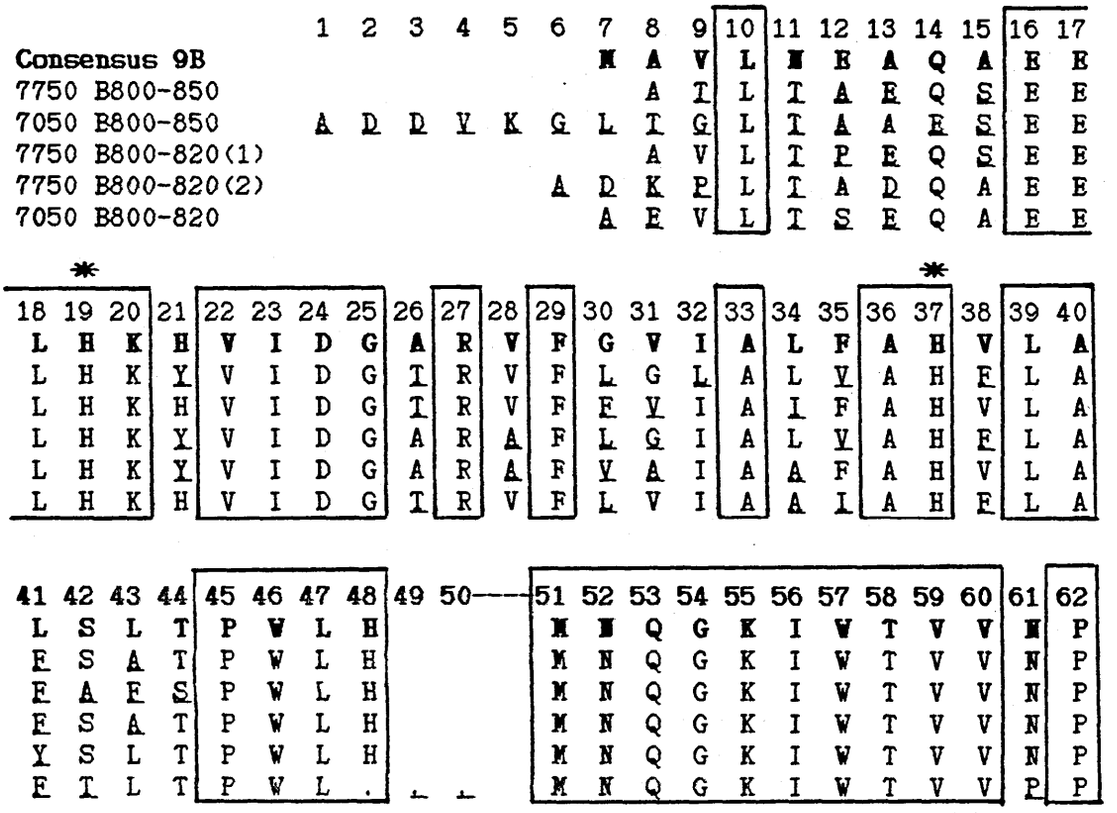

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Consensus 9B M A V L M E A Q A E E
7750 B800-850 A T L T A E Q S E E
7050 B800-850 A D D V K G L T G L T A A E S E E
7750 B800-820(1) A V L T P E Q S E E
7750 B800-820(2) A D K P L T A D Q A E E
7050 B800-820 A E V L T S E Q A E E
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
L H K H V I D G A R V F G V I A L F A H V L A
L H K Y V I D G T R V F L G L A L V A H F L A
L H K H V I D G T R V F F V I A I F A H V L A
L H K Y V I D G A R A F L G I A L V A H F L A
L H K Y V I D G A R A F V A I A A F A H V L A
L H K H V I D G T R V F L V I A A I A H F L A
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
L S L T P W L H M M Q G K I W T V V M P
F S A T P W L H M N Q G K I W T V V N P
F A F S P W L H M N Q G K I W T V V N P
F S A T P W L H M N Q G K I W T V V N P
Y S L T P W L H M N Q G K I W T V V N P
F T L T P W L . M N Q G K I W T V V P P

Figure 9.25

Upstream of both coding regions are possible Shine-Delgarno sequences. The sequences GAGAGG and GGAG are 6 and 8bp upstream of the presumptive alpha and beta coding regions respectively. The GGAG sequence follows directly after the termination codon of the beta coding region ie. TAAGGAG.

The deduced amino acid sequences have been compared to those of the light harvesting polypeptides from *R. acidophila* strains 7050 and 7750. This is shown in figure 9.25 and table 9.5.

		Amino Acid Differences		
R. acidophila strain	Antenna complex type	Polypeptide Type BETA	ALPHA	TOTAL
7750	B800-850	13	0	13
7050	B800-850	21	0	21
7750	B800-820[1]	12	0	12
7750	B800-820[2]	12	0	12
7050	B800-820	14	1	15

Table 9.5

The scores shown represent how the native polypeptides differ with respect to the presumptive polypeptides. The differences shown for the alpha polypeptide were determined only for that piece of sequence which was available. I have made the assumption that the C-terminal sequence is missing through not being subcloned and therefore have not scored against the missing sequence.

The table above suggests that the presumptive sequences are most like those of either of the two B800-820 sequences from strain 7750. However, having said that it also quite clear that there is only 77% identity between the presumptive and native sequences.

The presumptive beta polypeptide appears to be a hybrid of all of the native sequences. By this I mean, that if any position of the presumptive polypeptide is examined, the residue which is found can

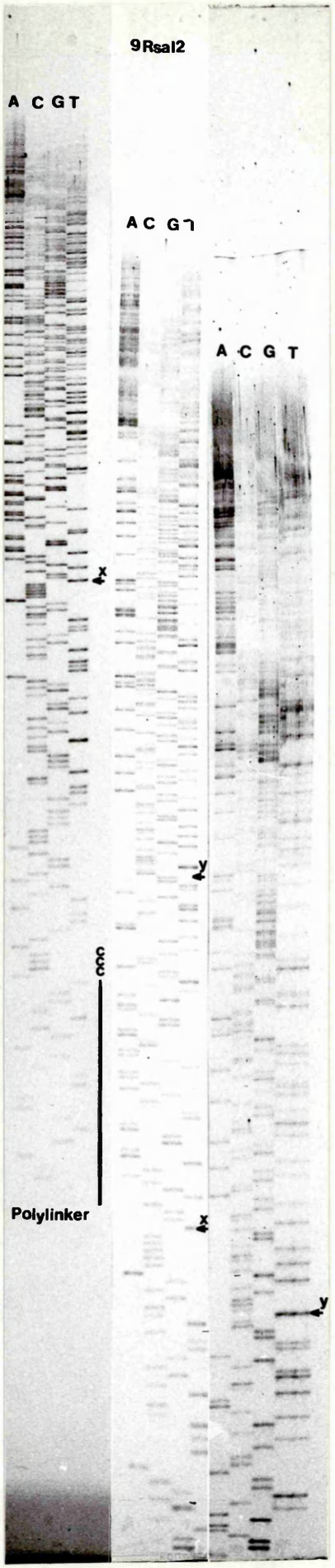
also be found in the same position in at least one of the other polypeptides. There is one exception to this and that is position 30 where glycine (G) is found in the presumptive polypeptide and leucine (L), or phenylalanine (F) or valine (V) are found in the native polypeptides. However, even with the lack of identity to any particular polypeptide sequence the presumptive polypeptide sequence still has the overall appearance of a light harvesting polypeptide.

When making the comparisons between the native and presumptive polypeptides it must be remembered that the native sequences are those of polypeptides which have been successfully isolated and puified. They are not necessarily all the light harvesting polypetides which the organism is capable of synthesising. I think it is possible that the presumptive polypeptides encoded in consensus sequence 9B have not as yet been isolated as native polypeptides. This hypothesis will be discussed in greater detail in chapter 11.

Consensus sequence 9C

Consensus sequence 9C was derived from the melding of five contigs, 9AluI1, 9HincII1, 9HincII2, 9HpaII3 and 9RsaI2. The sequencing gel autoradiograph of the 9RsaI2 sequencing gel, which encodes most of the two presumptive polypeptides is shown in figure 9.2. The sequences of all five contigs are presented in figure 9.27.

The schematic representation in figure 9.27f shows how the contigs overlapped to form the consensus sequence. From this figure and the consensus sequence itself, figure 9.28, it can be seen that both strands of the consensus, from nucleotide 81-204 and 312-460 have been sequenced. The sequence between nucleotide 204 and 312, though not confirmed by sequencing both DNA strands, has been checked by sequencing the one strand from two independent clones containing two different restriction fragments. The first 81 and last 74 nucleotides have not been checked in some way.



≃2hrs ≃4hrs ≃7hrs

Figure 9.26

Shown opposite are three gel autoradiographs of a template which gave contig sequence 9RsaI2. This contig encodes all of a presumptive beta and most of an alpha light harvesting polypeptide.

The gels have been run for ≃2, ≃4 and ≃7 hours under the conditions described previously.

The nucleotide sequence taken from these autoradiographs is shown overleaf.

X and Y are reference points in the sequence.

The junction between the subcloned RsaI2 fragment and the mp19 polylinker is shown by CCC.

```
  1  GCGAAGCGCT CAGCCGAGAA CGGAGGTCTC GATCGCGGCG ACCTTCTTGC
 51  CACCCTGCCA GTAGGCCGGG AACCAGGTGG TGTGCGACAG GATCGCAACA
101  TGCACGAGGA TCGCGATCAC GGTCACGGAG CCGAGGAGCA GCGGGAGGCC
151  AATCGAAGGA TTGACGACGG TCCAGATCTT GCCTTGGTTC ATTGTCGGAA
201  TTCCTTAGTG CAGCCAG
```

A

```
  1  CCCGGGTGTT CCTGGGCCTC GCGCTGGTCG CGCATTTCCT CGCGTTCTCC
 51  GCGACCCCCT GGCTGCACTA AGGAATTCCG ACAATGAACC AAGGCAAGAT
101  CTGGACCGTC GTCAATCCTT CGATTGGCCT CCCGCTGCTC CTCGGCTCCG
151  TGACCGTGAT CGCGATCCTC GTGCATGTTG CGATCCTGTC GCACACCACC
201  TGGTTCCCGG CCTACTGGCA GGGTGGCAAG AAGGTCGCCG CGATCGAGAC
251  CTCCGTTCTC GGCTGAGCGC TTCGCTGACG TC
```

B

```
  1  TGCGGAGCAA TCCGAAGAAC TGCACAAGTA CGTTATCGAC GGCACCCGGG
 51  TGTTCCTGGG CCTCGCGCTG GTCGCGCATT TCCTCGCGTT CTCCGCGACC
101  CCCTGGCTGC ACTAAGGAAT TCCGACAATG AACCAAGGCA AGATCTGGAC
151  CGTCGTCAAT CCTTCGATTG GCCTCCCGCT GCTCCTCGGC TCCGTGACCG
201  TGATCGCGAT CCTCGTGCAT GTTGCGATCC
```

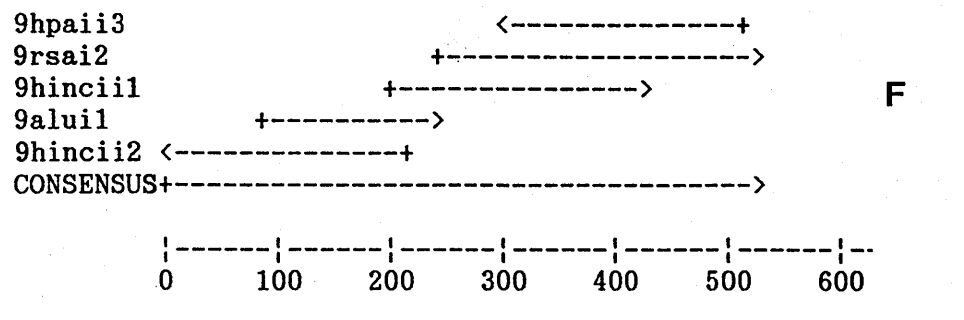
C

```
  1  CCACCTTCTG CACGGCGCGC CGTGCATTGG TCGCTAGCCA CAAGCGAATA
 51  ATCGCGTGCG TTCGCGCACG AGAAAATTGT GGTGACTTAC AAGAGGTGGA
101  CAATGGCGAC GTTGACTGCG GAGCAATCCG AAGAACTGCA CAAGTACGTT
151  AT
```

D

```
  1  TTGCTCCGCA GTCAACGTCG CCATTGTCCA CCTCTTGTAA GTCACCACAA
 51  TTTTCTCGTG CGCGAACGCA CGCGATTATT CGCTTGTGGC TAGCGACCAA
101  TGCACGGCGC GCCGTGCAGA AGGTGGGGTC ACTATAGCTT GACACTTTTT
151  ACTGTCAATA TCGAATGACA TTTTTTTTAA TGCGACCGCA TTGTCGCAAC
201  ATGCGGACCG GACAGACGC
```

E



F

Figure 9.27

Consensus sequence 9C was obtained by melding the five contig sequences shown opposite. The contigs were 9HpaII3 (A); 9RsaI2 (B); 9HincII1 (C); 9AluI1 (D); and 9HincII2 (E).

Note that contig 9HpaII3 and 9HincII2 were obtained by sequencing from one strand and the other contigs by sequencing the opposite strand. This is illustrated in the schematic diagram (F).

Consensus sequence 9C is presented overleaf.

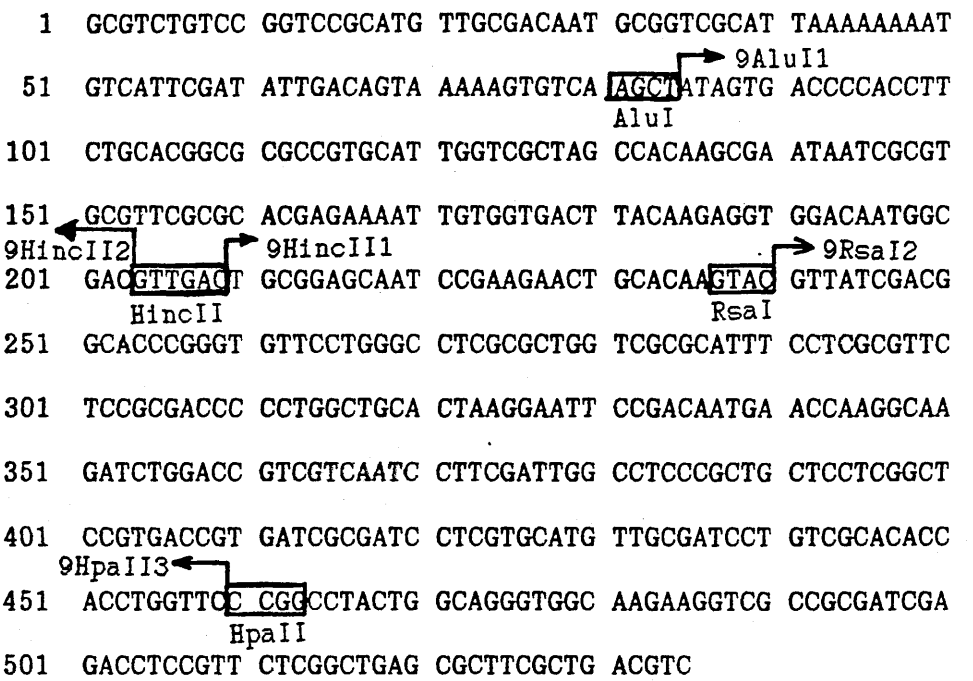


Figure 9.28

Shown above is consensus sequence 9C which was generated by melding the contigs shown in Figure 9.27. Where possible the restriction sites which permitted the sub-cloning of the contig fragments are boxed. The direction in which the fragments were sequenced is arrowed.

```
        10                  30                  50
         .         .         .         .         .         .
GCGTCTGTCCGGTCCGCATGTTGCGACAATGCGGTCGCATTAAAAAAAATGTCATTCGAT

        70                  90                  110
         .         .         .         .         .         .
ATTGACAGTAAAAAGTGTCAAGCTATAGTGACCCCACCTTCTGCACGGCGCGCCGTGCAT

       130                 150                 170
         .         .         .         .         .         .
TGGTCGCTAGCCACAAGCGAATAATCGCGTGCGTTCGCGCACGAGAAAATTGTGGTGACT

       190                 210                 230
         .         .         .         .         .         .
TACAAGAGGTGGACAATGGCGACGTTGACTGCGGAGCAATCCGAAGAACTGCACAAGTAC
     S.D.       MetAlaThrLeuThrAlaGluGlnSerGluGluLeuHisLysTyr
                M  A  T  L  T  A  E  Q  S  E  E  L  H  K  Y

       250                 270 BETA            290
         .         .         .         .         .         .
GTTATCGACGGCACCCGGGTGTTCCTGGGCCTCGCGCTGGTCGCGCATTTCCTCGCGTTC
ValIleAspGlyThrArgValPheLeuGlyLeuAlaLeuValAlaHisPheLeuAlaPhe
V  I  D  G  T  R  V  F  L  G  L  A  L  V  A  H  F  L  A  F

       310                 330                 350
         .         .         .         .         .         .
TCCGCGACCCCCTGGCTGCACTAAGGAATTCCGACAATGAACCAAGGCAAGATCTGGACC
SerAlaThrProTrpLeuHisEnd S.D.       MetAsnGlnGlyLysIleTrpThr
S  A  T  P  W  L  H  *              M  N  Q  G  K  I  W  T

       370                 390                 410
         .         .         .         .         .         .
GTCGTCAATCCTTCGATTGGCCTCCCGCTGCTCCTCGGCTCCGTGACCGTGATCGCGATC
ValValAsnProSerIleGlyLeuProLeuLeuLeuGlySerValThrValIleAlaIle
V  V  N  P  S  I  G  L  P  L  L  L  G  S  V  T  V  I  A  I

       430                 450 ALPHA           470
         .         .         .         .         .         .
CTCGTGCATGTTGCGATCCTGTCGCACACCACCTGGTTCCCGGCCTACTGGCAGGGTGGC
LeuValHisValAlaIleLeuSerHisThrThrTrpPheProAlaTyrTrpGlnGlyGly
L  V  H  V  A  I  L  S  H  T  T  W  F  P  A  Y  W  Q  G  G

       490                 510                 530
         .         .         .         .         .
AAGAAGGTCGCCGCGATCGAGACCTCCGTTCTCGGCTGAGCGCTTCGCTGACGTC
LysLysValAlaAlaIleGluThrSerValLeuGlyEnd
K  K  V  A  A  I  E  T  S  V  L  G  *
```

Figure 9.29

Consensus sequence 9C was translated in the six possible reading frames one of which revealed the amino acid sequences shown above. These encode a presumptive alpha and beta light harvesting polypeptide. Possible Shine-Delgarno sequences are signified by S.D.

Figure 9.30

The presumptive 9C beta (positions 7-50) and alpha (positions 51-110) polypeptides have been aligned with native B800-850 and B800-820 beta and alpha polypeptides from *R. acidophila* strains 7050 and 7750.

The reference positions of alignment are the conserved histidine residues (*) at positions 19, 37 and 81.

The presumptive polypeptides are presented in bold typeface.

Positions where there is absolute residue conservation have been blocked.

At other positions, where the native polypeptides differ from the presumptive polypeptides the differences have been signified by underlining the relevant residue eg. A.

Where the type of amino acid residue at a particular position is unknown, this is signified by (.).

The differences between the presumptive and native polypeptides have been quantified and are presented in the difference score table shown in the text.

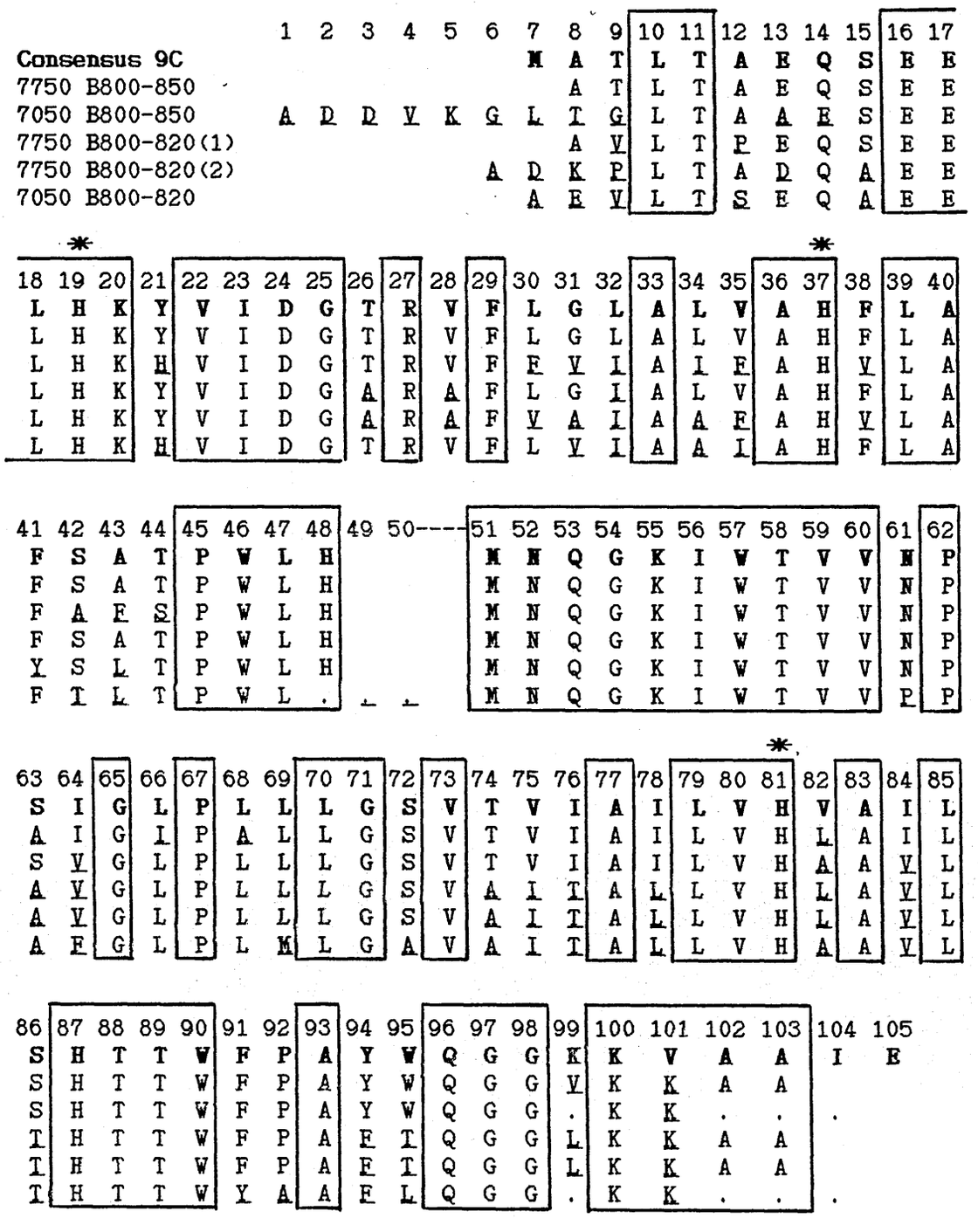

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Consensus 9C M A T L T A E Q S E E
7750 B800-850 A T L T A E Q S E E
7050 B800-850 A D D V K G L T G L T A A E S E E
7750 B800-820(1) A V L T P E Q S E E
7750 B800-820(2) A D K P L T A D Q A E E
7050 B800-820 A E V L T S E Q A E E
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
L H K Y V I D G T R V F L G L A L V A H F L A
L H K Y V I D G T R V F L G L A L V A H F L A
L H K H V I D G T R V F F V I A I F A H V L A
L H K Y V I D G A R A F L G I A L V A H F L A
L H K Y V I D G A R A F V A I A A F A H V L A
L H K H V I D G T R V F L V I A A I A H F L A
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62
F S A T P W L H M N Q G K I W T V V N P
F S A T P W L H M N Q G K I W T V V N P
F A F S P W L H M N Q G K I W T V V N P
F S A T P W L H M N Q G K I W T V V N P
Y S L T P W L H M N Q G K I W T V V N P
F T L T P W L . M N Q G K I W T V V P P
63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85
S I G L P L L L G S V T V I A I L V H V A I L
A I G I P A L L G S V T V I A I L V H L A I L
S V G L P L L L G S V T V I A I L V H A A V L
A V G L P L L L G S V A I T A L L V H L A V L
A V G L P L L L G S V A I T A L L V H L A V L
A F G L P L M L G A V A I T A L L V H A A V L
86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105
S H T T W F P A Y W Q G G K K V A A I E
S H T T W F P A Y W Q G G V K K A A
S H T T W F P A Y W Q G G . K K . . .
T H T T W F P A F T Q G G L K K A A
T H T T W F P A F T Q G G L K K A A
T H T T W Y A A F L Q G G . K K . . .
106 107 108 109 110
T S V L G

Figure 9.30

Note, the HincII site at position 204 (GTTGAC) initially does not appear to possess a twofold axis of symmetry. However, the recognition sequence of this enzyme is GTPy↓PuAC.

The consensus sequence was then translated in six reading frames, one of which generated the amino acid sequences of the presumtive alpha and beta light harvesting polypeptides shown in figure 9.29.

The region encoding these polypeptides is 325bp in length inclusive of start and termination codons. A 12bp non-coding region lies between the regions encoding the two polypeptides. Both coding regions have the intiation codon ATG. The alpha and beta coding regions are terminated by TGA and TAA respectively. Upstream of each coding region lies a possible Shine-Delgarno sequence. The sequences GAGG and GGAA are found 6bp and 8bp upstream of the beta and alpha initiation codons respectively. The sequence GGAA follows immediately after the termination codon of the beta polypeptide giving the sequence TAAGGAA.

The deduced amino acid sequences were then compared to those of the native B800-850 and B800-820 polypeptides of *R. acidophila* strains 7050 and 7750. This comparison is shown in figure 9.30 and the differences presented in table 9.6.

		Amino Acid Differences		
R. acidophila Strain	Antenna Complex Type	Polypeptide Type Beta	Alpha	Total
7750	B800-850	0	6	6
7050	B800-850	21	4	25
7750	B800-820[1]	5	12	17
7750	B800-820[2]	16	13	29
7050	B800-820	14	17	31

Table 9.6

The difference scores in table 9.6 show how many differences there are between the native polypeptides with respect to the presumptive polypeptides.

From the above table it can be seen that the presumptive polypeptide sequences are most like those of the native B800-850 antenna complex from *R. acidophila* strain 7750. The primary sequences of the two beta polypeptides are identical, and there are 6 differences between the alpha polypeptides. The alpha polypeptide of the B800-850 complex of strain 7050 has only 4 differences to the presumptive alpha polypeptide. However, the differences between their beta polypeptides is such that the total difference score is 25.

Like the other presumptive alpha polypeptides described in this chapter, there is a carboxy terminal sequence, in this case, of 7 residues (I-E-T-S-V-L-G) which is not found in any of the native polypeptides. A possible role for such sequences and other features of the polypeptides will be considered in chapter 11.

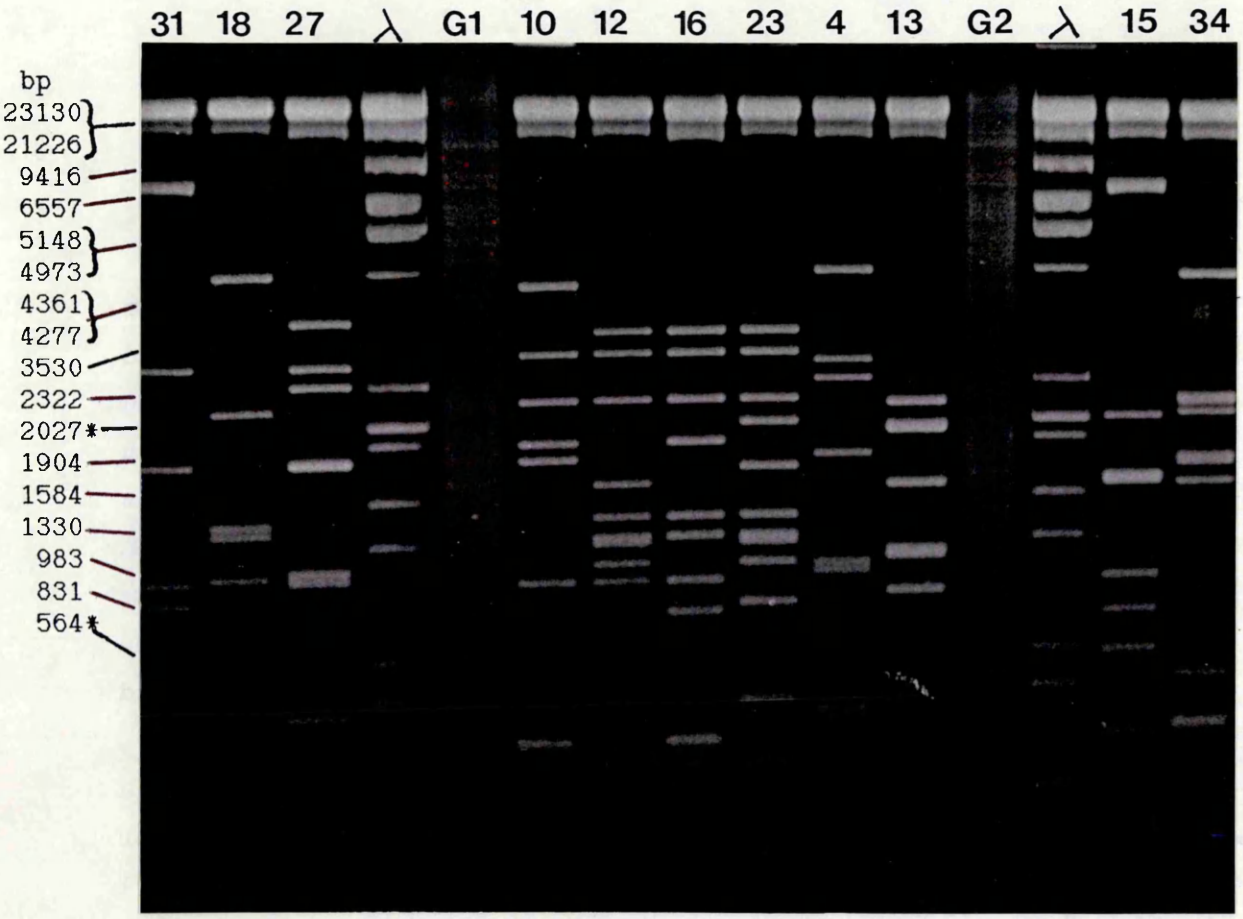
Chapter 10 VERIFYING THE SOURCE OF THE GENOMIC CLONE INSERTS

The following strategy was used to verify that the genomic clones contained inserts derived from *R. acidophila* strain 7050 genomic DNA. All of the genomic clones and *R. acidophila* strain 7050 genomic DNA were digested with SalI+EcoRI. The DNAs were run on gels, which were then blotted and the filters probed with radiolabelled pLHIISB18. If the genomic inserts contained *R. acidophila* derived DNA then positively hybridising bands of the same size would be visualised in both the clone and genomic lanes. The gel photographs and autoradiographs are shown in figures 10.1 and 10.2.

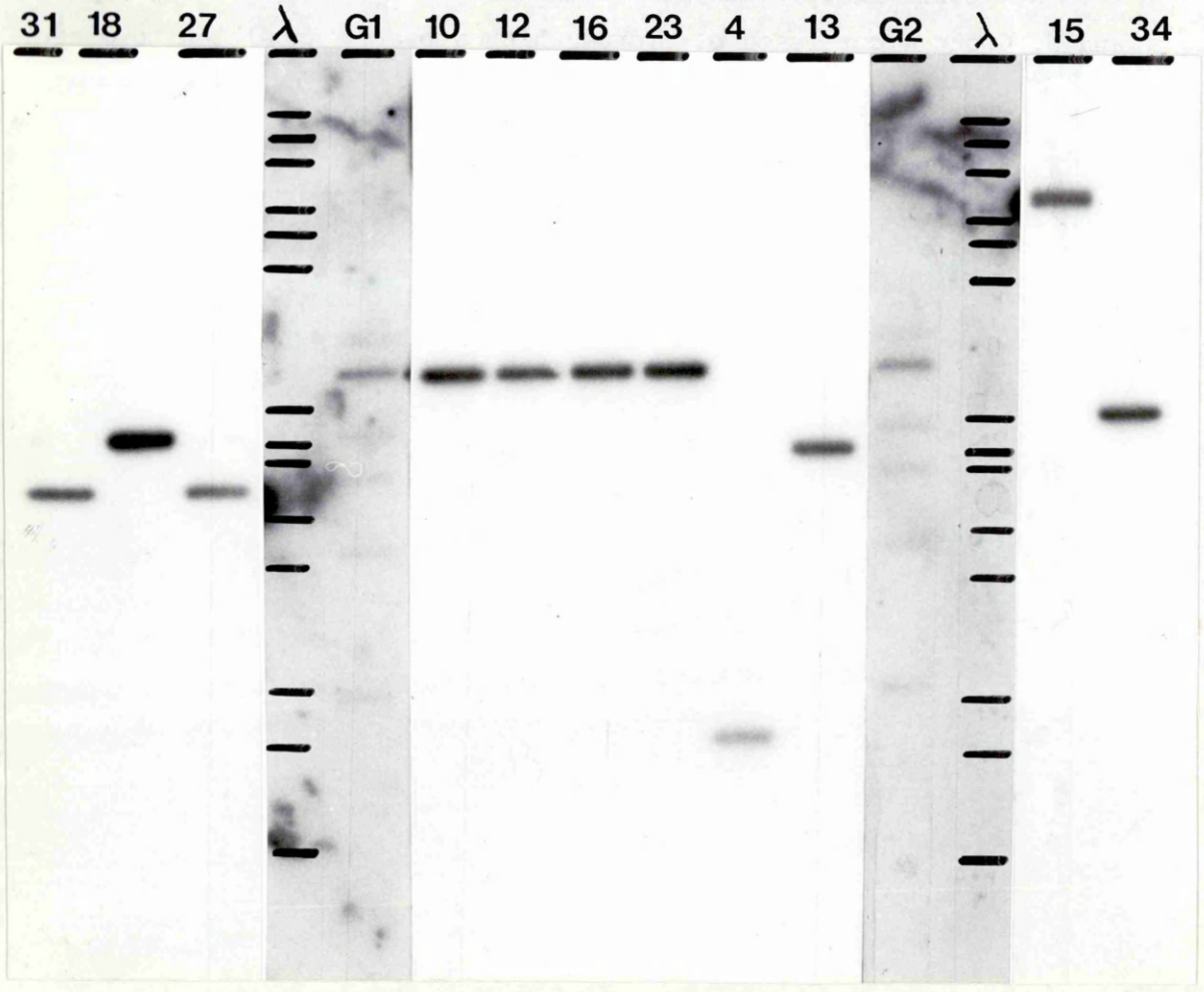
On the basis of restriction pattern, the genomic clones had been placed in four groups, see chapter 7. Members within group 1 overlapped with other members within that group, but not with clones from any other group. The same situation was also found in groups 2 and 3. Group 4 consisted of clones which appeared unrelated to other members within group 4 and unrelated to members within the other three groups ie. a group of "unique individuals".

Clones from groups 3 and 4 were run on the gel shown in figure 10.1a. The group 3 clones were numbers 10, 12, 16, and 23 and all share a positively hybridising 2.4kb SalI band, figure 10.1b. This appears to coincide with a positively hybridising band present in genomic lane G1, suggesting that this group of clones contain strain 7050 derived inserts.

It was believed that group 4 clones (the remainder on this gel) fell into two classes; a) those which contained intact light harvesting genes, and b) those in which the genes had been cleaved during Sau3AI partial digestion at the time of generating inserts. The first group would have positively hybridising fragments which would have counterparts in the *R. acidophila* genome. The second group


31 18 27 λ G1 10 12 16 23 4 13 G2 λ 15 34
bp
23130
21226
9416
6557
5148
4973
4361
4277
3530
2322
2027*
1904
1584
1330
983
831
564*

a


31 18 27 λ G1 10 12 16 23 4 13 G2 λ 15 34

b

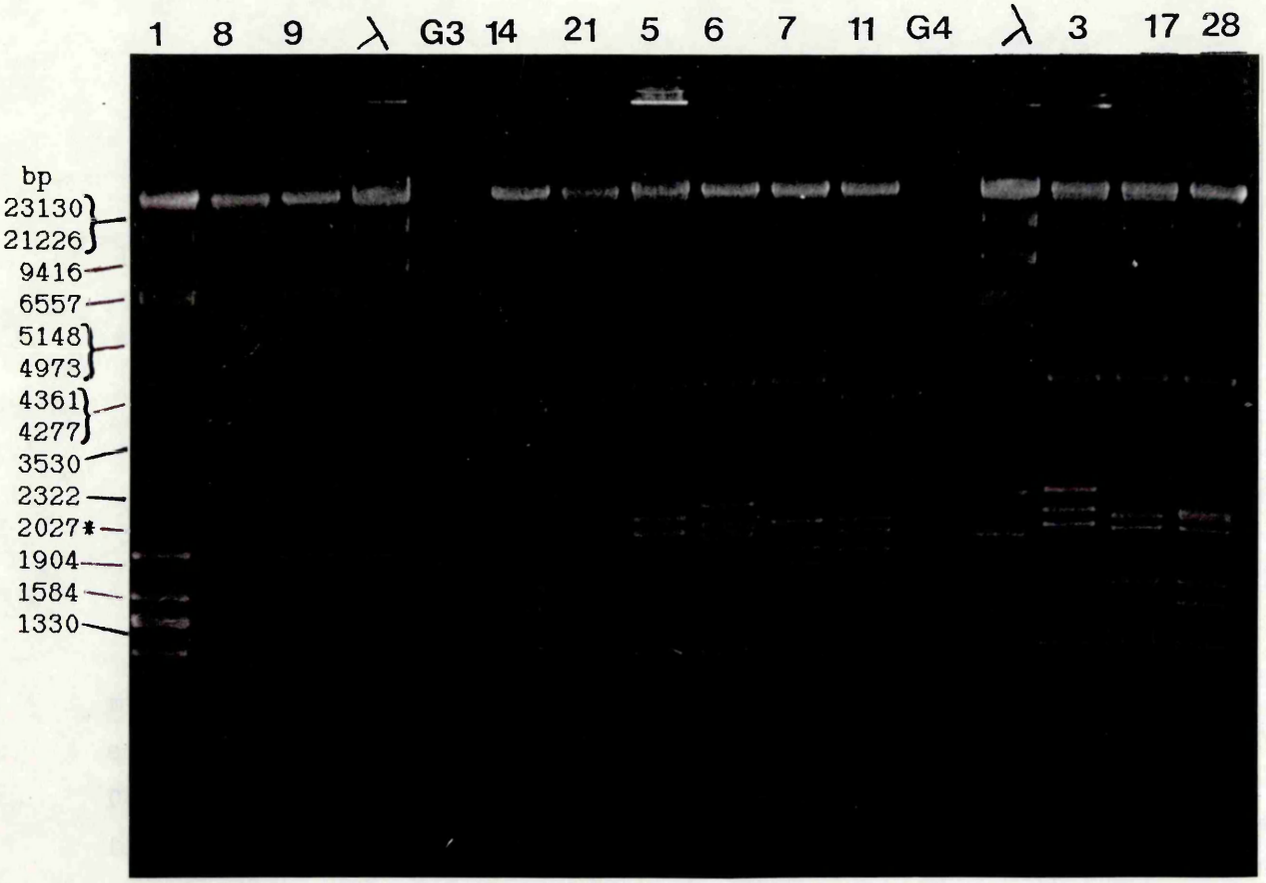
Figure 10.1a+b

To verify the source of the genomic clone inserts, clones from groups 3, (Nos. 10, 12, 16, and 23),and 4,(the remainder of the clone numbers), and genomic DNA from *R. acidophila* strain 7050 (G1+2) were digested with SalI+EcoRI and run on a 0.8% agarose gel shown in (a).

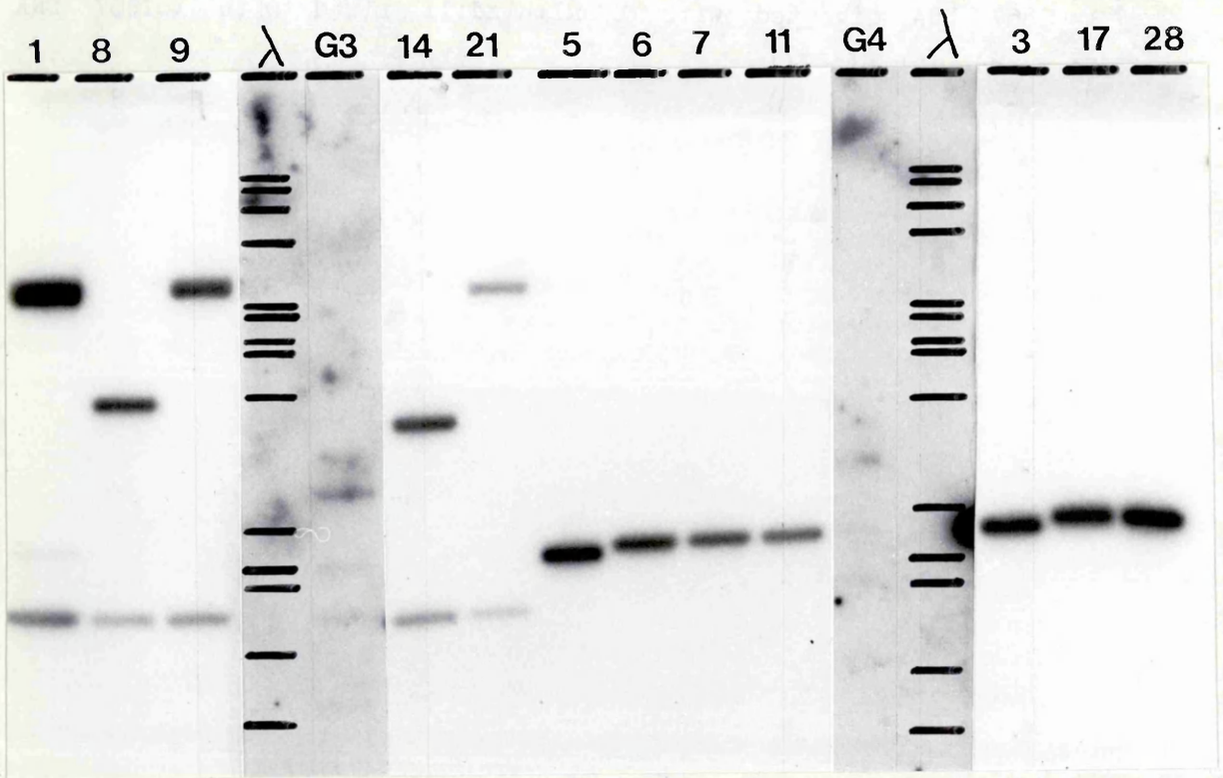
λ=cI857 DNA digested with EcoRI+HindIII mixed with λcI857 DNA digested with HindIII.

The gel was blotted and the filter probed with radiolabelled pLHIISB18 DNA. The resulting autoradiograph is shown in (b).
Positively hybridising clone bands were then compared to those formed in the genomic lanes.

The position of the λ size markers shown on the gel have been marked with black felt pen on the autoradiograph.


1 8 9 λ G3 14 21 5 6 7 11 G4 λ 3 17 28
bp
23130
21226
9416
6557
5148
4973
4361
4277
3530
2322
2027*
1904
1584
1330

a


1 8 9 λ G3 14 21 5 6 7 11 G4 λ 3 17 28

b

Figure 10.2a+b

To verify the source of the genomic clone inserts, clones from groups 1 (Nos. 1, 8, 9, 14, and 21) and 2 (the remainder of the clone numbers) and genomic DNA from *R. acidophila* strain 7050 (G3+4) were digested with SalI+EcoRI and run on a 0.8% agarose gel shown in (a).

λ=cI857 DNA digested with EcoRI+HindIII mixed with λcI857 DNA digested with HindIII.

The gel was blotted and the filter probed with radiolabelled pLHIISB18 DNA. The resulting autoradiograph is shown in (b).
Positively hybridising clone bands were then compared to those formed in the genomic lanes.

The position of the λ size markers shown on the gel have been marked with black felt pen on the autoradiograph.

would have positively hybridising fragments which were smaller than those found in the genome.

Clones 31, 18, 27, and 34 appeared to have positively hybridising bands which had counterparts in genomic lanes G1 and G2. These clones may therefore contain intact light harvesting genes. However, they may also contain cleaved light harvesting genes which by coincidence are of the same size as the positively hybridising genomic bands.

Clones 4 and 13 have positively hybridising bands which do not coincide with genomic bands. These were considered to have resulted from the internal cleavage of light harvesting genes and therefore were smaller than the genomic bands from which they had arisen.

A conclusive way of determining the source of these clones would have been the following. A clone could have been digested with EcoRI+SalI then run on a low melting point agarose gel. Restriction bands which were thought to be representative of the *in vivo* state could have been excised and radiolabelled. The EcoRI+SalI digested clone and genomic DNAs could then be run out next to each other on a gel, blotted then probed with the radiolabelled fragment. If the genomic DNA had given rise to the clone insert, adjacent positively hybridising bands would be found in the two DNA tracks. This process could be repeated for the other group 4 clones.

Clone 15 was another group 4 clone which did not have a positively hybridising band in alignment with a visible genomic band. However, in this case the size of the positively hybridising band was larger than those found within the genome. This clearly presented a problem which I will later return to.

Clones of group 2, ie. 5, 6, 7, 11, 3, 19 and 28 had a positively hybridising SalI band of 2.2kb, see figure 10.2b. There were also positively hybridising band of this size in genomic lanes G2 and G3.

Again suggesting that clones from this group contained inserts derived from strain 7050.

Group 1 clones ie. 1, 8, 9, 14 and 21 fell into two classes. Clones 1, 9 and 21 were shown to contain 1.9 and 5.0kb positively hybridising EcoRI fragments, *see* chapter 7. These fragments were thought to be representative of those found in the *R. acidophila* genome. Clones 8 and 14 contained positively hybridising bands of 1.9 and ≈2.5kb. The smaller band was the same 1.9 kb EcoRI band found in the other group 1 clones. The larger 2.5kb band was composed of EcoRI-SalI fragments and was thought to have resulted from internal Sau3AI cleavage of the 5kb fragment when the inserts were being prepared. Therefore, 1.9 and 5.0kb positively hybridising genomic bands were expected to be found, but not those of 2.5kb.

As expected a 1.9kb positively hybridising band was found in genomic track G3. However, a positively hybridising 5kb genomic fragment was not visible. As the 5kb band of the group 1 clones, and the 5kb band of clone 15 from group 4 was not visible, it was thought that these clones must have arisen through the inserts having been derived from another strain of bacteria.

I now believe that this assumption was incorrect and that the failure to observe a positively hybridising 5kb genomic band was due to poor DNA transfer to the filter during blotting. It would be fair to ask why have the positive 5kb clone bands been visualised, but not those of the genomic DNA?

It should be remembered that ≈0.5µgs of clone DNA and ≈2µgs of genomic DNA were loaded in the respective wells. Therefore the amount of DNA per genomic band available for hybridisiation will be very much smaller than a clone counterpart. The relative amounts of material available for hybridisation can be calculated approximately as follows.

SalI and EcoRI both recognise hexanucleotide sequences therefore on average they will each cut every 4^6bp ie. 4096bp. As there are two enzymes a cut, on average, will occur every 2048bp. If we take the size of the *R. acidophila* genome to be the same size as that of *E. coli* (4.2×10^6bp), then digestion with EcoRI+SalI will generate about 2000 gel bands. As 2µgs of DNA were loaded each band would contain ≃1ng of DNA.

On the other hand ≃0.5µgs of clone DNA were loaded per lane, and each clone has ≃5 bands. Therefore, each band would have ≃100ngs of DNA. In order for the clone bands and genomic bands to have the same signal strength autoradiography would need to be carried out for 100 times as long for the genomic tracks. The clone tracks were autoradiographed for 18 hours, therefore 75 days would be required for the genomic tracks (assuming there was no decay of the probe). As the genomic tracks were only given 5 days autoradiography it is perhaps not surprising that the 5kb genomic band is not visible.

However, the smaller positive genomic bands are visible, and showed up after 5 days autoradiography, why then not the 5kb genomic band? If DNA transfer to the filter was uneven ie. the small fragments transferred more efficiently than the large fragments then the small fragments will show up more clearly after autoradiography. As there was such a vast excess of clone DNA, even if differential transfer did occur there would be sufficient clone DNA to give a strong positive hybridisation signal. However, it is unlikely that under the same conditions the genomic DNA bands would be visualised.

As was stated earlier the lack of the positively hybridising 5kb band had suggested that the library contained inserts derived from another strain of bacteria. As *R. acidophila* strain 7750 could survive the low pH of Pfennig's medium it was thought that this may have been the contaminant. To test this hypothesis genomic DNA was isolated from *R. acidophila* strain 7750. In addition genomic DNA was also prepared from all other photosynthetic bacterial species being used in the lab at the time of library construction. DNA from

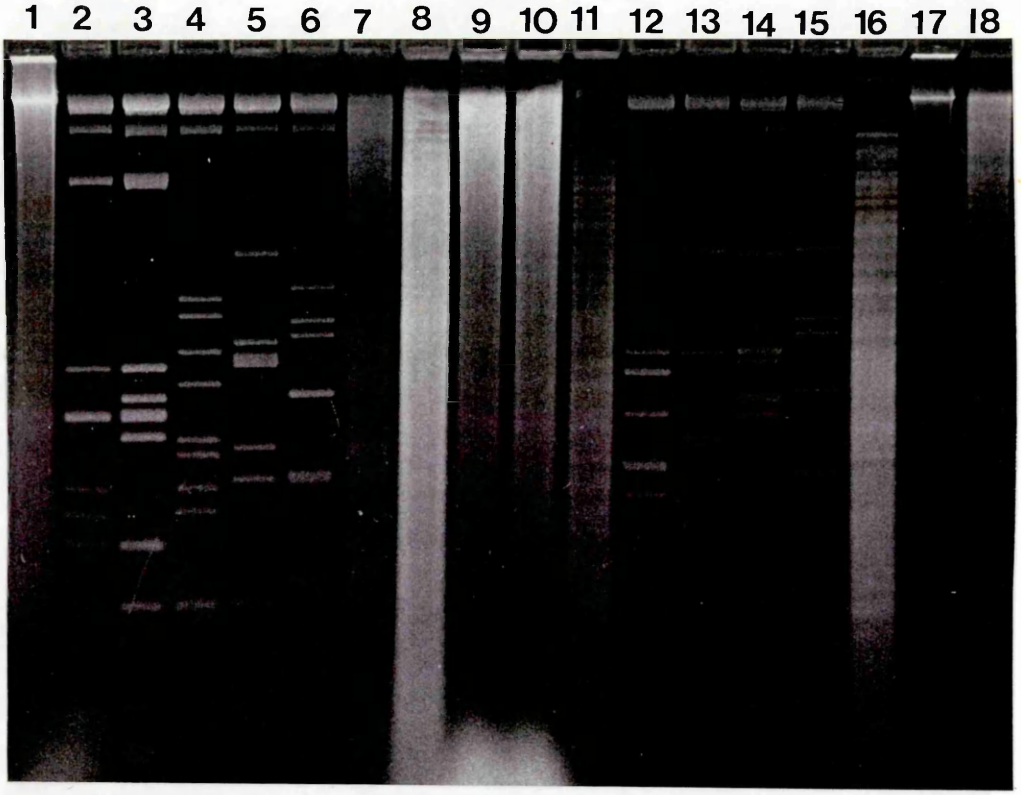


Figure 10.3a

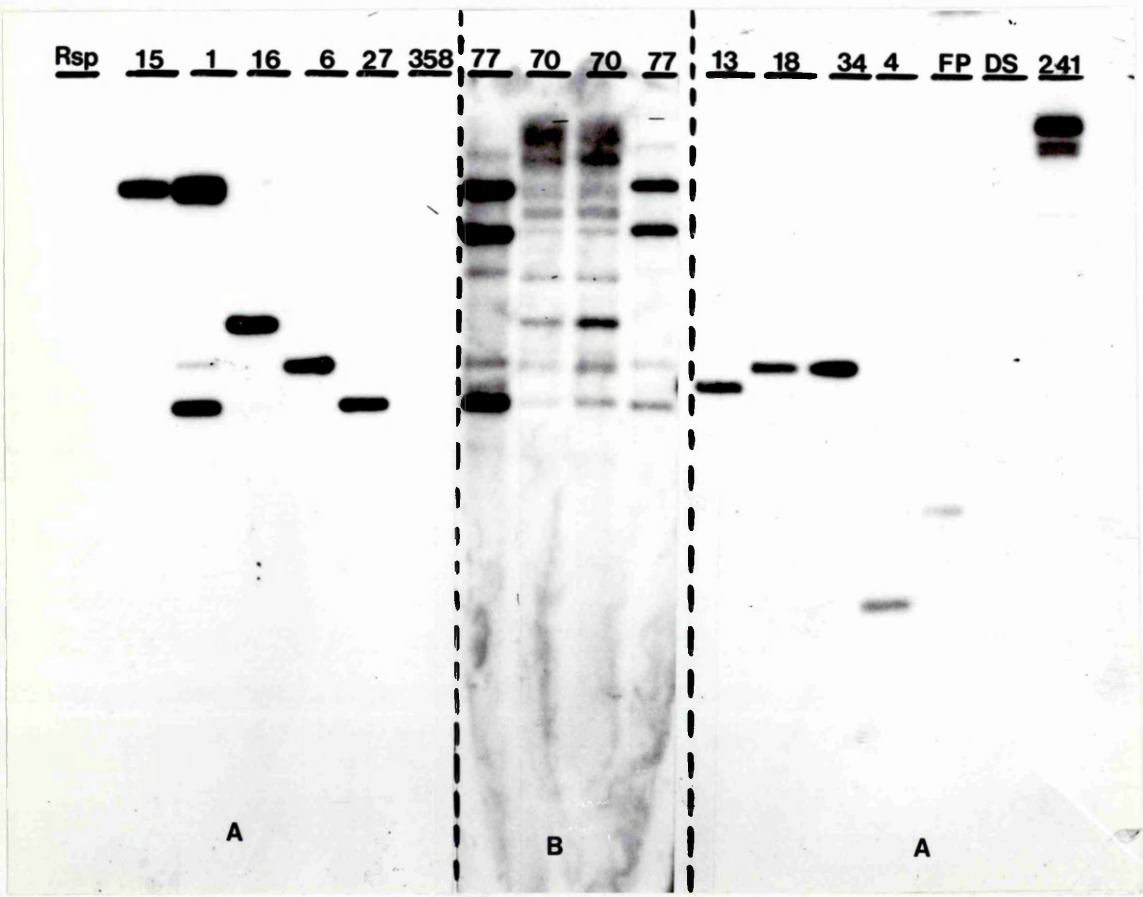


Figure 10.3b

Figure 10.3a

To determine the source of the clone inserts, genomic DNAs from a variety of bacterial species (and strains), and a representative sample of genomic clone DNAs were digested with EcoRI+SalI.
The digests were run on the gel shown opposite. The lanes contain the following DNAs.

Lane 1 : *Rhodospirillum rubrum* genomic DNA.
Lane 2 : Genomic clone 15 DNA.
Lane 3 : Genomic clone 1 DNA.
Lane 4 : Genomic clone 16 DNA.
Lane 5 : Genomic clone 6 DNA.
Lane 6 : Genomic clone 27 DNA.
Lane 7 : *E.coli* strain Q358 genomic DNA.
Lane 8+11 : *R. acidophila* strain 7750 genomic DNA.
Lane 9+10 : *R. acidophila* strain 7050 genomic DNA.
Lane 12 : Genomic clone 13
Lane 13 : Genomic clone 18
Lane 14 : Genomic clone 34
Lane 15 : Genomic clone 4
Lane 16 : *R. palustris* strain French genomic DNA.
Lane 17 : *R. palustris* strain DSM genomic DNA.
Lane 18 : *R. sphaeroides* strain 2.4.1 genomic DNA.

The gel was blotted and the filter probed with radiolabelled pLHIISB18. The resulting autoradiograph is shown in Figure 10.3b.

Note panels A and B have been exposed for 18 hours and 5 days respectively.

E. coli strain Q358 was also used, as it was considered possible that it may have contaminated the vector and was hybridising positively with the probe. All of these bacteria were plated for single colonies, then used in batch culture for DNA isolation. Representative clone and genomic DNAs were digested with EcoRI+SalI, run on a gel, blotted and probed with radiolabelled pLHIISB18 DNA. The resulting gel photograph and filter autoradiograph are shown in figure 10.3.

Discussion

The autoradiograph in figure 10.3 suggests that the positively hybridising clone fragments from groups 1-3 are comparable in size to those found in the genomic DNAs of *R. acidophila* strains 7050 and 7750. This evidence suggests that the library may have been constructed from a mixture of insert DNAs from these two strains.

In addition, there is a greater number of positively hybridising genomic bands than have been isolated as clones. This suggests that the library may be incomplete and/or that partial digestion of the genomic DNA has occurred. These points will now be dealt with in turn.

Clone 1, a representative of group 1, and clone 15, a representative of group 4, have positively hybridising bands which can be accounted for, in terms of size, with those found within the 7050 genome. However, in the case of clone 1 the ratio of the signal strength between the positivly hybridising 5kb and 1.9kb bands cannot be accounted for by the hybridisation signal given by strain 7050 genomic DNA. Only when the positively hybridising bands of strain 7750 genomic DNA are examined can the positively hybridising bands of clone 1 be fully accounted for both in terms of size and relative signal strength.

It should be pointed out that initially clone 1 was used as the representative from group 1 (rather than clone 9) for sequencing.

It should be pointed out that initially clone 1 was used as the representative from group 1 (rather than clone 9) for sequencing. However, the nucleotide sequences from this clone translated into, for example, 15 amino acids of a light harvesting polypeptide, then "nonsense" primary sequence was revealed. This suggested that the insert of this clone may have undergone some form of rearrangement. The positively hybridising bands (especially the faintest one, arrowed) therefore, may not be representative of the genomic state.

Therefore, in terms of size all of the genomic clones have positively hybridising bands with counterparts in the genome. Though it would seem that in terms of intensity the bands in clone 1 cannot be accounted for in full. However, this method of looking at single clones in isolation, and comparing the intensity of its positively hybridising bands to those found within the genome may be flawed. My argument to support this is as follows. Let us assume all of the clones shown have arisen from the one organism and that clone 27 is representative of the *in vivo* state. Then as the positively hybridising band of this clone is 1.9kb, the signal from clone 27 should be added to the signal from the 1.9kb band of clone 1 before the comparison to the genomic bands are made. Therefore, in order to make comparisons of intensity ratios there must be complete confidence that bands of the same size but from different genomic origins do not overlap. If this does occur then the signal strength of the positively hybridising bands within the genome is the sum of the signal strength of all of the positively hybridising bands (of the same size) within the clones.

In addition to checking for contamination from *R. acidophila* strain 7750 the possibility of contamination from *Rsp. rubrum*, *R. palustris* strains French and DSM, *Rb. sphaeroides* strain 2.4.1 and *E. coli* strain Q358 was also investigated.

As expected *Rsp. rubrum* DNA did not hybridise positively with the pLHIISB18 probe as it lacks LHII type complexes. There was also an apparent lack of signal from *E. coli* strain Q359. *R. palustris*

strain French did show a positively hybridising band, however, this did not coincide with any of the positively hybridising bands from the clones. It should be noted that panels A and B on the autoradiograph in figure 10.2 recieved different exposure times. On longer exposure to that shown additional faint positively hybridising bands showed up in the *R. palustris* strain French lane. However, all of these bands were smaller than those of the clones. This longer exposure has not been shown. The *R. palustris* strain DSM DNA has not cut in this experiment. A subsequent experiment showed that two faint positively hybridising genomic bands were present, though neither of them coincided with positively hybridising bands from the genomic clones. *Rb. sphaeroides* strain 2.4.1 hybridised strongly with the probe. This was to be expected as it was from this organism that the probe was derived. It can be seen however that the positively hybridising banding pattern does not coincide with those of the clones. This combined evidence suggests that the library was not contaminated with DNA from other bacterial species.

Perhaps the most surprising feature of the autoradiograph is the eleven positively hybridising bands found within the strain 7050 genomic DNA tracks. Spectrophotometric studies found only three types of antenna complex in this organism [Cogdell *et al.*, 1983]. However, if the assumption is made that one band equals one complex then this suggest that there are considerably more complexes than initially envisaged.

It could be argued that many of the bands are the result of internal clevage of genes. However, even if it was assumed that all genes had been cleaved once, this still suggests that five different types of genes encoding these complexes are found. Such a model does not take into account possible clustering of genes such as the three alpha-beta gene pairs found in clone 9, as described in the previous chapter. If this is accounted for eleven different types of antenna complex may be an under estimate! This conjecture of course assumes that the genomic DNA is not partially digested and that the probe hybridises only to light harvesting genes.

If the DNA had been digested fully then this suggested that there were more positively hybridising genomic bands than could be accounted for by the clones. This suggested that either the library was incomplete, or that the screening had not been carried out extensively enough to allow all types of positive clone to be found. As ≈22,000 plaques had been screened then the second possibility was considered unlikely. Instead it was thought that host restriction may have occurred in a non-random fashion eliminating certain groups of recombinant clones. This may have accounted for the lower than expected titre of the unamplified library, see chapter 5.

In order to check that partial digestion of the DNA had not occured and that the relative intensities of the clone bands were representative of the genome the following experiment was carried out.

Rechecking the source of the genomic clones

The previous experiment was repeated using only genomic DNAs from strains 7050 and 7750. Fresh genomic DNA samples were prepared from each strain. The clone and genomic DNAs were digested with EcoRI+SalI as before then ethanol precipated. 50ngs of clone and 5µgs of genomic DNA were loaded onto the gel, see chapter 3.8. The idea behind this was to have approximately the same amount of DNA in each band in both the genomic and clone tracks. This would allow more direct comparisons of positively hybridising band intensities to be made and avoid the neccesity of cutting up the filters for autoradiography due to the "overloading" of clone DNA. In addition each clone DNA would be in a lane adjacent to both 7050 and 7750 genomic DNA. This would allow a better comparison of the positively hybridising band sizes to be made. The gel was run, blotted and probed as described previously. The gel photograph and autoradiograph are shown in figure 10.4.

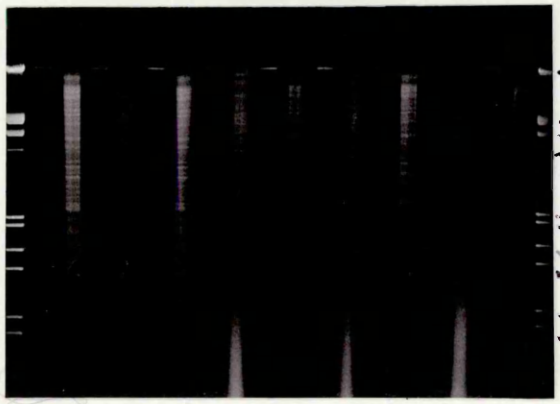


Figure 10.4a

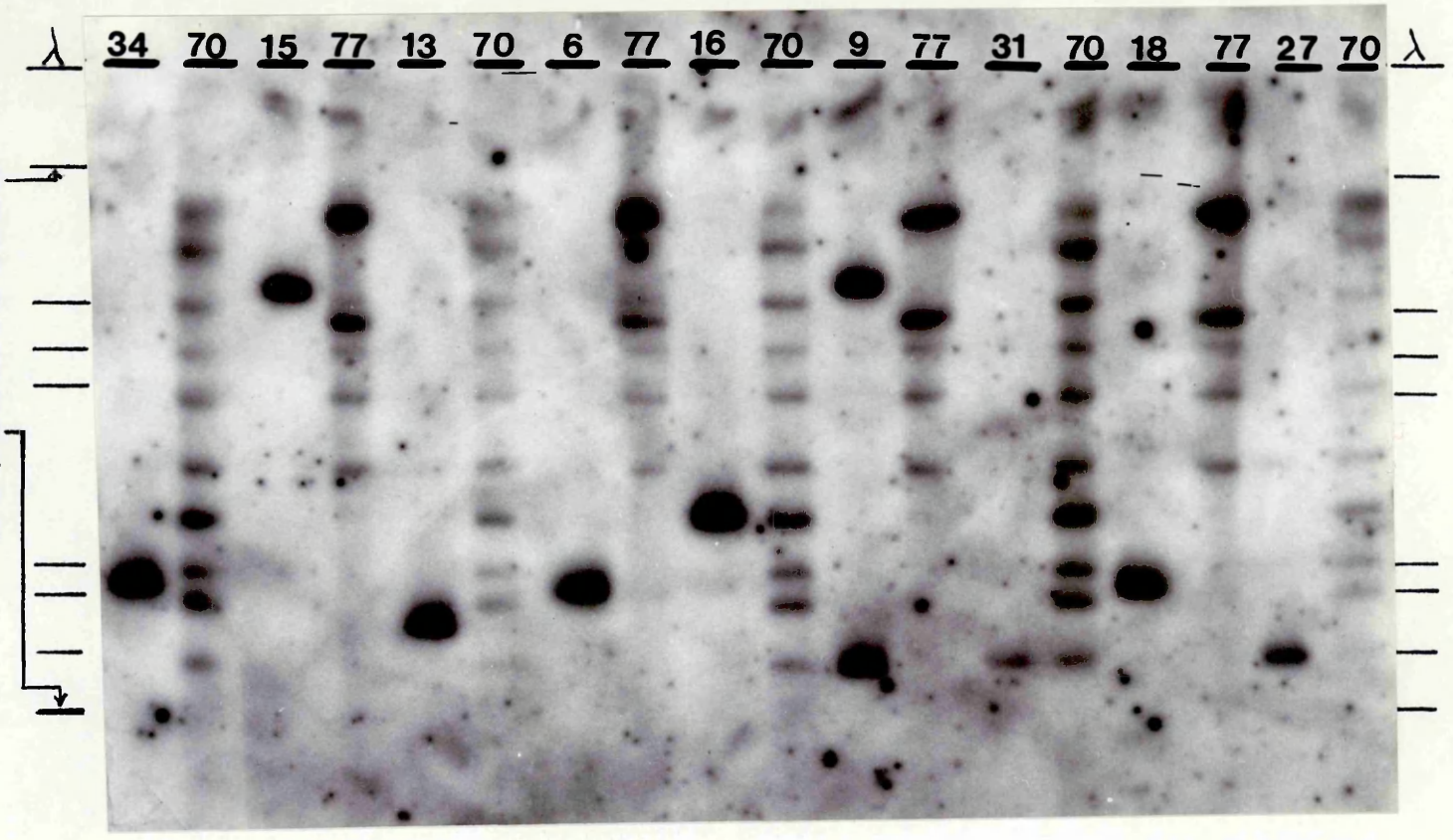


Figure 10.4b

Figure 10.4

Genomic DNAs from *R. acidophila* strains 7050 and 7750, and representative clone DNAs from Groups 1 - 4 were digested with EcoRI+SalI. The DNAs were ethanol precipitated, and sufficient DNA, 50ngs of clone, and 5µgs of genomic, loaded into the wells to give approximately the same number of DNA fragments in each DNA band. The gel photograph is shown opposite, Figure 10.4a.

Lanes 1+20: cI857 DNA cut with EcoRI+HindIII.
Lanes 3,7,11,15, & 19: 5µgs of *R. acidophila* strain 7050 genomic DNA.
Lanes 5,9,13, & 17: 5µgs of *R. acidophila* strain 7750 genomic DNA.
Lane 2: genomic clone 34.
Lane 4: genomic clone 15.
Lane 6: genomic clone 13.
Lane 8: genomic clone 6.
Lane 10: genomic clone 16.
Lane 12: genomic clone 9.
Lane 14: genomic clone 31.
Lane 16: genomic clone 18.
Lane 18: genomic clone 27.

The gel was blotted and the filter probed with radiolabelled pHHIISB18. The autoradiograph is shown opposite, Figure 10.4b.

λ = cI857 DNA EcoRI+HindIII "markers".
70 = *R. acidophila* strain 7050 genomic DNA
digested with EcoRI+SalI.
77 = *R. acidophila* strain 7750 genomic DNA
digested with EcoRI+SalI.
All other numbers are clone numbers shown in Figure 10.4a.

From the autoradiograph it can be seen that only clones 16, 31 and 27 have positively hybridising bands with counterparts in the strain 7050 genomic tracks. The other clones have bands which are close to, but not the same as, the bands in the 7050 genomic tracks. However, clones 34, 13, 6 and 9 and 18 have bands which are closer in size to those of strain 7050 than 7750. The positively hybridising band of clone 15 suggests this clone could have arisen from either strain.

However, closer examination of the data suggests that differential migration of the various DNAs has occurred. The positively hybridising bands of clone 9 for example have sizes of 5 and 1.3kb, rather than the 5 and 1.9kb expected from previous gels. Clone 34 has a positively hybridising band of ≃1.9kb rather than 2.3kb found previously. Whereas the positively hybridising band of clone 16 is ≃2kb, the size expected from previous work.

The only explanation for this anomolous migration behaviour is that through the ethanol precipitation differing amounts of salt have remained in the DNA solutions. This could lead to differences in the migration rate of the DNAs.

This perhaps can be illustrated further by examining the banding pattern of the genomic DNAs. All tracks of strain 7050 genomic DNA have the same postively hybridising banding pattern. This was to be expected as these DNAs were ethanol precipitated together then aliquoted into the gel tracks. The same method was also used for strain 7750 genomic DNA.

Now compare the banding patterns of the two types of genomic DNA relative to one another. Recall that in the previous experiment clone 1 had a banding pattern that could be found in the DNA from both strains, figure 10.3. This similar banding pattern has now gone. This is particularily noticable if the genomic 7050 track adjacent to clone 16 is examined. In this track there is a positively hybridising band of the same size as that found in clone 16. Below this band are two smaller bands of 1.6-1.9kb. These bands

were also found in strain 7750 in the previous experiment, see figure 10.3. However, these bands are not evident in the 7750 tracks in this experiment, figure 10.4.

Although this result does not show that the clone inserts were derived from strain 7050, it does suggests that the number of positively hybridising bands found within the genomic DNA were not a result of partial digestion.

As the quality of the data are so poor, there is as yet no definitive evidence to show that the genomic clones contain inserts of *R. acidophila* strain 7050 DNA. This clearly is an area which needs to be clarified before further work on these clones is undertaken. However, from the data available, (see figure 10.3) there is no evidence to suggest that the library has been constructed from DNA from any other strain.

CHAPTER 11 DISCUSSION

The aim of this work was to clone the light harvesting genes of *R. acidophila* strain 7050 . In this final chapter I would like to examine how successful this work has been in achieving that objective.

The data presented previously suggests that light harvesting genes from *R. acidophila* strain 7050, have been cloned and their nucleotide sequences determined. Evidence to support this is as follows. Firstly, recombinant lambda and mp19 clones were isolated on the basis of positive hybridisation to pLHIISB18 which encodes the B800-850 polypeptides of *Rb. sphaeroides*. In addition, the nucleotide sequences determined subsequently showed considerable identity to the nucleotide sequence of pLHIISB18 using Bestfit analysis. Translation of these sequences, generated amino acid sequences which were shown to have considerable identity to the light harvesting polypeptides of *R. acidophila*. These sequences encoded the polypeptides in the order beta-alpha, the same order found in all other light harvesting genes examined by other workers. All of the polypeptides have the initiation codon ATG, upstream of which lies a presumptive ribosome binding site. Each polypeptide is terminated by TAA or TGA.

In this chapter I have looked more critically at the data and asked, and hopefully answered, four fundamental questions: a) if the presumptive genes are transcribed can they also be translated ie. how good are the Shine-Delgarno sequences? b) are the translated products likely to lead to functional light harvesting polypeptides? c) what are the possible roles of the carboxy-terminal extensions? and d) are these genes encoded in the genome of *R. acidophila* strain 7050?

Transcription

It would be fair to begin by asking why make the assumption that the genes are transcribed? After all, a common working definition of a gene is; a stretch of DNA that is transcribed into the RNA which codes for a single polypeptide chain [Alberts *et al*, 1983]. Why therefore not start at the level of transcription?

Two main problems were associated with determining the theoretical starting point of transcription. Firstly only in consensus sequences 16, 9B and 9C were there sequences available for analysis 5' to the beta polypeptide initiation codon. Secondly the sequences which were available did not extend that far upstream having lengths of 185, 195 and 76bp respectively.

It had been shown that in *Rb. sphaeroides* the promoter for the *puc* genes lay within a region 211bp upstream of *pucB (-211bp)*. It had also been suggested (from comparison of the *Rb. sphaeroides* and *Rb. capsulatus* nucleotide sequences) that the promoters could be localised to a region of homology between -96 to -165bp upstream of the *puc*B genes in *Rb. sphaeroides* [Kiley and Kaplan, 1987]. If these sequences were genuine promoter sequences, and the spacing between the promoter and the structural genes was critical to function, then immediately the upstream region of consensus 9B was too short (-76bp) to find a promoter of this type.

However, Bestfit analyses of the regions 5' to the presumptive *puc*B genes of sequences 16, 9B and 9C were made to the sequences upstream of the *puc*B genes from *Rb. capsulatus* and *Rb. sphaeroides* strains 2.4.1 and NCIB 8253 [Ashby *et al*, 1987; Kiley and Kaplan, 1987 and Youvan and Ismail 1985]. Note the upstream sequences of two *R. sphaeroides* strains were used in these analyses as their upstream regions have different sequences.

The results of these Bestfit comparisons suggested that there was no significant alignment between the presumptive and native upstream

sequences. The Bestfit analyses were then repeated using only the regions from -96 - -165 and -99 - -174bp upstream of *pucB* in *Rb. sphaeroides* 2.4.1 and *Rb. capsulatus* respectively. Again the results suggested that there was no significant alignment between the sequences. Finally it was noted that between the *Rb. spaheroides* strain 2.4.1 and *Rb. capsulatus* that there was a relatively well conserved piece of sequence in the promoter region.

CCCATAnTGCGnnnnnCnnGGnCGGATCA

The bases in upper case letters were conserved, the n's, denoting differences between the sequences. This sequence was used in Bestfit analyses as a "probe" for promoter sequences in the region 5' to the presumptive *puc*B genes. As a control, the region 5' to the *pucB* genes of *Rb. sphaeroides* and *Rb. capsulatus* were also "probed". In addition, the weighting of the Bestfit scoring system was adjusted as follows: any gaps = 0, average mismatch (to a conserved base) = -0.5, average match (to a conserved base) = 1.0, and match of n to any base = 0. This effectively scored in the favour of sequence alignment of the conserved bases to any other similar sequence because; a) mismatches had half the (opposite sign) value of matches and b) gaps in the alignment and matches to n were ignored. Using this system the control promoter sequences were perfectly matched. However, even under such non-stringent matching conditions, a significant alignment between the probe and the region 5' to the presumptive *puc*B genes could not be found.

Bestfit analysis was also carried out between the upstream sequences and the sequence encoding *puf*Q and its oxygen regulated promoter [Adams *et al*, 1989] in the hope that an analogous system of regulation may exist. Unfortunately however, alignments between these sequences were also insignificant.

The reason for this lack of alignment was considered to be one of the following. Either the presumptive genes were not transcribed from a promoter like those of *Rb. sphaeroides* and *Rb. capsulatus*, or they were transcribed from promoters of this type but they lay toofar upstream ie. in the regions which had not been sequenced , to

be found by the Bestfit analysis. Clearly the second hypothesis could not be tested with the available sequence data. However, testing the first was possible by carrying out Bestfit analyses between the sequences 5' to the the presumptive *puc*B genes. The result of this suggested that considerable similarity existed between the sequences eg. an alignment showing 64% identity, extending over 80bp, was found between consensus 16 and 9C. These similarities can be further illustrated by comparing the 53 bases upstream of the presumptive *puc*B initiation codon of the three consensus sequences, see figure 11.1.

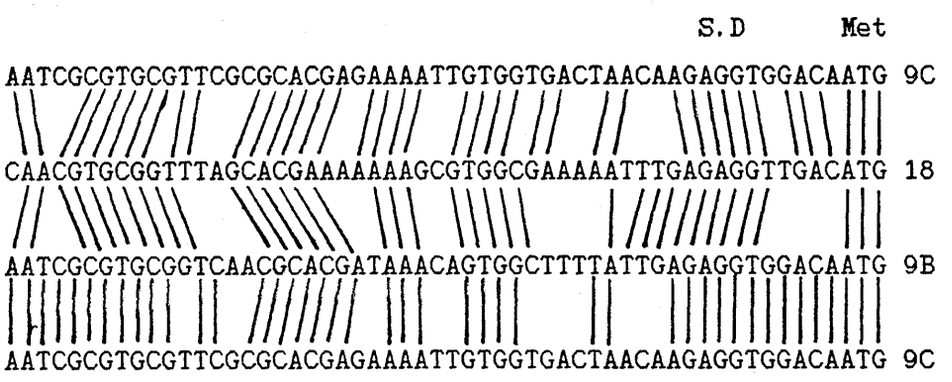


Figure 11.1

Bestfit analysis of the 53bp upstream regions of consensus sequences 18, 9B and 9C. The three sequences are aligned at the presumptive initiation codon ATG which codes for Met. The Shine-Delgarno sequences are also shown at S.D. Bestfit has been used to compare each of sequence to the other two. The Bestfit alignments have then been transferred into the form shown in which a vertical line represents matching bases (at the same position in two sequences) and a sloping line represents matching bases at a different position, (the equivalent of an up or downstream gap in the alignment). Note sequence 9C has been presented twice for ease of comparison.

From this figure it can be seen that there appears to be considerable sequence similarity within the regions shown. Indeed,

between sequences 18 and 9C this level of similarity extends to over 120bp upstream of the initiation codon. This degree of similarity suggests that for some reason a relatively high level of sequence conservation exists immediately upstream of the presumptive *puc*B genes. As yet the function of these sequences is unknown, however, possiblities clearly include promoter or regulatory sequences.

Translation

Translation initiation requires an mRNA molecule to possess features which allow recognition by the ribosome. Do these features occur within the presumptive *puc* gene transcripts?

When Shine and Dalgarno sequenced the 3' end of 16S rRNA, they found that it contained nucleotides complementary to each of the then known translational initiation regions, just 5' to the initiation codon [Shine and Dalgarno, 1974]. They hypothesised that base pairing between the mRNA and rRNA was an important step in the ribosomal selection of translation initiation sites. There is considerable evidence to support this hypothesis [Gold *et al.*, 1981; Grunberg-Manago, 1980; Steitz, 1979].

The "classic" translational initiation site has two main features, the initiation codon (which may be either AUG or more infrequently GUG) and ≈7bp upstream the polypurine sequence AGGAGGA (or variation of it) now known as the Shine-Dalgarno (S.D) sequence. Shown overleaf in figure 11.2 are the possible translation initation sites of the presumptive polypeptides.

From this figure above it is clear that none of the presumptive S.D sequences are identical to the "classic" S.D sequence, though they are all a variation of it. However, are the sequences close enough to allow translation?

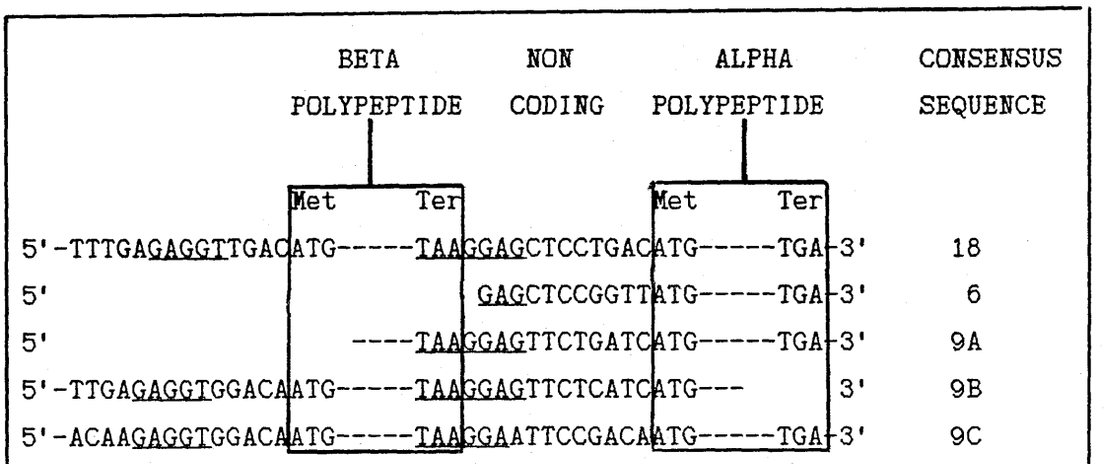


Figure 11.2

Met = Initiation codon; Ter = termination codon;
ATG---TG/AA = either a beta or alpha polypeptide.
Possible Shine Delgarno sequences are underlined
Note that for convienience the mRNA sequences have been written as DNA.

Stormo *et al* characterised the Shine-Dalgarno sequences of 124 gene beginings from *E. coli* and its bacteriophage.. This information was used to devise a set of rules which help distinguish gene beginings from other sites in a library of over 78,000bp of mRNA [Stormo *et al*, 1982]. However, the beginings of the presumptive genes were known ie. ATG. What was required was a method of determining whether or not the S.D sequences were "genuine". To do this I have turned the rules on their head. Instead of of using the rules to predict gene beginings, the gene beginings have been used to determine if the the S.D sequences are in a suitable upstream position and have a suitable sequence to permit translation.

The region 20bp upstream of the initiation codon was searched for sequence which had the best complementarity with the 3' end of 16S rRNA. These sequences are underlined in figure 11.2. Written as the complementary DNA the 16S rRNA which was used in the comparison is:

5'-T A A G G A G G T G A-3'

I then applied the rules for ribosome binding sites to the underlined sequences. Part of the Stormo *et al* rule table is reproduced below:

RULE NUMBER	DESCRIPTION[1] (S.D[2]-SPACE-ATG)	GENES(%) FOUND	NONGENES FOUND
0	--- ATG	124 (100)	2548
2	AGG GGA 3E6N ATG GAG	105 (85)	1419
6	AGGA GGAG 4E5N ATG GAGG	83 (67)	44

1. N is any base, E is any base or no base, eg. 4E5N is 5-9 unspecified bases.
2. S.D is the searched for Shine-Delgarno sequence.

Table 11.1

The rules begin at rule 0, which is the least stringent in its search for gene beginings. Using this rule Stormo *et al* found in their library all 124 genuine genes, but they also found 2548 nongenes. Using this rule all the presumptive genes were considered as true genes. However, many other sites within their sequences would also be considered as gene beginings. Clearly rule 0 is a poor rule for finding gene beginings and predicting S.D sites.

Rule 6 was the second most stringent in its searches for gene beginings. Using this rule Stormo *et al* found 83 (67%) genes in their library but only 44 nongenes, a ratio of ≈2:1. All of the complete presumptive S.D sequences (ie. excluding consensus 6)

comply to these rules suggesting that there is a 2:1 probability in favour of the presumptive genes being geniune genes with genuine ribosome binding sites. I am unaware of any work which has attempted to better this predictive value.

However, it must be remembered that all of these rules were determined from *E. coli* sequences. It is not known how well these rules can be used to predict true gene beginings in other organisms. Indeed the use of the 3' end of the 16S ribosome from *E. coli* to pinpoint the presumptive S.D sequence to allow the use of rules 0-7 is perhaps questionable.

Another flaw in these prediction systems is that they only compare sequence patterns. They take no account of secondary structure within the mRNA and evidence is available which suggests that these structures play an important role in the inhibition of translation [Iserentant and Fiers, 1980 and Fiers, 1979]. Although, there is no evidence to suggest that mRNA structure halts translation, only that it reduces the rate.

Functionality

The question now arises, if the genes are transcribed and translated, are the translation products functional light harvesting polypeptides?

In the following passages I have recapped on the general structural features of light harvesting polypeptides, then determined how well the presumptive polypeptides comply with them. For comparison, the presumptive polypeptide amino acid sequences are shown against those of known B800-850 and B800-820 light harvesting polypeptides, see figures 11.3 and 11.4. The conserved histidine residues at positions 19 and 37 in figure 11.3, and 81 in figure 11.4 have been used as reference points to align the polypeptides [Theiler and Zuber, 1984, Youvan and Ismail, 1985; Tadros *et al.*, 1987].

[1] = Amino acid sequence of consensus 16 beta polypeptide.

[2] = Amino acid sequence of consensus 9B beta polypeptide.

[3] = Amino acid sequence of consensus 9A beta polypeptide.

[4] = Amino acid sequence of consensus 9C beta polypeptide.

[5] = *R. acidophila* strain 7750 B800-850 beta polypeptide amino acid sequence [Bissig *et al*, 1988].
[6] = *R. acidophila* strain 7050 B800-850 beta polypeptide amino acid sequence [Rene Brunisholz, personal communication].
[7] = *R. acidophila* strain 7750 B800-820 beta 1 polypeptide amino acid sequence [Rene Brunisholz, personal communication].
[8] = *R. acidophila* strain 7750 B800-820 beta 2 polypeptide amino acid sequence [Rene Brunisholz, personal communication].
[9] = *R. acidophila* strain 7050 B800-820 beta polypeptide amino acid sequence [Rene Brunisholz, personal communication].

[10] = *R. palustris* strain 1e5, B800-850 beta polypeptide (a) amino acid sequence [Tadros and Waterkamp, 1989].
[11] = *R. palustris* strain 1e5, B800-850 beta polypeptide (b) amino acid sequence [Tadros and Waterkamp, 1989].
[12] = *R. palustris* strain 1e5, B800-850 beta polypeptide (c) amino acid sequence [Tadros and Waterkamp, 1989].
[13] = *R. palustris* strain 1e5, B800-850 beta polypeptide (d) amino acid sequence [Tadros and Waterkamp, 1989].

[14] = *R. palustris* strain French B800-850 (β_1) beta polypeptide amino acid sequence [Evans, 1989].
[15] = *R. palustris* strain French B800-850 (β_2) beta polypeptide amino acid sequence [Evans, 1989].
[16] = *R. palustris* strain French B800-850 (β_3) beta polypeptide amino acid sequence [Evans, 1989].

[17] = *Rb. sphaeroides* strain 2.4.1, B800-850 beta polypeptide amino acid sequence [Kiley and Kaplan, 1987].
[18] = *Rb. sphaeroides* strain NCIB 8253, B800-850 beta polypeptide amino acid sequence [Ashby *et al*, 1987].

[19] = *Rb. capsulatus* B800-850, beta polypeptide amino acid sequence [Youvan and Ismail, 1985].

Figure 11.3 continued.

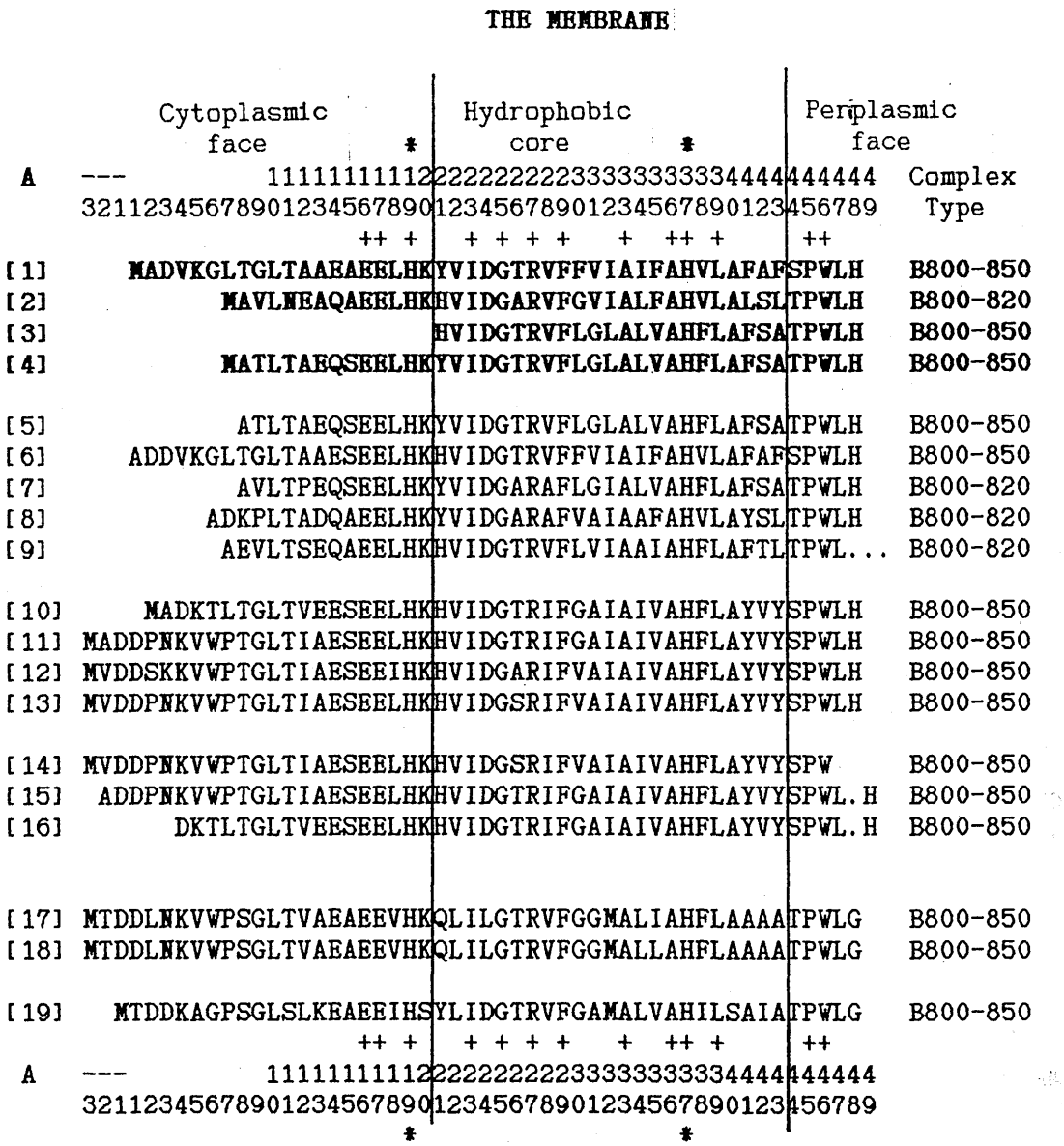


Figure 11.3
This shows a comparison of the amino acid sequences of the presumptive BETA polypeptides to the B800-850 and B800-820 **BETA** polypeptides from a variety of photosynthetic bacteria. The square bracketed numbers refer to the list opposite.

The presumptive polypeptides are presented in bold type. They have been designated B800-850 or B800-820 according to the *R. acidophila* amino acid sequences which their difference totals are "most like", see difference tables in chapter 9.

* = The position of the conserved histidine residues.
+ = A position where there is absolute amino acid conservation.

Note: The amino acid residue positions are presented on line **A**. With the exception of -1, -2 and -3 ie. - - -, these numbers coincide
1 2 3
with those presented in previous figures.

[1] = Amino acid sequence of consensus 16 alpha polypeptide.

[2] = Amino acid sequence of consensus 6 alpha polypeptide.

[3] = Amino acid sequence of consensus 9B alpha polypeptide.

[4] = Amino acid sequence of consensus 9A alpha polypeptide.

[5] = Amino acid sequence of consensus 9C alpha polypeptide.

[6] = *R. acidophila* strain 7750 B800-850 alpha polypeptide amino acid sequence [Bissig *et al*, 1988].
[7] = *R. acidophila* strain 7050 B800-850 alpha polypeptide amino acid sequence [Rene Brunisholz, personal communication].
[8] = *R. acidophila* strain 7750 B800-820 alpha polypeptide amino acid sequence [Rene Brunisholz, personal communication].
[9] = *R. acidophila* strain 7050 B800-820 alpha polypeptide amino acid sequence [Rene Brunisholz, personal communication].

[10] = *R. palustris* strain 1e5, B800-850 alpha polypeptide (a) amino acid sequence [Tadros and Waterkamp, 1989].
[11] = *R. palustris* strain 1e5, B800-850 alpha polypeptide (b) amino acid sequence [Tadros and Waterkamp, 1989].
[12] = *R. palustris* strain 1e5, B800-850 alpha polypeptide (c) amino acid sequence [Tadros and Waterkamp, 1989].
[13] = *R. palustris* strain 1e5, B800-850 alpha polypeptide (d) amino acid sequence [Tadros and Waterkamp, 1989].

[14] = *R. palustris* strain French B800-850 (α_1) alpha polypeptide amino acid sequence [Evans, 1989].
[15] = *R. palustris* strain French B800-850 ($\alpha 2$) alpha polypeptide amino acid sequence [Evans, 1989].
[16] = *R. palustris* strain French B800-850 (α_3) alpha polypeptide amino acid sequence [Evans, 1989].
[17] = *R. palustris* strain French B800-850 (α_4) alpha polypeptide amino acid sequence [Evans, 1989].

[18] = *Rb. sphaeroides* strain 2.4.1, B800-850 alpha polypeptide amino acid sequence [Kiley and Kaplan, 1987].
[19] = *Rb. sphaeroides* strain NCIB 8253, B800-850 alpha polypeptide amino acid sequence [Ashby *et al*, 1987].

[20] = *Rb. capsulatus* B800-850, alpha polypeptide amino acid sequence [Youvan and Ismail, 1985].

Figure 11.4 continued

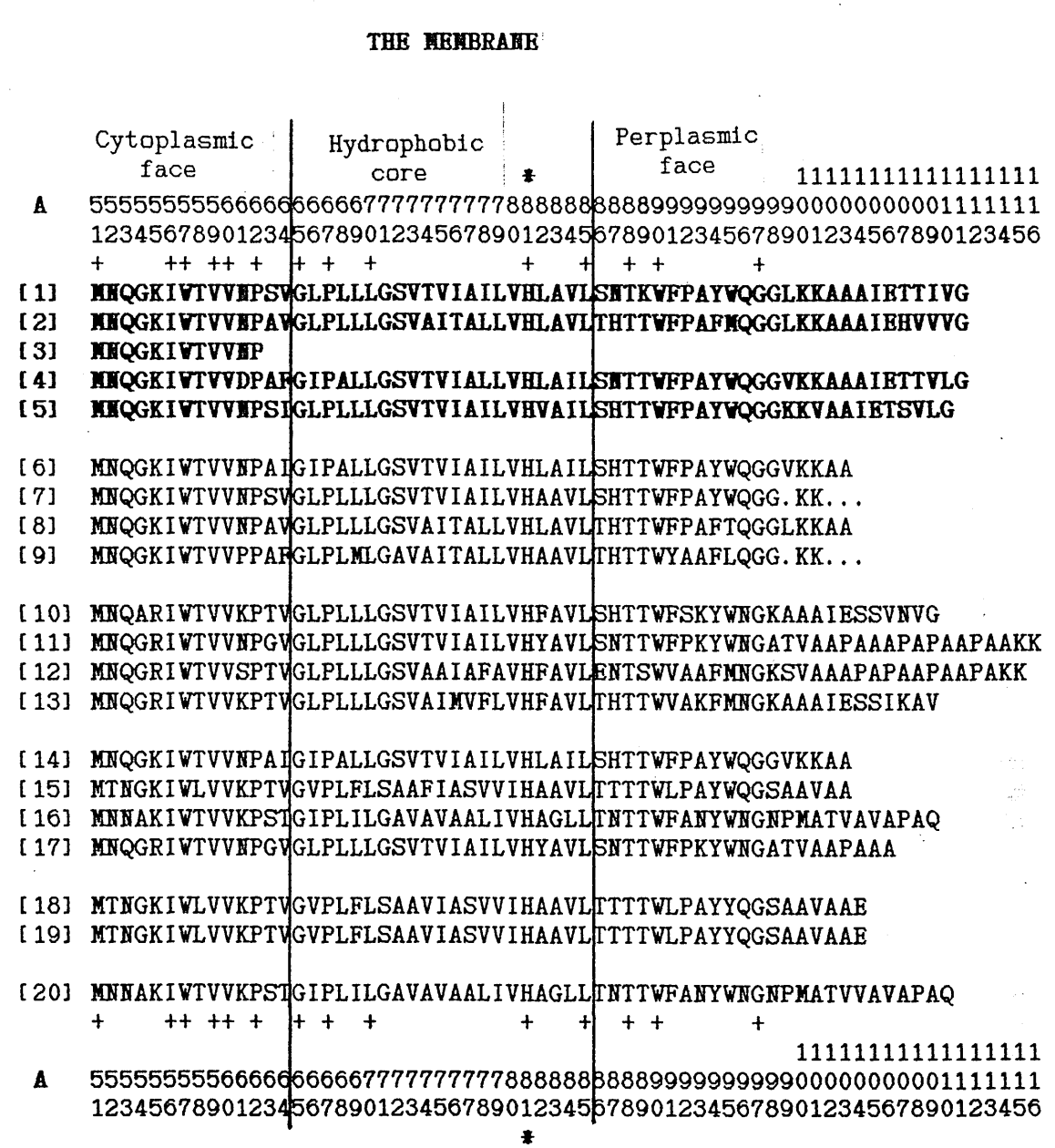


Figure 11.4 This shows a comparison of the amino acid sequences of the presumptive **ALPHA** polypeptides to the B800-850 and B800-820 **ALPHA** polypeptides from a variety of photosynthetic bacteria. The square bracketed numbers refer to the list opposite.

The presumptive polypeptides are presented in bold type. They have been designated B800-850 or B800-820 according to the *R. acidophila* amino acid sequences which their difference totals are "most like", see difference tables in chapter 9. Polypeptides [2], [3], [8] and [9] are of the B800-820 type all other polypeptides are B800-850.

* = The position of the conserved histidine residue.
+ = A position where there is absolute amino acid conservation.

Note: The amino acid residue positions are presented on line **A**. These numbers coincide with those presented in previous figures.

The size of the polypeptides

The number of amino acid residues found in native light harvesting polypeptides has been determined by polypeptide sequencing and deduced from nucleotide sequences to be ≈50 [Zuber, 1986; Cogdell and Thornber, 1980; Thornber *et al.*, 1983].

Two of the three complete presumptive beta polypeptides (ie. excluding polypeptide [3], fig. 11.3) have 41 residues, the remaining one 48. All of the complete presumptive alpha polypeptides (ie. excluding polypeptide [4] fig. 11.4) have 51 residues. These sizes compare favourably with the polypeptides from other organisms. Note these sizes include the initiating methionine and exclude the carboxy terminal residues of the alpha polypeptides after position 104, see later.

Methionine is not found at the amino terminus of any of the native *R. acidophila* beta polypeptides, fig. 11.3, polypeptides [5]-[9]. However it is present at the begining of the alpha polypeptides. This would suggest that N-terminal processing occurs in the case of the beta polypeptides.

The number of residues in the presumptive beta polypeptides if translated then processed would then probably be 40 or 47. The original number of residues in the alpha polypeptides (51) is probably correct.

Note that all of the amino acid sequences in figures 11.4 and 11.3, with the exception of the native polypeptides from *R. acidophila* were obtained from translated nucleotide sequences. Comparison of these sequences with those obtained by protein sequencing (data not shown) suggest that the degree of N-terminal processing of the beta polypeptides is variable among the photosynthetic bacteria.

In *Rb. capsulatus* for example fmet is the amino terminal residue in the polypeptide sequence, suggesting that processing does not occur.

However, in the carotenoidless mutant of *Rb. sphaeroides* ,R26, (sequence not shown) threonine is found at the N-terminus and the polypeptide contains 50 amino acids. In strain 2.4.1 there is a degree of heterogeneity at the N-terminus, with some chains being identical to R26 and others having an additional methionine at the N-terminal position [Theiler *et al*, 1984].

There is no evidence which suggests N-terminal processing of either LHI or LHII alpha polypeptides occurs other than deformylation of the N-terminal formyl-methionine. However, even deformylation is not carried out in all cases, for example f-met is the N-terminal residue of the α-B890 polypeptide in *Rsp. rubrum strain G9 [Brunisholz et al* 1989].

The amino acid composition of the presumptive polypeptides: the hydrophobic core.

All of the light harvesting polypeptides examined have a hydropathy plot which exibits a tripartite character [Brunisholz *et al.*, 1985; Zuber, 1985; Youvan and Ismail, 1985]. There is usually a central hydrophobic core of ≈23 amino acids, while at either end of the polypeptide there are polar amino acids [Brunisholz, 1984].

In figures 11.3 and 11.4 the hydrophobic core is delimited by vertical lines. These limits were taken from those described as the membrane boundary in other works [Zuber 1985, Brunisholz *et al.*, 1985]. From this it can be seen that 23 and 21 residues of the beta and alpha polypeptides respectively lie within this membrane boundary. These numbers of residues are found in both the native and presumptive polypeptides.

The type of residue found at any position within the presumptive polypeptide hydrophobic core can be found at the same position in at least one of the native polypeptides, with two exceptions. In presumptive beta polypeptide [2], position 41, figure 11.3, there is

a leucine residue. Leucine is not found at this position in any other native or presumptive polypeptide. In *R. palustris* and *R. acidophila* strains phenylalanine or tyrosine are found at this position.

The codon for leucine is TTA, therefore a conversion to phenylalanine (TT^{C}_{T}) requires a single base change. Conversion to tyrosine (TA^{T}_{C}) requires two base changes. The R-groups of leucine and phenylalanine class these amino acids as hydrophobic, and that of tyrosine as polar but uncharged. However, the main differences lie in the bulk of the R groups.

Both phenylalanine and tyrosine possess aromatic ring systems not found in leucine. It has been suggested that these ring systems interact with bacteriochlorophyll affecting the absorption maxima of the pigment [Brunisholz and Zuber, 1987]. This can be exemplified by examining positions 94 and 95 of the *R. acidophila* alpha polypeptides [6]-[9], figure 11.4. The B800-850 complexes of strains 7050 [7] and 7750 [6] have identical absorption spectra. When strain 7050 is grown under low light conditions, or strain 7750 at low temperature (25°C), B800-820 complexes are synthesised which have phenylalanine and leucine or phenylalanine and threonine at positions 94 and 95 of strains 7050 and 7750 respectively. As these are the only major differences in the C-terminal residues it has been suggested that the aromatic R groups specifically, are responsible for the differences in absorption maxima of these complexes.

There is also a cluster of aromatic residues at positions 41 and 43 of the beta polypeptides. However, the possession of such residues at these positions does not seem a prerequisite to function. In *Rb. sphaeroides* and *Rb. capsulatus*, for example, B800-850 complexes have alanine at position 41 and 43. Alanine is classed as a hydrophobic residue and possess a small R group $-CH_3$. This suggests that although presumptive beta polypeptide [3] has leucine at position 41

this does not prevent the polypeptide from carrying out a functional role.

Perhaps much more important in this respect is the possession of the residues proline and tryptophan at positions 45 and 46. These have been found in the presumptive polypeptides and in all LHI and LHII antenna polypeptides examined sofar.

The second exception is found in presumptive alpha polypeptide [5], position 82, figure 11.4. Valine (GTT) is found in this position rather than leucine (TT^{A}_{G} or CTN) or alanine (GCN). The change from valine to either of the other two residues would require only a single base change.

Although these two residues are well conserved at this position, a change to valine may not have a considerable functional effect. All three residues are classed as hydrophobic, and all have similar R groups in terms of size and composition ie. alanine $-CH_3$, valine $-CH[CH_3]_2$, leucine $-CH_2CH[CH_3]_2$. They also have relatively similar hydropathy indices: Ala -4.2 (1.8); valine -8.4 (4.2) leucine -10.1 (3.8). Their main differences appear to lie in their ability to form alpha helices, see table 11.2. Both alanine and leucine are considered as strong helix formers, whereas valine is only a helix former [Chou *et al.*, 1978]. However, I think it would be difficult to predict from current knowledge the effect such a change would have on protein function.

Description	Amino Acids
Strong helix formers	Met, Ala, Glu, Leu.
Helix formers	Ile, Trp, Lys, Val, Phe, Gln.
Weak helix formers	His, Asp.
Indifferent	Arg, Thr, Ser, Cys.
Weak helix breakers	Asn, Tyr.
Strong helix breakers	Pro, Gly.

Table 11.2 The bias of amino acids to alpha helix formation.

The amino acid composition of the presumptive polypeptides: the hydrophilic regions.

The hydrophilic sections of the presumptive polypeptides also have a strong resemblance to those of the native polypeptides. The carboxy terminal regions of the presumptive beta polypeptides, for example, all have counterparts in the native polypeptides of *R. acidophila* and/or the polypeptides of other photosynthetic bacteria.

At the beta N-terminal region, all of the presumptive residues at any particular position have a counterpart in a native *R. acidophila* polypeptide, with one exception.

At position 11 of polypeptide [3], figure 11.3, asparagine is found, rather than threonine which is present in all the other polypeptides except that of *Rb. capsulatus*. Asparagine is encoded by AAC, therefore the transition to threonine (ACN) would require a single base change from A to C. Changing to serine, AG^{C}_{T} [17] would also require a single base transition, A to G.

All of these residues are classed as polar but uncharged, all are indifferent to alpha helix formation. The main differences appear to lie in their hydropathy indices Asp 31.0 (-3.5); thr 3.8 (-0.7) and ser 6.3 (-0.8) and the bulk of their side groups. Asparagine has a relatively large amide group $-CH_2CONH_2$, whereas the other two both have smaller aliphatic hydroxyl side chains, thr $-CHOHCH_3$; ser $-CH_2OH$.

The alpha polypeptides are virtually identical in the amino acid sequences of their hydrophilic domains. All the N-terminal regions have 14 residues, and all the presumptive residues at a particular position can be found in a native polypeptide at the same position, with the exception of the aspartic acid residue at position 61, polypeptide [4] figure 11.4.

The aspartic acid residue is found in the presumptive polypeptide where asparagine, proline, lysine and serine are found in the native polypeptides. These latter four amino acids differ widely in all aspects of their biochemistry, for example proline is a strong helix breaker whereas lysine is a helix former. Proline is non-polar, serine and asparagine are polar but uncharged and lysine is polar and positively charged. Aspartic acid is polar, negatively charged and a weak helix former.

Among the alpha polypeptides (between positions 51 and 104) this position (61) is the joint second most variable in its residue composition. As the residues found in the native polypeptides do not seem to conform to any particular pattern it is difficult to interpret any effect on function by having aspartic acid at this position.

At the carboxy terminal end of the alpha polypeptides there may be absolute identity between the residues found in the presumptive polypeptide and those found at the same position in the native *R. acidophila* polypeptides. I use the phrase may be, because at position 99 of presumptive polypeptide [5], figure 11.4, there is a lysine residue which is not found at that position in any of the polypeptides for which sequence at that position is available. However, in *R. acidophila* polypeptides [7] and [9] there are two residues of unknown type at this position. It is clearly possible that they may be lysine residues.

Another structural characteristic of antenna polypeptides is the presence of the sequence AxxxxH, the histidine of which is believed to bind bacteriochlorophyll [Theiler and Zuber, 1984; Youvan and Ismail 1985; Tadros *et al.*, 1987]. Two of these sequences are found in beta polypeptides and one in the alpha.

The correct number of these sequences are found in the correct position within all the presumptive polypeptides with one exception. In presumptive beta sequence [4], fig 11.3, serine rather than

alanine (at position 15) is found five residues upstream of the conserved histidine. However, an identical situation is found in most of the *R. acidophila* and all the *R. palustris* beta polypeptides. Perhaps the use of AxxxxH as a diagnostic structural characteristic is no longer as useful as it once was?

Comparison of the amino acids sequences of the presumptive and native polypeptides has shown that there are considerable similarities between the two groups. This is further illustrated in figures 11.3 and 11.4 by examining positions of absolute residue conservation, marked +. From these it can be seen that there are 13 and 14 absolutely conserved positions (ie. only one residue found) in the beta and alpha polypeptides respectively. There are no instances where a residue is absolutely conserved in the native polypeptides but the conservation is disrupted at that position by the presumptive polypeptides.

C-terminal sequences of the presumptive alpha polypeptides

It has been shown that there is considerable identity between the primary sequences of the native and presumptive polypeptides of *R. acidophila*. However, a major difference exists between their C-terminii. In the native polypeptides the C-terminus is at position ≃104 (figure 11.4) whereas in the presumptive polypeptides there is an extension beyond this position of 6-7 residues.

These extensions occur after an alanine rich region and begin with isoleucine followed by a glutamic acid residue. There are then three or four residues of which at least one is valine, the extention terminating with glycine. If the Kyte and Doolittle indices of the residues in the extensions are summed, the extensions are found to be slightly hydrophilic. However, this seems mainly due to the weighting given to the second residue in the extension,

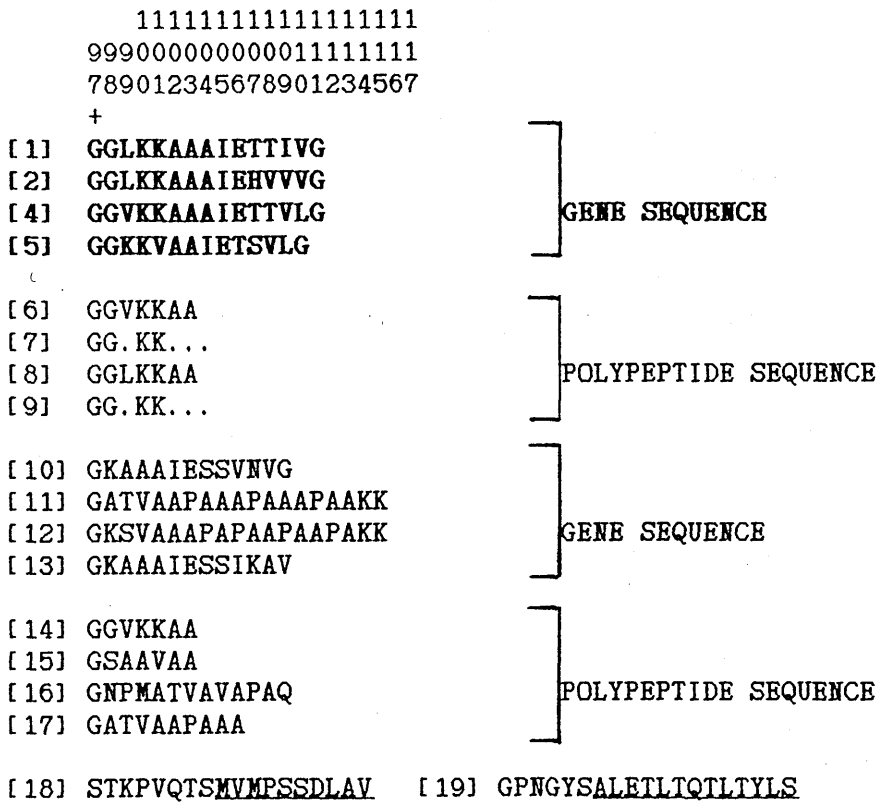


Figure 11.5 The C-terminal Sequences of the Light Harvesting Alpha Polypeptides

With the exception of [18] and [19], the sequences shown correspond in numbering to that shown in figure 11.4.

[18] and [19] are the alpha and beta B880 polypeptides from *Rsp. rubrum*. The underlined regions of these two sequences were not found in the native polypeptides but were found encoded in the structural gene sequences [Berard *et al*, 1986].

[1]-[5] are presumptive alpha polypeptides.
[6]-[9] are native *R. acidophila* alpha polypeptides.
[10]-[13] are *R. palustris* strain le5 alpha polypeptides derived from structural gene sequences.
[14]-[17] are *R. palustris* strain French alpha polypeptide sequences derived by amino acid sequencing.

aspartic acid (+31). This can be seen below where the average hydropathy index of each position has been calculated:

I	E	X	X	X	X	G
-10.5	+31	+6.4	+5.5	-8.6	-9.25	0

+= hydrophilic, - = hydrophobic, X = unspecified residue

Such extensions are not confined to *R. acidophila*. Similar extensions occur in the alpha polypeptides of *R. palustris* and in both the alpha and beta polypeptides of *Rsp. rubrum* [Evans, 1989; Tadros and Waterkamp, 1989; Berard *et al.*, 1986]. These extensions are shown in figure 11.5

Comparison of the extensions of *R. palustris* with those of the presumptive polypeptides show that these extensions are also preceded by an alanine rich region. In alpha polypeptides [10] and [13], figure 11.4, the extensions begin with the residues isoleucine and aspartic acid. The former polypeptide is terminated by glycine as in the presumptive polypeptides. The extensions of polypeptides [10] and [13] are also slightly hydrophilic.

The extensions of polypeptides [11] and [12], however, differ from [10] and [13] in that they are ≃10 residues in length, composed mainly of alanine, proline and lysine residues and are much more hydrophilic ≃ +65 Kyte and Doolittle units [Evans, 1989; Tadros and Waterkamp 1989].

Comparison of the *R. palustris* and presumptive polypeptide extensions with those of *Rsp. rubrum* suggests that there is little similarity between their primary sequences. The extensions found on both alpha and beta B880 polypeptides are overall slightly hydrophilic [Berard *et al.*, 1986].

The light harvesting polypeptides of photosynthetic bacteria are unusual in that they are orientated with their N-terminal and C-

terminal sequences facing the cytoplasm and periplasm respectively [Brunisholz *et al.*, 1984]. In the vast majority of membrane proteins examined this orientation is reversed [Gerber *et al.* 1977; Abdulaev *et al.*, 1978; Walter *et al.*, 1979; Wickner, 1979, 1980; Benson and Silhavy, 1983].

R. acidophila, *R. palustris* and *Rsp. rubrum* are also relatively unusual in possesssing C-terminal sequences which appear to be processed. C-terminal extensions have been shown to occur in only a few proteins, notably proglucagon (Tager and Steiner, 1973; Patzelt *et al.*, 1979) and procollagen [Fesler *et al.*, 1979, 1981]. While the physiological function of the carboxy extension in proglucagon has not been elucidated, in procollagen it appears to play a disulphide-bridging role in the assembly of the procollagen trimer.

C-terminal extensions have also been found in α-lactalbumin in the rat. However in this, and other cases, the additional amino acids do not appear to serve as precursor extensions, and are thought to have been the result of mutation in the termination codon [Prasad *et al.* 1982].

In most membrane polypeptides and polypeptides secreted across membranes, processing occurs at the N-terminus and is believed to be involved in the transportation of the polypeptide partially or completely across the membrane.

Perhaps the best known method by which polypeptide transport may cross the membrane is the "signal hypothesis" first proposed for export of protein across the endoplasmic reticulum [Blobel and Doberstein, 1975]. In this process a protein destined for export or insertion is synthesised as a larger precursor with 15-30 additional amino acids at the N-terminal end of the molecule. This peptide extension, the signal sequence, was proposed to initiate binding of the translation complex to the membrane. This binding results in the formation of a transient pore through which the nascent peptide chain passes as synthesis proceeds ie. co-translational export. The

signal sequence is then removed by a specific protease (signal peptidase) before the synthesis of the protein has been completed.

Though apparently widespread for the exportation or insertion of many membrane proteins, co-translational insertion of the light harvesting polypeptides probably does not occur. The main argument against this method is that the extensions, if they are indeed signal sequences occur at the C-terminal end of the polypeptides. Translation would therefore need to be complete before the signal sequences could be recognised by the signal recognition protein. This therefore suggests that posttranslational rather than co-translational insertion occurs.

Proteins specified by nuclear DNA, destined for localisation in the mitochondria, chloroplast and peroxisome are exported by a different mechanism to that seen with the rough ER. This was first demonstrated with the chloroplast protein ribulose-1,5,-bisphosphate carboxylase [Dobberstein *et al.*, 1977; Chua and Schmidt, 1978; Highfield and Ellis, 1978]. Chloroplast proteins typically are synthesised by cytoplasmic ribosomes as higher molecular weight precursors. They possess N-terminal extensions, termed transit sequences that are both necessary and sufficient for transport [Gregory *et al.*, 1986; Cashmore, 1984,]. These sequences allow the proteins to traverse the chloroplast envelope and insert into the thylakoid membranes ie. posttranslational export [Gregory *et al.*, 1986].

One of the few polypeptides which undergoes C-terminal processing is a chloroplast protein [Marder *et al.*, 1984]. However, this protein is encoded in the chloroplast genome [Grebanier *et al.*, 1978]. The 32-kilodalton protein of photosystem II lies partially exposed at the stromal surface of the thylakoid membrane and is one of the major products of light dependent protein synthesis [Eaglesham and Ellis, 1974]. However, it does not accumulate due to its rapid turnover and represents less than 1% of steady state thylakoid proteins. The 32-kDa protein is suggested to have an important

function on the reducing side of photosystem II [Mattoo *et al.*, 1981, 1984].

In the plant, processing was shown to be a posttranslational event occurring on the photosynthetic membranes [Edelman and Reisfield, 1980; Grebanier *et al.*, 1978]. The 32-kDa protein is the only membrane protein encoded within the chloroplast genome for which a precursor form has been absolutely established [Marder *et al.* 1984]. The C-terminal extension of this protein is 12 residues in length and interestingly the last 10 residues give the sequence:

A A I E A P S T N G.

Which is similar to: A A I E X X X X G the sequence of the C-terminal extensions found in the presumptive polypeptides. The relevance, if any, of these similarites is not clear.

Although there are similarities in that the C-terminii of the presumptive and 32-kDa protein are processed and face out of the photosynthetic membrane there are probably major differences in their mode of insertion and overall topography. The 32-kDa polypeptide does not, as in the antenna polypeptides, span the membrane, but sits partially buried in it with its C-terminus exposed to the stroma. As the polypeptide is synthesised in the stroma, this suggests that the N-terminal sequence is responsible for leading the insertion into the membrane, leaving the exposed C-terminus. The C-terminus therefore does not pass through the membrane as in the antenna polypeptides. Topologically, this would be equivalent to the antenna polyeptides being synthesised in the periplasmic space then being inserted, N-terminal end leading, into the photosynthetic membrane.

The model for the maturation and insertion of the 32-kDa into the membrane is based on the membrane trigger hypothesis of Wickner (1979) [Marder *et al.*, 1984]. Which in turn is based on the method of insertion into the membrane of the thoroughly studied major coat protein of f1 (M13). This protein exists transiently in the inner membrane before incorporation into phage particles. When present in

the inner membrane it spans the membrane once with its N-terminal sequence facing out [Wickner, 1979].

Wickner *et al* suggested that a leader (signal) peptidase (which cleaves a leader sequence on the coat protein) is the only cellular component required for coat protein localisation in the membrane. To account for the postranslational export mechanism, Wickner (1979) proposed that the signal sequence promotes folding of the the completed precursor into a soluble export-competent conformation. Upon exposure to a hydrophobic environment, the protein is triggered into a different conformation which allows spontaneous membrane insertion. Cleavage of this sequence would drive the reaction and make it irreversible [Wickner 1979; Watts *et al.*, 1981].

The membrane trigger hypothesis has found favour with a number of workers and has been implicated in the insertion of the B880 complex of *Rsp. rubrum* and the B800-850 complex of *Rb. capsulatus* [Berard *et al.*, 1986; Dierstein, 1984]. In the case of Berard *et al.* insertion by a co-translational mechanism was also envisaged. However, they failed to describe how the carboxy terminus of the polypeptide could enter the membrane first if it is translated last.

Dierstein (1984) on the other hand uses only part of the trigger hypothesis. As *Rb. capsulatus* B800-850 polypeptides lack a cleavable leader sequence it is proposed that on contact with the membrane the polypeptide undergoes a conformational change and is inserted. However, it is not explained how the polypeptides remain within the mebrane without the removal of a cleavable leader to stabilise the insertion. It would however, be reasonable to argue that the tripartite character of the polypeptide itself would be sufficient to stabilise the insertion by acting like a "halt or stop transfer sequence". Such sequences are found in a range of membrane polypeptides and as their name suggests they prevent the entire polypeptide from being transferred across the membrane. The hypothesis of Dierstein also does not include a role for the 14-kDa colourless polypeptide which is believed to be involved in the

antenna assembly and insertion process [Drews and Feick, 1978]. A similar protein, 15A has been implicated in the assembly of the B800-850 antenna complex of *Rb. sphaeroides* [Kiley and Kaplan, 1988].

The question therefore arises, what is the role of the carboxy terminal extensions which are found in *R. acidophila* and *R. palustris*?

It is interesting to note that the C-terminal regions beyond the membrane interface of all the alpha polypeptides (shown in figure 11.4) are between 14 and 26 residues longer than the corresponding C-terminii of the beta polypeptides (figure 11.3). This suggests that a long C-terminus is not required for insertion of the beta polypeptides.

However, it may be that insertion of a beta polypeptide occurs only after it has combined with an alpha polypeptide, (perhaps after Bchl has been bound to the conserved histidine residues) through the hydrophobic attraction of the alpha helices. This sharing of a C-terminus leader may then allow insertion of the heterodimer into the membrane. This may also account, in part, for the 1:1 ratio of the two polypeptides found in the minimal unit of these complexes.

In addition, having the long carboxy terminus on the alpha polypeptide would also allow time for the beta polypeptide to associate with the newly forming alpha polypeptide. If the beta polypeptide had the long carboxy terminus it is conceivable that it could be inserted into the membrane without "waiting" for the translation of an alpha polypeptide. The carboxy terminal extension may then assist conformational changes in the polypeptides and allow them to enter the membrane as in the membrane trigger hypothesis. However, it may also interact with an as yet unknown membrane protein which could facilitate positioning and insertion.

Cleavage of the extension sequence may then stabilise the polypeptides within the membrane. However, once spanning the membrane, if the polypeptides are in their energetically most stable state, the subsequent cleavage may only be a method of recycling the extensions which are no longer required once insertion has occured.

This hypothesis does not preclude the existence of proteins such as 15A or the 14-kDa polypeptide which may assist in the insertion process. It merely suggests that long C-terminiii on the alpha polypeptide are required for insertion and that in some species ie. *R. acidophila*, *R. palustris*, and *Rsp. rubrum* these terminii are removed after insertion. It should also be noted that these three organisms may assemble and insert their antenna polypeptides by a completely different mechanism to *Rb. sphaeroides* and *Rb. capsulatus* and that it may be unwise to generalise too much on the basis of polypeptide similarities. However, which parts, (if any) of this hypothesis are correct could be tested experimentally, relatively easily by site-directed mutagenesis.

The genes sequenced are from *R. acidophila* strain 7050.

Combined Southern blot analysis of clone and genomic DNA was used to verify the source of the genomic clone inserts. The result (figure 10.3) suggested that all of the positively hybridising bands found in the genomic clones could be accounted for in the genomic DNA of *R. acidophila* strain 7050. However, as the evidence was not as absolutely conclusive as was hoped I would like to introduce evidence from the method of bacterial culture and from the polypeptide sequences which suggests that the inserts were derived from strain 7050.

Bacterial culture

There are two pieces of evidence which suggest strain 7050 was cultured for the isolation of genomic inserts. Firstly, the cells were grown under a low light regime at 30°C which gave them a deep red-purple colouration. Of the species of bacteria being cultured in the laboratory at that time this colour was unique to this strain.

Of the other species of bacteria placed under this low light regime only *R. palustris* (strains French and DSM) appeared to thrive. The other species tended to form a sediment on the bottom of the culture bottle after a few days. However, although *R. palustris* grew well under these conditions, the colouration of the culture was still significantly different to that of *R. acidophila* strain 7050.

R. acidophila was always cultured in Pfennig's medium which has a pH of 5.2. This medium selects for acid loving bacteria. *R. palustris* was normally grown in C-succinate medium pH 6.8. Transfer of *R. palustris* to Pfennig's medium appears to result in cessation of cell growth [personal observation].

It is possible however, that the original starter culture (*R. acidophila* strain 7050) had become contaminated with strain 7750. If this were the case, strain 7750 would have been capable of growth in the selective medium. However, it probably would have grown more slowly than strain 7050 because at 37°C it does not appear to generate low light complexes. The more rapid growth of strain 7050, and hence more rapid increase in its colour intensity may have masked the presence of strain 7750.

The combination of low light conditions and a low pH selective medium suggest that the cells cultured for the construction of the library were those of *R. acidophila* strain 7050. However, the presence of strain 7750 cannot be completely excluded.

The nucleotide and primary amino acid sequences of the polypeptides

With the exception of *R. palustris* strain DSM, primary amino acid sequences of the light harvesting polypeptides were available for all other species used in the genomic zoo blot, see figure 11.3. These sequences were compared to those of the presumptive polypeptides and a search for regions of dissimilarity made. These comparisons suggest that the presumptive polypeptides have greatest primary sequence similarity with the polypeptides of *R. acidophila*.

The results of these comparisons are discussed below. For brevity only the major differences have been cited.

Comparison of the presumptive polypeptides with the B880 polypeptides of *Rsp. rubrum* strain S1.

Rsp. rubrum strain S1 contains the one type of antenna complex ie. LHI, B880 which is encoded by a single copy of the alpha and beta genes [Berard *et al.*, 1986]. Although it seemed unlikely that lambda recombinants containing LHI genes had been selected with pLHIISB18 this comparison has been made as an additional check.

The main differences found were:

a) The B880 beta polypeptide has a carboxy terminal extension of 19 residues. C-terminal extensions were not found in the presumptive beta polypeptides.

b) Between positions -3 - 10 and 21 - 27 (figure 11.3) the residue found at any position within the B880 beta polypeptide lacks identity with residues at the same position within any of the presumptive beta polypeptides.

c) The N-terminus of the B880 alpha polypeptide has two residues less than the presumptive alpha polypeptides.

d)The sequence QGG (positions 96, 97 and 98) is well conserved in the presumptive and *R. acidophila* polypeptides, however it is not found in the B880 alpha polypeptide.

Comparison of the presumptive polypeptides with the B800-850 polypeptides of *Rb. sphaeroides* strain 2.4.1.

The main differences found were:
a) The sequence $^{H}_{Y}$VID is found at positions 21-24 in the presumptive beta polypeptides. In the beta B800-850 polypeptide of *R. sphaeroides* the sequence QLIL is found, figure 11.3.

b) The carboxy terminal residue of the presumptive beta polypeptides is histidine. In the beta B800-850 polypeptide of *R. sphaeroides* the terminal residue is histidine, figure 11.3.

c) The sequence LLGSV is found at positions 69-73 in the presumptive alpha polypeptides. Sequence LFLSA is found at these positions in the B800-850 alpha polypeptides of *R. sphaeroides*, figure 11.4.

d) All of the presumptive alpha polypeptides have sequences which extend beyond position 104. However, this is the C-terminal position for the B800-850 alpha polypeptides from *R. sphaeroides*.

Note the LHI polypeptides (B875) from this strain were also compared to the presumptive polypeptides. The differences found were even greater and have not been presented.

Comparison of the presumptive polypeptides with the LHII complexes from *R. palustris* strain French.

Finding differences between *R. palustris* strain French and the presumptive polypeptides was more difficult than for any other species. This difficulty arose due to the similarity between the *R. palustris* and *R. acidophila* polypeptides. This can be exemplified by examining alpha polypeptide sequences [6] and [14] from these two

organisms, figure 11.4. It can be seen that these two sequences are identical!

This clearly creates a problem. If *R. palustris* and *R. acidophila* sequences cannot be resolved from one another then it is impossible to resolve *R. palustris* and presumptive sequences even if the presumptive sequences arose through cloning *R. acidophila* DNA. Fortunately the sequences exemplified above are extreme in their identity and the following differences were found.

a) In the beta polypeptides of both strains of *R. palustris* shown ie. [10]-[13] and [14]-[16] figure 11.3, isoleucine is absolutely conserved at position 28. In both the presumptive and *R. acidophila* polypeptides either valine or alanine are found at this position.

b) In the beta polypeptides of both strains of *R. palustris*, alanine is absolutely conserved at position 31, figure 11.3. With the exception of *R. acidophila* polypeptide [8] valine or glycine are found at this position in both the presumptive and *R. acidophila* beta polypeptides.

c) In the beta polypeptides of both strains of *R. palustris* the sequence YVYS is absolutely conserved at positions 41-44, figure 11.3. At these positions the sequences FSAT (x2), FAFS and LSLT were found in the presumptive polypeptides and FSAT (x2), FAFS, YSLT, and FTLT in the *R. acidophila* polypeptides.

d) In the presumptive and *R. acidophila* alpha polypeptides the sequence QGG is absolutely conserved between positions 96-99, figure 11.4. With the exception of polypeptide [14], this sequence is not found at these positions in either strain of *R. palustris*.

e) With the exception of presumptive alpha polypeptide [5], lysine is found at positions 100 and 101 in the presumptive and *R. acidophila* polypeptides, figure 11.4. In alpha polypeptide [5]

lysine is found only at position 100. In *R. palustris* lysine is only found at these positions in polypeptide [14].

An additional piece of evidence which suggests that *R. palustris* strain French did not give rise to the presumptive polypeptide sequences comes from the construction of the library. Recall that two genomic libraries were constructed (ie. *R. acidophila* and *R. palustris*) in a parallel experiment. However, the *R. palustris* library when plated on *E. coli* strains Q358 and Q359 gave rise to a large number of recombinant plaques. The *R. acidophila* library required the use of strain NM621 before a useful number of recombinants could be recovered. This suggests that the genomic inserts did not originate in *R. palustris* strain French.

In addition to the differences outlined above, there are also other differences which can be seen simply by comparing the presumptive polypeptide sequences to those of the other organisims, see figures 11.3 and 11.4.

Unfortunately, the primary sequences of the LHII polypeptides from *R. palustris* strain DSM were not available for comparison. It may be however, that isoleucine and alanine are absolutely conserved at positions 28 and 31 (figure 11.3) in all strains of *R. palustris*. Perhaps suggesting that strain DSM may not have given rise to the presumptive polypeptides.

Such inferences should however be treated with caution. Residues absolutely conserved in one strain may not be conserved in another strain of the same species. Indeed *R. palustris* can be used to illustrate this. Arginine is absolutely conserved at position 55 of strain I1S, whereas in strain French arginine and lysine residues are found at this position, figure 11.4.

The combined evidence of Southern blotting, the method of growing the bacteria, and the comparison of the polypeptide sequences

suggests that the presumptive polypeptide sequences were derived from *R. acidophila* strain 7050.

However, there are differences between the primary sequences determined by amino acid sequencing of the polypeptides from strain 7050 and those deduced from the nucleotide sequences of the clones. This can be seen in the difference score tables presented in chapter 9. How can these sequence differences be reconciled?

Firstly, the deduced presumptive polypeptide sequences were compared to *R. acidophila* sequences which had been obtained by polypeptide sequencing. Though accurate, polypeptide sequencing is not infallible. In total 7 differences were found between consensus sequence 16 and the B800-850 polypeptides of strain 7050. A difference of this order is in keeping with the errors which may occur with polypeptide sequencing [Rene Brunisholz, personal communication]. The differences which were found in the other comparisons however, are probably too great to be considered as polypeptide sequencing errors.

Let us make the assumption that the polypeptides encoded in consensus 16 are the same polypeptides which were found in the B800-850 complex of strain 7050. Then only the B800-830 polypeptides from this strain remain for a comparison to be made with the other presumptive polypeptides. However, other workers have shown that at least one other type of complex (B800-850 type II) is also found in this strain [Cogdell *et al*, 1983]. It is possible that one or more of the presumptive polypeptides are of this type.

It must also be remembered that the not all of the antenna complex types in this strain may have been revealed by spectral and protein analysis. Those isolated previously have been obtained under particular types of environmental conditions ie. cultured in Pfennig's medium at a temperature of 30°C in either high or low light conditions. It is seems feasible that changing any of these parameters may lead to the production of different, and as yet

undiscovered types of complexes. Indeed a precedent for this has been set by *R. acidophila* strain 7750. When placed in low light at 25°C this strain can synthesis B800-820 complexes which it cannot do at 30°C. If such a situation also exists in strain 7050 it seems reasonable to suggests that the presumptive polypeptides may belong to this "undiscovered" class.

In addition genomic Southern blot analysis suggests that strain 7050 may contain a much larger repertoire of antenna polypeptides than initially envisaged, see figures 10.3 and 10.4. When strain 7050 genomic DNA is digestsed with EcoRI+SalI then blotted, and probed with pLHIISB18, ≃11 positively hybridising bands are found. If we now assume that only LHII encoding genes are visualised then this suggests that there are 11 genes each encoding an alpha and beta polypeptide.

It could be argued that some of these bands has resulted from the internal cleavage of genes ie. one gene generating two positive fragments. This type of situation was found in lambda clone 9 where internal cleavage with EcoRI resulted in two positively hybridising fragments. Let us assume that all of the bands have arisen as a result of this mechanism. This would reduce the number of genes to 5. However this is nearly twice the number originally envisaged from spectral analysis and does not take into account any clustering of genes which may occur.

CONCLUSION

The evidence presented suggests that genomic clones encoding light harvesting genes of *R. acidophila* strain 7050 have been isolated. One of these clones, number 16, encodes an alpha and a beta light harvesting polypeptide which have primary amino acid sequences most like those of the B800-850 complex from strain 7050.

Clone 6 has been shown to to encode an alpha polypeptide which has an amino acid sequence most like that of the B800-820 alpha

polypeptide of strain 7750. It is thought that a corresponding beta polypeptide lies 5' to the alpha polypeptide. An additional alpha-beta polypeptide gene pair may also occur within this clone on a positvely hybridising 2.2kb SstI fragment.

Clone 9 contains at least three alpha-beta gene pairs. Consensus 9A and 9C each encode an alpha and a beta polypeptide most like the B800-850 complexes of strain 7750. Consensus 9B encodes an alpha and a beta polypeptide which is most like the B800-820 complex from strain 7750. However, the differences between the 9B and B800-820 polypeptides are so great that this may be a new type of polypeptide.

Note that although the presumptive polypeptides have been called "most like", for example, in the case of consensus 6, the B800-820 complex from strain 7750, this is neither an acceptance of their genomic origin nor a statement on their spectral properties. I therefore propose that the genes and the complexes which their polypeptides form should be named as follows:

Clone	Consensus	Most Like	Complex	Genes
16	16	B800-850 strain 7050	LHII1	*puc*BA1
6	6	B800-820 strain 7750	LHII2	*puc*A2
9	9A	B800-850 strain 7750	LHII3	*puc*BA3
9	9B	B800-820 strain 7750	LHII4	*puc*BA4
9	9C	B800-850 strain 7750	LHII5	*puc*BA5

.ooOoo.

APPENDIX 1 **MEDIA**

For the culture of strains of *Rhodopseudomonas acidophila*:

(1.1) Pfennig's Medium [Pfennig, 1969]:

The medium was used for batch culture of *Rhodopseudomonas acidophila* and was made as follows. The following were added to ≈800mls of water;

Potassium di-hydrogen orthophosphate	1.0g
Magnesium sulphate.$7H_2O$	0.4g
Sodium chloride	0.4g
Sodium succinate	1.5g
Calcium chloride.$2H_2O$	0.05g
Succinic acid	0.065g
Ammonium chloride	0.5g
0.1% w/v Ferric citrate solution	5.0ml
Trace element solution (see below)	10.0ml

The pH was adjusted to 5.2 with HCl then the volume made up to 1 litre with water. The medium was sterilised by autoclaving at 15psi for 20 minutes then stored at 4°C until required.

Trace element solution:

The following were added to ≈800mls of water;

E.D.T.A (di-sodium salt)	0.5g
Ferrous sulphate.$7H_2O$	0.01g
Manganous chloride.$4H_2O$	0.003g
Cobalt chloride.$2H_2O$	0.02g
Calcium chloride.$2H_2O$	0.001g
Nickel chloride.$6H_2O$	0.002g
Sodium molybdate.$2H_2O$	0.003g

The volume was made up to 1 litre with water and the pH adjusted to 3-4 with 10N NaOH.

(1.2) Agar Stabs and Agar Plating Medium:

The same medium was used for stabs and plating and was composed of the following:

Concentration base*	20ml
di-potassium hydrogen orthophosphate (1M)	10ml
Potassium di-hydrogen orthophosphate (1M)	10ml
10% w/v Ammonium sulphate	5ml
Sodium or potassium sulphate (1M)	10ml
Growth factors**	1ml
Agar (Difco)	15g

The volume was made up to 1 litre and the medium autoclaved at 15psi for 20 minutes before use [A. Gardner personal communication, Bose 1963].

Concentration base*:

To 500ml of water was added:

Nitriloacetic acid	10g
Magnesium sulphate	14.45g
Calcium chloride.$2H_2O$	3.4g
Ammonium molybdate	0.0092g
Ferrous sulphate.$7H_2O$	0.099g
Nicotinic acid	0.05g
Aneurine hydrochloride	0.025g
Biotin	0.0005g
Metos 44***	250ml

The pH of the solution was adjusted to 6.8 (at which point it became clear) and the final volume made up to 1 litre with water.

Growth factors**:

To 100ml of water was added:

Biotin	0.02g
Sodium hydrogen carbonate	0.5g
Nicotinic acid	0.1g
Aneurine hydrochloride	0.05g
4-aminobenzoic acid	0.1g (see overleaf)

The solids were boiled to dissolve then the volume made up to 1 litre with water.

Metos 44***:

To 100ml of water was added:

EDTA	0.5g
Zinc sulphate.$7H_2O$	2.19g
Ferrous sulphate.$7H_2O$	1.0g
Manganous sulphate.$7H_2O$	0.308g
Copper sulphate.$5H_2O$	0.0784g
Colbalt nitrate.$6H_2O$	0.0496g
disodium tetraborate.$10H_2O$	0.0354g
Concentrated sulphuric acid	2 drops

MEDIA

For the culture of *E. coli* strains:

(1.3) L-broth:

L-broth is a rich medium used for the growth of E. coli. It was used whenever large numbers of cells or high phage titres were required eg. the isolation of plasmid pLHIISB18 and EMBL3 vector DNA. The medium was made as follows, to ≈800ml of water was added;

Bacto-tryptone	10g
Yeast extract	5g
Sodium chloride	10g

The pH was adjusted to 7.2 with 10N NaOH, the volume made up to 1 litre with water then the medium sterilised by autoclaving at 15psi for 15 minutes. For the growth of EMBL3 and its derivatives magnesium sulphate was added to a final concentration of 10mM before use [Maniatis *et al.*, 1982; Kaiser and Murray, 1985].

(1.4) L-broth agar:

As L-broth with the addition of 15g of Bacto-agar.

(1.5) L-broth top layer:

As L-broth with the addition of 6.5g of Bacto-agar or gel quality agarose.

(1.6) BBL-broth:

Relative to L-broth, BBL-broth is a less rich medium. It was used for the culture of *E. coli* to increase the size of EMBL3 and EMBL3 recombinant plaques. The increase in plaque size occurs because the poorer medium reduces the rate of cell growth giving more time for plaque formation. The medium was made as follows, to ≈800ml of water was added:

Trypticase	10g
Sodium chloride	5g

The pH was adjusted to 7.2 with 10N NaOH, the volume made up to 1litre with water then the medium sterilised by autoclaving at 15psi for 15 minutes [Kaiser and Murray, 1985].

(1.7) BBL-broth agar:

As BBL-broth with the addition of 10g of Bacto-agar.

(1.8) BBL-top layer:

As BBL-broth with the addition of:

Bacto-agar or gel quality agarose	6.5g
Magnesium sulphate	10mM

(1.9) 2x YT:

To 600ml of water added:

Tryptone	16g
Yeast extract	10g
Sodium chloride	10g

Adjusted the pH to 7.6 with 10N NaOH and made the volume up to 1 litre. Sterilised by autoclaving at 15psi for 15 minutes [Bethesda Research Laboratories, 1983].

(1.10) H agar:

H agar formed the bottom layer of plates used for the the growth of *E. coli* strain JM101 infected with M13 [Bethesda Research Laboratories, 1983]. The medium was composed of the following:

2% Water agar	400ml
DM Salts	100ml
20% Glucose	4ml
20% Casamino acids	20ml
Vitamin B1 (1mg/ml)	2ml

These ingredients were made as follows:

Water agar:

20g of Taiyo agar was added to 800ml of water and the volume made up to 1 litre. The agar was melted and the solution dispensed as 400ml aliquots then autoclaved at 15psi for 20 minutes.

DM (Davis and Mingioli) Salts:

To 500ml of water was added:

Dipotassium hydrogen orthophosphate K_2HPO_4	28g
Potassium dihydrogen orthophosphate KH_2PO_4	8g
Ammonium sulphate $(NH_4)_2SO_4$	4g
Tri-sodium citrate $C_6H_5Na_3O_7.H_2O$	1g
Magnesium sulphate $MgSO_4.7H_2O$	0.4g

The volume was made up to a litre and the solution dispensed as 100ml aliquots before being sterilised at 15psi for 15 minutes.

20% Glucose:

200g of glucose were added to 700ml of water. The volume was made up to 1 litre then the solution dispensed into 50ml aliquots before being sterilised at 5psi for 10 minutes.

Casamino Acids:

200g of Casamino acids (Difco) were added to 750ml of water. The volume was made up to 1 litre and the solution dispensed into 20ml aliquots before sterilisation at 15psi for 15 minutes.

Vitamin B1:

100mg of vitamin B1 were dissolved in 100ml of sterile distilled water. The solution was filter sterilised through a 0.22μm filter (Sartorius) then stored at 4°C until required.

(1.11) Minimal H-agar:

This was used for the growth and maintainance of *E. coli* strain JM101 colonies. The lack of proline in the medium forces the retention of the F episome which encodes the proline biosynthetic genes. The composition of the medium is the same as that of H-agar, however, the 20% casamino acids are omitted [Bethesda Reasearch Laboratories, 1983].

(1.12) 0.6% Top Agar:

0.6% top agar was the top layer used for the growth M13 infected E.coli strain JM101. In addition to the agar it contained the chromogenic substrate X-gal and the lac operon inducer IPTG [BRL 1983].

0.6% top agar was composed of:	0.6% water agar	50ml
	X-Gal solution	1ml
	IPTG solution	1ml

The mixture was stored at 4°C in the dark until required. The ingredients were composed of:

0.6% agar:

6g of Taiyo agar were added to 800ml of water then the volume made up to 1 litre. The agar was dissolved by heating and the solution dispensed into 50ml aliquots. The medium was autoclaved at 15psi for 15 minutes.

X-Gal solution:

25mg of X-Gal were dissolved in 1ml of dimethylformamide then added to the molten 0.6% agar.

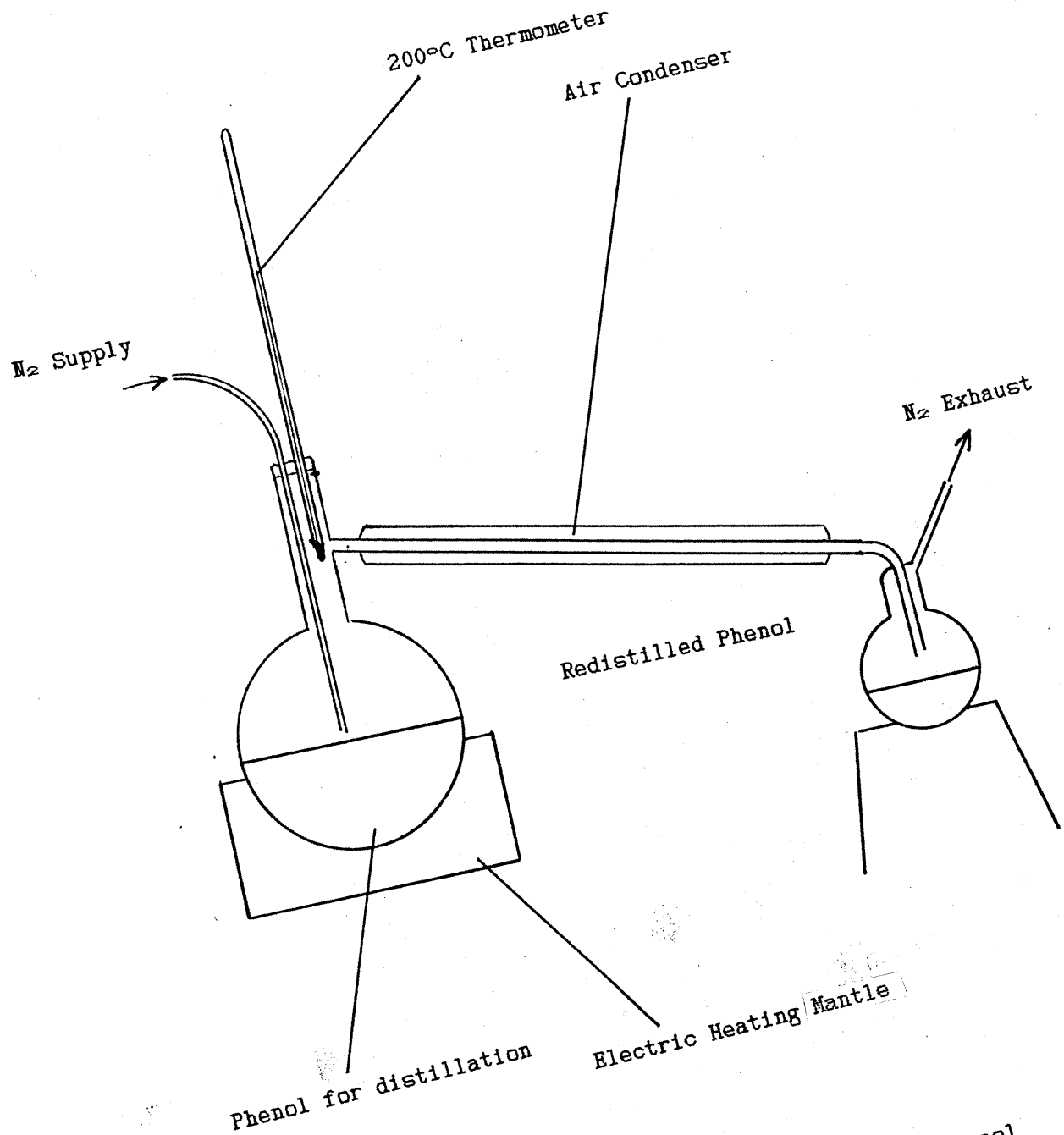


Figure A1: Apparatus for the Redistillation of Phenol

IPTG solution:

28mg of IPTG were dissolved in 1ml of water then added to the molten 0.6% agar.

DNA EXTRACTION AND PURIFICATION MATERIAL

(1.13) Phenol:

Phenol refers to redistilled phenol, which was prepared as follows. Bottles of solid phenol were placed in a 68°C water bath until their contents had melted. The liquid phenol was poured into a round bottomed flask which formed part of the equipment shown in figure A1. Once the equipment was assembled the nitrogen supply was turned on and the flow of gas adjusted until it just caused ripples on the surface of the phenol. The phenol was then redistilled at 160°C. The first 50mls of distillate were discarded, the following ≈800mls were run into a round bottomed flask containing 50ml of 1M Tris.HCl pH 8.0, the last ≈50mls of undistilled phenol were not collected. 8-hydroxyquinoline was added to the redistilled phenol to a concentration of 0.1% w/v, then the mixture stored as 50ml aliquots at -20°C until required. Before use the phenol was extracted once with an equal volume of 1M Tris.HCl pH 8.0 then with equal volumes of 0.1M Tris.HCl pH 8.0 until the pH of the aqueous phase was >7.6. The equilibrated phenol was stored at 4°C when not in use [Maniatis et al 1982].

(1.14) Phenol-chloroform:

Phenol-chloroform was prepared by mixing equal volumes of redistilled phenol and chloroform. The phenol-chloroform mixture was stored at 4°C when not in use.

(1.15) Water saturated butanol:

To a volume of butanol an equal volume of sterile distilled water was added, the mixture shaken thoroughly then allowed separate into two phases. The upper, butanol containing phase was then ready for use.

(1.16) Dialysis tubing:

The tubing was cut into ≃50, 30cm lengths then boiled for 15minutes in 1.5 litres of 1% w/v EDTA. The liquid was poured off and the process repeated with the same volume of 0.1% w/v EDTA solution. The liquid was poured off and the tubing boiled twice for 15 minutes in 1.5 litres of 1% w/v sodium bicarbonate solution. The liquid was poured off and the tubing boiled twice for 15 minutes in 1.5 litres of water. The tubing was stored at 4°C in 10mM EDTA, 25% v/v ethanol until required. This is a modification of the method of Schleif and Wensink, 1981.

(1.17) 3M Sodium acetate pH 5.2:

408.1g of sodium acetate.$3H_2O$ were dissolved in ≃800ml of water. The pH was adjusted to 5.2 with glacial acetic acid then the volume made up to 1 litre with water. The solution was dispensed into aliquots then sterilised by autoclaving at 15psi for 15 minutes [Maniatis *et al.*, 1982].

(1.18) Sephadex G-50:

10g of Sephadex G-50 was added to 500ml of TE. The mixture was autoclaved at 15psi for 15 minutes before use.

GENERAL BUFFERS AND SOLUTIONS:

(1.19) Phage buffer:

Phage buffer was used for the storage of phage stocks and for dialysis of phage suspensions. To 700ml of water was added:

Disodium hydrogen orthophosphate: dodecahydrate ($Na_2HPO_4.12H_2O$)	18g
Potassium dihydrogen orthophosphate KH_2PO_4	3g
Sodium chloride	5g
Magnesium sulphate solution (0.1M)	10ml
Calcium chloride solution (50mM)	2ml
1% Gelatin solution (see below)	1ml

The buffer was dispensed into 100ml aliquots then sterilised at 15psi for 15 minutes.

1% Gelatin solution:

10g of gelatin was added to 750ml of water. The volume was made up to 1 litre and the solution sterilised at 5psi for 10 minutes.

(1.20) TE:

TE was used for the storage and isolation of DNA and was composed of.

Tris.HCl pH 8.0	10mM
EDTA pH 8.0	1mM

(1.21) 20x SET:

5 litres of 20x SET were made by dissolving 876.6g of sodium chloride, 37.2g of EDTA and 0.4M Tris.HCl pH 7.8 in 4 litres of water. The pH was adjusted to 7.8 with HCl then the volume made up to 5 litres. The solution was dispensed into 500ml aliquots and autoclaved at 15psi for 15 minutes before use [Mason and Williams, 1985].

(1.22) 20x SSC:

175.3g of sodium chloride and 88.2g of sodium citrate were dissolved in ≈800ml of water. The pH was adjusted to 7.0 with 10N NaOH then the volume made up to 1 litre with water [Maniatis *et al* 1982].

SOLUTIONS FOR NUCLEIC ACID HYBRIDISATION

See also (1.22) and (1.23) above.

(1.23) 100x Denhardts:

10g of Ficoll, 10g of polyvinylpyrrolidone and 10g of BSA (Pentax Fraction V) were added to ≈350ml of water. The mixture was stirred until the solids had dissolved (about 1 hour) then the volume made up to 500ml with water. The solution was dispensed into 20ml aliquots which were stored at -20°C [Maniatis *et al.*, 1982, Mason and Williams, 1985].

(1.24) 10mg/ml Sheared Herring sperm DNA:

1g of herring sperm DNA was dissolved in 100ml of water. To shear the DNA the solution was passed several times through an 18swg needle then sonicated for 15 minutes. An aliquot was run on a 1.0% agarose gel to check that the DNA had been sheared to less than 1kb. The solution was dispensed into 2ml aliquots which were stored at -20°C until required. Immediately before use the DNA was denatured by boiling for 15 minutes [Mason and Williams, 1985].

(1.25) Prehybridisation solution:

200ml of prehybridisation solution were prepared from the following volumes of stock solutions:

	Vol. added	Working concentration.
20x SET	40ml	4x
100x Denhardt's soln.	20ml	10x
10% w/v SDS	2ml	0.1% w/v
10mg/ml Sheared herring sperm DNA	4ml	200µg/ml.

The final volume was made up to 200ml with water. The prehybridisation solution was stored at 4°C until required. Before use the solution was warmed in a microwave oven to allow dissolution of solid SDS [Mason and Williams, 1985].

(1.26) Hybridisation solution:

Hybridisation solution consists of prehybridisation solution after the radiolabelled probe has been added.

ELECTROPHORESIS BUFFERS AND SOLUTIONS

(1.27) 10x TBE:

To 1 ltre of water was added:

Tris base	219g
Boric acid	110g
EDTA	19g

The mixture was stirred until the solids had dissolved then the volume made up to 2 litres with water. The working strength of solution was 1x TBE [Maniatis *et al.*, 1982].

(1.28) 50x TAE:

When bands were to be excised from low melting point agarose gels and the DNA purified by Genecleaning, TAE was used as an electrophoresis buffer. The use of TBE in these circumstances results in reduced recovery of DNA from the Geneclean process.

242g of Tris base was dissolved in ≈800ml of water. 57.1ml of glacial acetic acid and 100ml of 0.5M EDTA pH 8.0 were then added and the final volume made up to 1 litre with water. The working strength of solution was 1x TAE [Maniatis *et al.*, 1982].

(1.29) Loading dye:

To 40ml of water, 7.5g of Ficoll type 400, 25mg of Bromophenol Blue and 25mg of Xylene Cyanol were added. The mixture was stirred for 1 hour then filtered through a 0.45µm filter (Sartorius) to remove undissolved particles. The solution was dispensed into 1ml aliquots and stored at room temperature [Maniatis *et al.*, 1982].

(1.30) 10mg/ml Ethidium bromide solution:

A 10mg/ml stock solution was made by dissolving 1g of ethidium bromide in 100ml of water. With continuous stirring, dissolution took several hours. Undissolved ethidium bromide particles were removed by filtering the solution through a 0.45µm filter (Sartorius). Aliquots were dispensed into bottles wrapped in foil to exclude light. For staining gels the stock solution was diluted to ≈0.5µg/ml.

(1.31) 8% Acrylamide:

20mls of 40% acrylamide stock (see below), 50g of urea and 5mls of 20x TBE were added to 35ml of water. The solution was stirred at 40°C until the urea had dissolved then 1ml of 10% ammonium persulphate and 20µls of TEMED added. The gel was poured immediately.

40% acrylamide stock was prepared by adding 190g of acrylamide and 10g of bis-acrylamide to ≈300ml of water. The total volume of the solution was made up to 500ml with water. The stock solution was filtered through 2V filter paper (Whatman) then through a 0.45µm

filter. The solution was stored in a light tight bottle at 4°C until required [Bethesda Research Laboratories, 1983].

DIDEOXYSEQUENCING REAGENTS:

(1.32) Sequenase version 2.0 kit:

The following reagents were included in the sequencing kit.

5x Sequenase Buffer:

200mM Tris.HCl pH 7.5
100mM $MgCl_2$
250mM NaCl

Primer:

Universal 5'-GTAAAACGACGGCCAGT-3' 0.5 pmol/µl
(-40) 5'-GTTTTCCCAGTCACGAC-3' 0.5 pmol/µl

DTT 0.1M

Enzyme Dilution Buffer:

10mM Tris.HCl pH 7.5.
5mM DTT.
0.5mg/ml BSA

5x Labelling Mix:

7.5µM dGTP
7.5µM dCTP
7.5µM dTTP

ddG Termination Mix:

80µM dGTP, 80µM dATP, 80µM dCTP, 80µM dTTP, 8µM ddGTP and 50mM NaCl.

ddA Termination Mix:

80µM dGTP, 80µM dATP, 80µM dCTP, 80µM dTTP, 8µM ddATP and 50mM NaCl.

ddT Termination Mix:
80μM dGTP, 80μM dATP, 80μM dCTP, 80μM dTTP, 8μM ddTTP and 50mM NaCl.

ddC Termination Mix:
80μM dGTP, 80μM dATP, 80μM dCTP, 80μM dTTP, 8μM ddCTP and 50mM NaCl.

Stop solution:

95% Formamide
20mM EDTA
0.05% Bromophenol Blue
0.05% Xylene Cyanol

RESTRICTION BUFFERS

(1.33) Bethesda Research Laboratories supplied their restriction enzymes with the appropriate 10x restriction buffer (ie. ReAct). The composition of these buffers were:

ReAct 2: 50mM Tris HCl pH 8.0; 10mM $MgCl_2$; 50mM NaCl.
ReAct 3: 50mM Tris HCl pH 8.0; 10mM $MgCl_2$; 100mM NaCl.
ReAct 4: 20mM Tris HCl pH 7.4; 5mM $MgCl_2$; 50mM KCl.

APPENDIX 2 REAGENTS AND EQUIPMENT

REAGENTS:

Common laboratory chemicals and reagents were purchased from a number of suppliers without preference to any particular source. Wherever possible AnAlar or Biochemical grade products were purchased. The suppliers were:

BDH, BDH Chemicals Ltd., Poole, England.
Formachem, Formachem (Research International) Ltd., Strathaven, Scotland.
Hopkins and Williams, Chadwell Heath, Essex, England.
May and Baker, Dagenham, England.
Riedel-De Haen AG, Seelze-Hannover, West Germany.

SPECIFIC REAGENTS:

Acrylamide (Electran): BDH.
Agarose (Type I Low EEO): Bethesda Research Laboratories.
Agarose (Low Melting Point): Bethesda Research Laboratories.
Ampicillin: Sigma.
ATP: Sigma.
dATP: Pharmacia
^{32}P-dATP: New England Nuclear.
^{35}S-dATP: New England Nuclear.
Bacto-Agar: Difco.
Bis-acrlamide: BDH.
Bromophenol blue: Sigma.
BSA (Pentax Fraction V): Sigma.
Casamino acids: Difco.
Dialysis tubing: Medicell.
Dimethylformamide: BDH.
DNase I, from bovine pancreas (crude extract): Sigma.
Ethidium bromide: Sigma.

Ficoll (Type 400): Sigma.

Filters (0.22μm and 0.45μm): Sartorius.

Filter paper (3MM and 2V): Whatman.

Geneclean kit: Bio 101 Inc.

dGTP: Pharmacia.

HEPES: Sigma.

Herring sperm DNA: Sigma.

Hexadeoxyribonucleotides: Pharmacia.

Hybond-N: Amersham International.

8-Hydroxyquinoline: BDH.

IPTG: Northumbria Biolologicals Ltd.

Klenow fragment: Bethesda Research Laboratories.

β-Mercaptoethanol: BDH

Nitrocellulose: Schleicter and Schuell

Nuclease free BSA 10mg/ml: Pharmacia.

Orange-G: Sigma.

PEG 6000: Sigma.

Polyvinylpyrrolidone: Sigma.

Proteinase K (from *Tritirachium album*): Sigma.

Restriction endonucleases: Bethesda Research Laboratories.

RNase A (type I-AS from bovine pancreas): Sigma.

Sephadex G-50: Pharmacia.

Sequenase version 2.0 kit: United States Biochemicals.

T4 Ligase: Bethesda Research Laboratories.

Taiyo agar: Difco.

Trypticase: Baltimore Baltimore Biological Laboratories.

Tryptone: Difco.

dTTP: Pharmacia.

Urea: Sigma.

Vitamin B1: Sigma.

X-Gal: IBI.

Xylene cyanol: Sigma.

Yeast extract: Difco.

EQUIPMENT

General:

Falcon tubes: Becton Dickinson Labware.

"Rock and Roller" as found in haematology labs for mixing blood samples: Denley.

Low Speed Centrifuges:

Beckman Model J2-21 used: JS-7.5 Swing out rotor (50ml Falcon tubes).
JA-10 Angle rotor (500ml tubes).
JA-14 Angle rotor (250ml tubes).
JA-20 Angle rotor (50mltubes).

MSE Microfuge: Angle rotor (0.5 and 1.5ml Eppendorf vials)

Jouan CR312: Swing out rotor (50ml Falcon tubes).

Ultra Centrifuges:

Beckman L8-55M, L8-55 and L7-55 all used the following rotors:
VTi-50 Vertical rotor (39ml tubes).
VTi-65 Vertical rotor (5.1ml tubes).
Ti-70.1 Angle rotor (13.5ml tubes).
SW-28 Swing out rotor (38.5ml tubes).

Gel and Electrophoresis equipment:

Large gel tanks made in University workshop.
Minigel tanks: EMBL 86 SM50, EMBL86 M100 and BRL Baby Gel tank.
DC Powerpacks: Model PAB 250-0.5 (Kikusui Electronics Corporation).

Transilluminator: 312nm wavelength generated by six, 15W tubes. (Vilber Lourmat, France)

Sequencing Powerpack: LKB 219 Power Supply (Pharmacia LKB Biotechnology).

Gel Drier: Model 483 Slab Drier (Biorad).

Autoradiography equipment:

Intensifying screens: Cronex, Quanta Fast Detail (E.I. Du Pont De Nemours).
Cassettes: Genetic Research Instrumentation Ltd.
Film: Kodak X-Omat S.

REFERENCES

Aagaard, J. and W. R. Sistrom. 1972. **Control of synthesis of reaction centre bacteriochlorophyll in photosynthetic bacteria.** Photochem. Photobiol. 15: 209-225.

Aaij, C. and P. Borst. 1972. **The gel electrophoresis of DNA.** Biochim. Biophys. Acta 269:192.

Abdulaev, G. E. , M. Yu Feigina, A. V. Kisekev. 1978. **Products of limited proteolysis of bacteriorhodopsin generate a membrane potential.** FEBS Letts 90: 190-194.

Adams, C. W., M. E. Forrest, S. N. Cohen and J. T. Beatty. 1989. **Structural and functional analysis of transcriptional control of the *Rhodobacter capsulatus puf* operon.** J. Bacteriol. 171: 473-482.

Adams, R. P. L., and R. H. Burdon. 1985. **The function of DNA methylation in bacteria and phage.** In "Molecular Biology of DNA Methylation." (Ed. Rich, A.) pp. 73-87. Springer-Verlag. New York.

Agliadis, I., and F. Reiss-Husson. 1983. **Several properties of the LH unit extracted with sodium dodecyl sulphate from *Rps. sphaeroides*: Preliminary characterisation.** Biochim. Biophys. Acta 724: 340-351.

Aiking, H. and G. Sojka. 1979. **Response of *Rhodopseudomonas capsulata* to illumination and growth rate in a light limited continuous culture.** J. Bacteriol. 139: 530-536

Alberti, M and J. E. Hearst. 1988. Cited in Zhu , Y. J. and J. E. Hearst, 1988. **Transcription of oxygen-regulated photosynthetic genes requires DNA gyrase in *Rhodobacter capsulatus*.** Proc. Natl. Acad. Sci. USA 85:4209-4213.

Alberts, B., D. Bray, J. Lewis, M. Raff. 1983. **The Molecular Biology of the Cell.** Garland Publishing, Inc. New York & London.

Allen, J. P., G. Feher, T. O. Yeates, H. Komiya and D. C. Rees. 1987a. **Structure of the reaction centre from *Rb. sphaeroides* R-26: the cofactors.** Proc. Natl. Acad. Sci. USA 84: 5730-5734.

Allen, J. P., G. Feher, T. O. Yeates, H. Komiya and D. C. Rees. 1987b. **Structure of the reaction centre from *Rb. sphaeroides*: the protein subunits.** Proc. Natl. Acad. Sci. USA 84: 6162-6166.

Allen, J. P., G. Feher, T. O. Yeates, D. C. Rees, J. Deisenhofer, H. Michel and R. Zuber. 1986. **Structural homology of reaction centres from *Rhodopseudomonas sphaeroides* and *Rhodopseudomonas viridis* as determined by X-ray diffraction.** Proc. Natl. Acad. Sci. USA. 83: 8589-8593.

Ames, G. F. L. and K. Nakaido. 1985. **Nitrogen regulation in *Salmonella typhinurium*. Identification of an ntrC protein-binding site and definition of a consensus binding sequence.** The EMBO Journal 4: 539-547.

Amesz. J. 1985. **Photosynthesis: Structure of membrane and membrane proteins.** Progress in Botany. 47: 87-104.

Angerhofer, A., R.J. Cogdell and M.F. Hipkins 1986. **A spectral characterisation of the light-harvesting pigment protein complexes from *Rhodopseudomonas acidophila*.** Biochim. Biophys. Acta 848: 333-341.

Ashby, M. K., S. A. Coomber and C. N. Hunter. 1987. **Cloning, nucleotide sequence and transfer of genes for the B800-850 light harvesting complex of *Rhodobacter sphaeroides*.** FEBS Lett. 213: 245-248.

Bakker, J. G. C., R. van Grondelle, and W. T. F. Den Hollander. 1983. **Trapping loss and annihilation of excitations in a photosynthetic system.** Biochim. Biophys. Acta 725: 508-518.

Bauer, C. E., and B. L. Marrs. 1987. **Oxygen control of bacteriochlorophyll biosynthesis occurs by regulating transcription through *puf*Q, a newly identified *puf* operon gene product.** J. Cell. Biochem. 11C:47.

Bauer, C. E., M. Eleuterio, D. A. Young and B. L. Marrs. 1987. in Proceedings of the Seventh International Congress of Photosynthesis. Ed. Biggins, J. Nijhoff Dordrecht, The Netherlands, vol. 4 pp669-705.

Bauer, C. E., Young, D. A. and Marrs, B. L. 1988. **Analysis of the *Rhodobacter capsulatus puf* operon. Location of the oxygen regulated-promoter region and the identification of an additional *puf*-encoded gene.** The Journal of Biological Chemistry 263: 4820-4827.

Belasco, J. G., J. T. Beatty, C. W. Adams, A. V. Gabain and S. N. Cohen. 1985. **Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic rxcA transcript.** Cell 40: 171-181.

Benson, S. A. and T. J. Silhavy. 1983. In "Gene Function in Prokaryotes" (Eds. Beckwith, J., J, Davies and J. A. Gallant) pp. 253-280. Cold Spring Harbour Laboratory , New York.

Berard, J., G. Belanger, P. Corriveau, and G. Gingras. 1986. **Molecular cloning and sequence of the B880 holochrome gene from *R. rubrum*.** J. Biol. Chem. 261: 82-87.

Bethesda Research Labs. 1980. Dideoxy Sequencing Manual.

Betler, M., and D. Freifelder. 1983. **Studies on the replication of *E. coli* phage λ DNA I. The kinetics of DNA replication and**

requirements for the generation of rolling circles. Virology 126: 168-182.

Biel, A. J. and B. L. Marrs. 1983. **Transcriptional regulation of several genes for bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata* in response to oxygen.** J. Bacteriol. 156: 686-694.

Biel, A.J. 1986. **Control of bacteriochlorophyll accumulation by light in *Rhodobacter capsulatus*.** J. Bacteriol. 168 :655-659.

Birnboim, H. C. and J. Doly. 1979. **A rapid alkaline extraction procedure for screening recombinant plasmid DNA.** Nucleic Acids Res. 7: 1513.

Bissig, I., R. A. Brunisholz, F. Suter, R. J. Cogdell and H. Zuber. 1988. **The complete amino acid sequences of the B800-850 antenna polypeptides from *Rhodopseudomonas acidophila* strain 7750.** Z. Naturforsch. 43c 77-83.

Bissig, I., R. Wagner, R. A. Brunisholz, G. Frank, and H. Zuber. 1989a. **Multiple forms of light harvesting polypeptides in *Chromatium vinosum*.** In "Symposium on Molecular Biology of Membrane-Bound Complexes in Phototropic Bacteria" pp105. Freiburg im Breislau. FRG.

Bissig, I., R. Brunisholz, G. Frank, and H. Zuber. 1989b. ***Ectothiorhodospira* contains four polypeptides with core antenna primary structure characteristics.** In "Symposium on Molecular Biology of Membrane-Bound Complexes in Phototropic Bacteria" pp106. Freiburg im Breislau. FRG

Blankenship, R. E. and R. C. Fuller. 1986. **Membrane topology and photochemistry of the green photosynthetic bacterium *Chloroflexus aurantiacus*.** In "Photosynthesis III" Encyclopedia of Plant Physiology New Series Volume 19. Photosynthetic Membranes and Light

Harvesting Systems. Eds. L. A. Staehelin and C. J. Arntzen. Springer-Verlag. Berlin, Heidelberg, New York, Tokyo. p390-421.

Blobel, G., and B. Dobberstein. 1975. **Transfer of proteins across membranes.** J. Cell. Biol. 67: 835-851.

Bolt, J. and J. Sauer. 1979. **Linear dichroism of light-harvesting bacteriochlorophyll proteins from *Rhodopseudomonas sphaeroides* in stretched polyvinyl alcohol films.** Biochim. Biophys. Acta. 546: 54-63.

Bolt, J., J. Sauer, J.A. Shiozawa and G. Drews. 1981. **Linear and circular dichroism of membranes from *Rhodopseudomonas capsulata*.** Biochim. Biophys. Acta. 635: 535-541.

Borisov, A.Y. 1978. **Energy migration mechanisms in antenna chlorophylls.** p323-331. In "The photosynthetic bacteria" (Eds. Clayton, R. K., and W.R. Sistrom). Plenum Press. New York.

Bose , S. K. 1963. **Media for aerobic growth of photosynthetic bacteria.** In "Bacterial Photosynthesis" (Eds. Gest, H., A. San Pietro and L.P. Vernon). pp.501. Yellow Springs. Ohio.

Breton, J., A. Vermeglio, A. Garrigos and M. Paillotin. 1981. In **"Photosynthesis Vol III: Structure and Molecular Organization of the Photosynthetic Apparatus."** (Ed. Akoyunoglou, G.). Balban. Int. Sci. Serv. Phil.

Breton, J., and A. Vermeglio. 1982. **Orientation of photosynthetic pigments *in vivo*** p153-194. In Govindjee (Ed.) Photosynthesis volI. Academic Press Inc. New York.

Breton, J. and E. Nabedryk. 1984. **Transmembrane orientation of α-Helices and the organisation of chlorophyll in photosynthetic pigment-protein complexes.** FEBS. Lett. 176: 355-359.

Brill, S. J. and R. Sternglanz. 1988. **Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants.** Cell. 54: 403-411.

Broglie, R. M., C. N. Hunter, P. Pelepelaire, R. A. Niederman, N. H. Chua and R.K. Clayton. 1980. **Isolation and characterisation of the pigment-protein complexes of *Rhodopseudomonas sphaeroides* by lithium dodecyl sulphate/polyacrylamide gel electrophoresis.** Proc. Natl. Acad. Sci. USA. 77: 87-91.

Brunisholz, R. A., P. A. Cuendet, R. Theiler and H. Zuber. 1981. **The complete amino acid sequence of the single light harvesting protein from chromatophores of *Rhodospirillum rubrum*-G-9+.** FEBS. Lett. 129: 150-154.

Brunisholz, R., F. Suter and H. Zuber. 1983. Eur. J. Biochem.

Brunisholz, R. A., F. Suter and H. Zuber. 1984a. **The light harvesting polypeptides of *Rhodospirillum rubrum* I. The amino acid sequence of the second LH polypeptide B880-β (B870-β) of *Rhodospirillum rubrum* S1 and the carotenoidless mutant G-9-+. Aspects of the two LH polypeptides B880-α (B870α) and B880-β (B870-β) and of the antenna complex B880 (B870) from *Rhodospirillum rubrum*.** Hoppe-Seylers Z. Physiol. Chem. 365: 675-688.

Brunisholz, R. A., V. Wiemken, F. Suter, R. Bachofen and H. Zuber. 1984b. **The light-harvesting polypeptides of *Rhodospirillum rubrum* II. Localization of the amino acid terminal regions of the LH-polypeptides of B870-α and B870-β and the reaction centre subunit L at the cytoplasmic side of the photosynthetic membrane of *Rhodospirillum rubrum* G-9-+.** Hoppe-Seylers Z. Physiol. Chem. 365: 689-701.

Brunisholz, R. A., F. Jay, F. Suter, and H. Zuber. 1985 **The light-harvesting polypeptides of *Rhodopseudomonas viridis*. The complete amino acid sequences of B1015-α B1015-β and B1015-γ.** Biol. Chem. Hoppe-Seyler. 366: 87-98.

Brunisholz, R. A., H. Zuber, J. Valentine, J. G. Lindsay, K. J. Woolley and R. J. Cogdell. 1986. **The membrane localisation of the B890-complex from *Rhodospirillum rubrum* and the effect of carotenoid on the conformation of its apoproteins exposed on the cytoplasmic surface.** Biochim. Biophys. Acta. 849: 295-303.

Brunisholz, R. A. and H. Zuber. 1987. **Primary structure analyses of bacterial antenna polypeptides.** In "Photosynthetic Light-Harvesting Systems" Proc. Int. Workshop, Freising, FRG. (Eds. Scheer, H. and S. Schneider), Walter De Gruyter, Berlin.

Brunisholz, R. A., I. Bissig, E. Niederer, F. Suter, and H. Zuber. 1987. **Structural studies on the light harvesting polypeptides of *Rps. acidophila*.** In "Progress in Photosynthesis Research II" (Ed. Biggins, J.) Martinus Nijhoff, The Netherlands.

Burgess, J. G., M. K. Ashby, S. A. Coomber and C. N. Hunter. 1988. **Deletion and plasmid complementation of the genes encoding the reaction centre and LHI antenna polypeptides of *Rhodobacter sphaeroides*.** In Abstracts of the VI International Symposium on Photosynthetic Prokaryotes. Noordwijkerhout, Netherlands. p205.

Campillo, A. J., R. C. Hyer, T. G. Monger. 1977. **Light collection and harvesting processes in bacterial photosynthesis investigated on a picosecond timescale.** Proc. Natl. Acad. Sci. USA. 74: 1997-2001.

Capaldi, R. A. 1979. Ed. Membrane Proteins in Energy Transduction. New York: Dekker.

Cashmore, A. R. 1984. **Structure and expression of a pea nuclear gene encoding a chlorophyll a/b binding polypeptide.** Proc. Natl. Acad. Sci. USA 81: 2960-64.

Chen, C-Y. A., J. T. Beatty, S. N. Cohen and J. G. Belasco. 1988. **An intercistronic stem-loop structure functions as an mRNA decay**

terminator necessary but insufficient for puf mRNA stability. Cell. 52:609-619.

Chory, J., T. J. Donohue, A. R. Varga, L. A. Staehelin and S. Kaplan. 1984. **Induction of photosynthetic membranes of *Rhodopseudomonas sphaeroides*: biochemical and morphological studies.** J. Bacteriol 159: 540-554

Chou, P. Y. and G. D. Fasman. 1978. **Empirical predictions on protein conformation.** Ann . Rev. Biochem. 47: 251-276.

Chua, N.-H., G. W. Schmidt. 1978. **Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose 1,-5,-bisphosphate carboxylase.** Proc. Natl. Acac. Sci. USA 75: 6110-14

Clark, W.G., E. Davidson and B. L. Marrs. 1984. **Variations of levels of mRNA coding for antenna and reaction centre polypeptides in *Rhodopseudomonas capsulata* in responses to changes in oxygen concentration.** J. Bacteriol. 157: 945-949.

Clayton, R. K. 1963. **Towards the isolation of a photochemical reaction centre in *Rhodopseudomonas sphaeroides*.** Biochim. Biophys. Acta 75: 312-323.

Clayton, R. K. 1966. **Spectrophotometric analysis of bacteriochlorophylls *in vivo* and *in vitro*.** Photochem. Photobiol. 5: 669.

Clayton. R. K. and B. J. Clayton. 1972. **Relations between pigments and proteins in the photosynthetic membrane of *Rhodopseudomonas sphaeroides*.** Biochim. Biophys. Acta. 501: 470-477.

Clayton, R. K. and R. Haselkorn. 1972. **Protein components of bacterial photosynthetic membranes.** J. Mol. Biol. 68:97-105.

Clayton, R. K. 1978. **Physiochemical mechanisms in reaction centres of photosynthetic bacteria.** In "The Photosynthetic Bacteria" (Eds. Clayton, R. K. and W.R Sistrom) Plenum Press, New YorK.

Clayton, R. K. 1980. 'Photosynthesis: Physical Methods and Chemical Patterns'. pp. 88-110. Cambridge University Press, Cambridge, England.

Cogdell, R. J., W. W. Parson, M. A. Kerr. 1976. **The type amount and location of carotenoids within reaction centres of *Rps. sphaeroides*.** Biochim. Biophys. Acta 430: 83-93.

Cogdell, R. J. and A. R. Crofts. 1978. **Analysis of the pigment content of an antenna pigment-protein complex from three strains of *Rhodopseudomonas sphaeroides*.** Biochim. Biophys. Acta. 502: 409-416.

Cogdell, R. J. and J. P. Thornber. 1979. **The preparation and chracterisation of different type of light-harvesting pigment-protein complexes from some purple bacteria.** In "Chlorophyll organisation and energy transfer in photosynthesis". Ciba Foundation Symp. 61 (New Series). (Ed. Wolstenholme, G. and P. W. Fitzsimons). pp61-79. Elsevier/Excerpta Medica.

Cogdell, R. J. and J. P. Thornber. 1980. **Light-Harvesting pigment-protein complexes of purple photosynthetic bacteria.** FEBS. Lett. 112: 1-8.

Cogdell, R. J., M. F. Hipkins, W. Macdonald and T.G. Truscott. 1981. **Energy transfer between the carotenoid and the bacteriochlorophyll within the B800-850 light-harvesting pigment protein complex of *Rhodopseudomonas sphaeroides*.** Biochim. Biophys. Acta. 634: 191-202.

Cogdell, R. J., J. G. Lindsay, J. Valentine and I. Durant. 1982. **A further characterisation of the B890 Light-harvesting pigment protein complex from Rhodospirillum rubrum strain S1.** FEBS Lett. 151: 151-154.

Cogdell, R. J. 1983. **Photosynthetic reaction centres.** Ann. Rev. Plant Physiol. 34: 21-45.

Cogdell, R. J., I. Durant, J. Valentine, J.G. Lindsay and K. Schmidt. 1983. **The isolation and partial characterisation of the light harvesting pigment-protein complement of *Rhodopseudomonas acidophila*.** Biochim. Biophys. Acta. 772: 427-435.

Cogdell, R. J. 1985. **Carotenoids in Photosynthesis.** Pure and Applied Chemistry. 57: 723-728.

Cogdell, R. J. and H. Scheer. 1985. **Circular dichroism of light-harvesting complexes from purple photosynthetic bacteria.** Photochem. Photobiol. 42: 669-689.

Cogdell, R. J. 1986. **Light-Harvesting complexes in the purple photosynthetic bacteria.** In the: Encyclopedia of Plant Physiology. New Series, vol. 19. Photosynthesis III. (Eds. Staehelin, L. A. and C. J. Arntzen.) pp252-259. Springer-Verlag. Berlin, Heidelberg, New York, Tokyo.

Cohen, S. N., A. C. Y. Chang, H. W. Boyer. 1974. **Replication and transcription of eukaryotic DNA in *E. coli*.** Proc. Natl. Acad. Sci. USA. 71: 1743-1747.

Cohen-Bazire, G. and W. R. Sistrom. 1966. In The Chlorophylls (Ed Vernon, L. P. and G. R. Seely) pp313. Academic Press. New York.

Cook, D. N., G. A. Armstrong and J. E. Hearst. 1989. **Induction of anaerobic gene expression in *Rhodobacter capsulatus* is not accompanied by a local change in chromosomal supercoiling as measured by a novel assay.** J. Bacteriol. 171: 4836-4843.

Dagert, M., and S. D. Ehrlich. 1979. **Prolonged incubation in calcium chloride improves the competence of *E. coli* cells.** Gene 6: 23.

Davis, J., T. J. Donohue and S. Kaplan. 1988. **Construction, characterisation and complementation of a puf- mutant of Rhodobacter sphaeroides.** J. Bacteriol. 170: 320-329.

Dawkins, D. J. 1988. **Isolation and characterisation of reaction centre and antennae conjugates from a range of bacteriochlorophyll a containing species of purple bacteria.** PhD. Thesis. Department of Botany. University of Glasgow.

Debus, R. J., G. Feher and M. Y. Okamura. 1985. **LM complex of reaction centres from *Rsp. sphaeroides* R-26. Characterisation and reconstitution with the H subunit.** Biochemistry 24: 2488-2500.

Dehardt, D. T., H. H. Dressler, and D. S. Ray. (Eds.) 1978. **Single stranded DNA phages.** Cold Spring Harbour Laboratory 11:724.

Deisenhofer, J., O. Epp, K. Miki, R. Huber and H. Michel. 1984. **X-ray strucure analysis of a membrane protein complex.** J. Mol. Biol. 180: 385-398.

Deisenhofer, J., H. Michel, R. Huber. 1985a. **The strucural basis of photosynthetic light reactions.** Trends in Biochemical Science : 10: 243-248.

Deisenhofer, J., O. Epp, K. Miki, R. Huber and H. Michel. 1985b. **Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3Å resolution.** Nature 318: 618-624.

Dierstein, R. 1983. **Biosynthesis of pigment-protein complex polypeptides in bacteriochlorophyll-less mutant of *Rhodopseudomonas capsulata* YS.** FEBS Lett. 160: 281-286.

Dierstein, R. 1984. **Synthesis of pigment-binding protein in toluene treated *Rhodopseudomonas capsulata* and in cell free systems.** Eur. J. Biochem. 138: 509-518.

Dobberstein, B., G. Blobel, N.-H, Chua. 1977. ***In vitro*** **synthesis and processing of a putative prrecursor for the small subunit of ribulose-1,5,-bisphosphsate carboxylase of *Chlamydomonas reinhardtii*.** Proc. Natl. Acad. Sci. USA 74: 1082-85.

Donohue, T. J., J. H. Hoger, and S. Kaplan. 1986. **Cloning and expression of the *Rhodobacter sphaeroides* reaction centre H gene.** J. Bacteriol. 168: 953-961.

Drews, G., R. Feick, A. Schumacher, N. N. Firsow. 1977. **Organisation and formation of the photosynthetic apparatus of *Rps. capsulata* and *Rps. palustris*.** In "Proceedings of the fourth international congress on photosynthesis" (Eds. Hall, D. O., J. Coombes and T. W. Goodwin). pp83-93. The Biochemical Society.

Drews, G., and R. Feick. 1978. **Isolation and characterisation of light-harvesting bacteriochlorophyll protein complexes from *Rhodopseudomonas capsulata*.** Ann. Microbiol.(Paris) 134B:151-158.

Drews, G. and J. Oelze. 1981. **Organisation and differentiation of membranes of phototrophic bacteria.** Adv. Micro. Physiol. 22: 1-92.

Drews, G. 1985. **Structure and functional organisation of light-harvesting complexes and photochemical reaction centres in membranes of phototrophic bacteria.** Microbiological Reviews. 49: 59-70.

Drews, G., 1986 **Adaptation of the bacterial photosynthetic apparatus to different light intensities.** Trends in Biochemical Sciences 11 pp255-257.

Drews, G. 1989. **Regulation of formation of light harvesting complexes in *Rhodobacter capsulatus*.** In Abstracts of Symposium of Molecular Biology of Membrane-Bound Complexes in Phototrophic Bacteria. Freiburg im Breisgau.1989. p18.

Eaglesham, A. R. J., and R. J. Ellis. 1974. **Protein synthesis in chloroplasts II. Light driven synthesis of membrane proteins by isolated pea chloroplasts.** Biochim. Biophys. Acta 335: 396-407.

Edelman, M., and A. Reisfeld. 1980. In "Genome Organisation and Expression in Plants" (Ed. Leaver, C.J.) pp.353-362. Plenum Press, New York.

Ehrlich, M and R. Y. A. Wang. 1981. **5-methylcytosine in eukaryotic DNA.** Science 212: 1350-1357.

Eimhejellen, K. E., O. Aasmundrud and A. Jenson. 1963. **A new bacterial chlorophyll.** Biochem. Biophys. Res. Commun. 10: 232-242.

Englehardt, H., W. Baumeister, and W. O. Saxton. 1983. **Electron microscopy of photosynthetic membranes containing bacteriochlorophyll b.** Arch. Microbiol. 135: 169-175.

Evans, M. B. 1989. **The structure and function of the light-harvesting antenna complexes from purple photosynthetic bacteria.** PhD. Thesis. Department of Botany. University of Glasgow.

Enquist, L., and A. Skalka. 1973. **Replication of lambda DNA dependent on the function of host and viral genes.** J. Mol. Biol. 75: 185.

Farchaus, J. W., H. Gruenberg and D. Oesterhelt. 1989. In Abstracts of Symposium of Molecular Biology of Membrane-Bound Complexes in Phototrophic Bacteria. Freiburg im Breisgau.1989. p26.

Farchaus, J. W. and D. Oesterhelt. 1989. **A *Rhodobacter sphaeroides puf* L, M and X deletion mutant and its complementation in trans with a 5.3kb *puf* operon shuttle fragment.** The EMBO Journal. 8: 47-54.

Feher, G., A. J. Hoff, R. A. Isaacson, L. C. Ackerson. 1975. **ENDOR experiments on chlorophyll and bacteriochlorophyll *in vivo* and in the photosynthetic unit.** Ann. N. Y. Acad. Sci. 244: 239-259.

Feher, G., and M. Y. Okamura. 1978. **Chemical composition and properties of reaction centres.** In The photosynthetic bacteria. (Eds. Clayton, R. K. and W. R. Sistrom). pp349-386. Plenum Press, New York.

Feher, G. and M. Y. Okamura. 1984. **Strucure and function of the reaction centre from *Rps. sphaeroides*.** In "Advances in Photosyntheis Research Volume II". (Ed. Sybesma, C.) pp 155-164. Nijkoff/Junk Publishers. The Hague. Boston. Lancaster.

Feick, R. and G. Drews. 1978. **Isolation and characterisation of light-harvesting bacteriochlorophyll: protein complexes from *Rhodopseudomonas capsulata*.** Biochim. Biophys. Acta. 501: 499-513.

Feinberg, A. P. and B. Vogelstein. 1983. **A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity.** Analytical Biochem. 132: 6-13.

Feiss, M., and A. Becker. 1983. **DNA packaging and cutting.** In "Lambda II" pp. 305-330. (Eds. Hendrix, R. W., J. W. Roberts and F. W. Stahl). Cold Spring Harbour, New York.

Fessler, L.I., N. P. Morris, and J. H. Fessler. 1975. **Procollagen: Biological scission of amino acid carboxyl extension peptides.** Proc. Natl. Acad. Sci. USA 72: 4905-4909.

FesslerL. I., R. Timpl, and J. H. Fessler. 1975. Proc. Natl. Acad. Sci. USA 256: 2531-2537.

*Fiers, W. 1979. Comprehensive Virology 13: 69-204.

Firsow, N. N. and G. Drews. 1977. **Differentiation of the intracytoplasmic membrane of *Rps. palustris* by variation of the oxygen partial pressure and light intensity.** Arch. Microbiol. 115: 299-306.

*Foote, C. S. and R. W. Denny. 1968. Ann. Chem. Soc. 90: 6233.

Francis, G. A. and W. R. Richards. 1980. **Localisation of photosynthetic membrane components of *Rhodopseudomonas sphaeroides* by a radioactive labelling procedure.** Biochemistry. 19: 5104-5111.

Frischauf, A., H. Lehrach, A. Poustka, and N. Murray. 1983. **Lambda replacement vectors carrying polylinker sequences.** J. Mol. Biol. 170: 827-842.

*Fuglistaller, P., R. Rumbeli, F. Suter and H. Zuber. 1984. Phsiol. Chem. 365: 675-

Furth, M. E. , and S. H. Wickner. 1983. **Lambda DNA replication.** In "Lambda II". (Eds. Hendrix, R. W., J. W. Roberts, and F. W. Stahl). Cold Spring Harbour Laboratory. New York.

Gabellini, N., U. Harisch, J. E. McCarthy, G. Havska and W. Sebald. 1985. **Cloning and expression of the *fbc* operon encoding the FeS protein cytochrome b and cytochrome c1 from *Rhodospseudomonas sphaeroides* b/c1 complex.** The EMBO Journal. 2: 549-553.

Gabellini, N. and W. Sebald. 1986. **Nucleotide sequence and transcription of the *fbc* operon from *Rhodopseudomonas sphaeroides*.** Eur. J. Biochem. 154: 569-579.

Gamper, H. B. and J. E. Hearst. 1982. **A topological model for transcription based on unwinding angle analysis of *E. coli* RNA polymerase binary, initiation and ternary complexes.** Cell. 29: 81-90.

* Title not available.

Gellert, M. 1981. **DNA topoisomerases.** Ann. Rev. Biochem. 50: 879-910.

Gerber, G. E., C. P. Gray, D. Wildenauer and H. G. Khorana. 1977. **Orientation of bacteriorhodopsin in *Halobacterium halobium* as studied by selective proteolysis.** Proc. Natl. Acad. Sci. USA 74: 5426-5430.

Giaever, G. N. and J. C. Wang. 1988. **Supercoiling of intracellular DNA can occur in eukaryotic cells.** Cell. 55: 849-856.

Gibbons, N. E. and R. G. E. Murray. 1978. **Proposals concerning higher taxa of bacteria.** International Journal of Systematic Bacteriology. 28: 1-6.

Gogel, G. E., P. S. Parkes, P. A. Loach, R. A. Brunisholz and H. Zuber. 1983. **The primary structure of a light harvesting bacteriochlorophyll-binding protein of wild type *Rhodospiriluum rubrum*.** Biochim. Biophys. Acta 746: 32-39.

Gold, L., D. Pribnow, T. Schneider. 1981. **Translation initiation in prokaryotes.** Ann. Rev. Microbiol 35: 365-403.

Grebanier, A. E. , D. M. Coen, A. Rich, and L. Bogorad. 1978. **Membrane proteins synthesised but not processed by isolated maize chloroplasts.** J. Cell. Biol. 78: 734-746.

Gregory, W. S., and M. L. Mishkind. 1986. **The transport of proteins into chloroplasts.** Ann. Rev. Biochem. 55: 879-912.

van Grondelle, R., H. J. M. Kramer, and C. P. Rijgersberg. 1982. **Energy transfer in the B800-850 carotenoid light-harvesting complex of various mutants of *Rhodophseudomonas sphaeroides* and *Rhodopseudomonas capsulata*.** Biochim. Biophys. Acta 682: 208-215.

van Grondelle, R., C.N. Hunter, J. G. C. Bakker and H. J. Kramer. 1983. **Size and structure of antenna complexes of photosynthetic bacteria as studied by singlet-singlet quenching of the bacteriochlorophyll fluorescence yield.** Biochim. Biophys. Acta 723: 30-36.

Grunberg-Manago, M. 1980. In "Ribosomes" (Eds. Chambliss, G., G. R. Craven, J. Davies) pp. 447-477. University Park Press.

Gussin, G. N., C. W. Ronson and F. M. Ausubel. 1986. **Regulation of nitrogen fixation genes.** Ann. Rev. Genet. 20: 567-591.

Hayashi, H., M. Miyao, and S. Morita. 1982. **Absorption and fluoresence spectra of light-harvesting bacteriochlorophyll-protein complexes from *Rhodopseudomonas palustris* in the near infra-red region. J. Biochem.** 91: 1017-1027.

Heber, U. and D. A. Walker. 1979. **The Chloroplast envelope - barrier or bridge?** TIBS. 4: 252-256.

von Heijne, G. 1981. **On the hydrophobic nature of signal sequences.** Eur. J. Biochem. 116: 419-922.

Highfield, P. E., and R. J. Ellis. 1978. **Synthesis and transport of the small subunit of chloroplast ribulose bisphosphate carboxylase.** Nature 271: 420-24.

Hinkle, P. C. and R. E. McCarthy. 1978. **How cells make ATP.** Sci. Am. 238(3): 104-123.

Holmes, W.M., T. Platt and M. Rosenberg. 1983. **Termination of transcription in *E.coli*.** Cell. 32: 1029-1032.

Hunter, C. N., R. A. Niederman, and R. K. Clayton. 1981. **Excitation energy transfer by antenna complexes isolated from *Rps. sphaeroides* with lithium dodecyl sulphate/ polyacrylamide gel electrophoresis.**

In "Proceedings of the fifth international congress on photosynthesis". (Ed. Akoyunoglou, G) pp.539-545. Balaban Int. Sci. Serv. Phil. P. A.

Hunter, C. N., J. D. Pennoyer, and R. A. Neiderman. 1982. **Assembly and structural organisation of pigment-protein complexes in membranes of *Rps. sphaeroides*.** In "Proceedings of the special FEBS meeting on cell function and differentiation". (Ed. Evangelopoulos, E.) Alan R. Liss. Inc. New York.

Hunter, C. N., H. J. M. Kramer and R. van Grondelle. 1985. **Linear dichroism and fluorescence emission of antenna complexes during photosynthetic unit assembly in *Rps. sphaeroides*.** Biochim. Biophys. Acta 807: 44-51.

Hunter, N. 1988. **Molecules of the green machine.** New Scientist 118: 60-63.

Imhoff, J. F., H. G. Truper, and N. Pfennig. 1984. **Rearrangement of the Species and Genera of the Phototrophic "Purple Nonsulphur Bacteria".** International J. of Systematic Bacteriol. 34: 340-343.

Iserentant, D., and W. Fiers. 1980. **Secondary structure of mRNA and efficiency of translation initiation.** Gene 9: 1-12.

Jackson, W. J., P. J. Kiley, C. E. Haith, S. Kaplan and R. C. Prince. **On the role of the light-harvesting B880 in the correct insertion of the reaction centre of *Rhodobacter capsulatus* and *Rhodobacter sphaeroides*.** FEBS. Lett. 215: 171-174.

Jay, F., M. Lambillotte, W. Stark and K. Muhlethaler. 1984. **The preparation and characterisation of the thylakoids of *Rps. viridis*.** The EMBO J. 3: 773-776.

Johnson, J. A., W. K. R. Wang and J. T. Beatty. 1986. **Expression of cellulase genes in *Rhodobacter capsulatus* by use of plasmid expression vectors.** J. Bacteriol. 167: 604-610.

Jolcine, G., F. Reiss-Husson. 1974. **Comparative studies on the two reaction centre preparations from *Rps. sphaeroides* Y.** FEBS Letts. 40: 5-8.

Jolcine, G. F. Reiss-Husson. 1975. **Studies on pigments and lipids in *Rps. sphaeroides* Y reaction centres.** FEBS Letts. 52: 33-56.

Kaiser, K., and N. E. Murray. 1985. **The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries.** In "DNA cloning vol. I". (Ed. Glover, D. M.) pp.1-47. IRL Press. Oxford. Washington D.C.

Kaplan. S. 1978. Cited in: The Photosynthetic Bacteria Eds. Clayton, P. K. and W. R. Sistrom. Plenum New York p809-839.

Kaplan, S. and B. L. Marrs. 1986. **Proposed nomenclature for photosynthetic prokaryotes.** ASM News 52: 242.

Kaplan, S. 1989. In Abstracts of: Symposium on Molecular Biology of Membrane-Bound Complexes in Phototrophic Bacteria. p19. Freiburg im Breisgau 1989.

Kaufmann, N., H. H. Reidl, J. R. Golecki, A. F. Garcia and G. Drews. 1982. **Differentiation of the membrane system in cells of *Rhodopseudomonas capsulata* after transition from the chemotrophic to phototrophic growth conditions. Arch. Microbiol. 131: 313-322.**

Keister, D. L. 1978. Respiration vs photosynthesis in the photosynthetic bacteria. (Eds. Clayton, R. K. and W. R. Sistrom) pp849-856. Plenum Press New York England.

Kell, D. B. 1984. **Diffusion of protein complexes in prokaryotic membranes: Fast, Free, Random or Directed?** TIBS. 9: 86-88.

Kiley, P. J., and S. Kaplan. 1987. **Cloning, DNA sequence, and expression of the *Rhodobacter sphaeroides* light-harvesting B800-850α and B800-850β genes.** J. Bacteriol. 169: 3268-3275.

Kiley, P. J. and S. Kaplan. 1987. **Cloning, DNA sequence and expression of the *Rhodobacter sphaeroides* light harvesting B800-850-α and B800-850-β genes.** J. Bacteriol. 169: 3268-3275.

Kiley, P. J., T. J. Donohue, W. A. Havelka and S. Kaplan. 1987. **DNA sequence and *in vitro* expression of the B875 light-harvesting polypeptides of *Rhodobacter sphaeroides*.** J. Bacteriol. 169: 742-750.

Kiley, P.J. and S. Kaplan. 1988. **Molecular Genetics of Photosynthetic Membrane Biosynthesis in *Rhodobacter sphaeroides*.** Microbiological Reviews 52: 50-69.

Kiley, P. J., A. Varga and S. Kaplan. 1988. **Physiological and structural analysis of light harvesting mutants of *Rhodobacter sphaeroides*.** J. Bacteriol. 170: 1103-1115.

Kirk, J. T. O. 1983. **Light and photosynthesis in aquatic ecosystems.** Cambridge University Press. Cambridge.

Klug, G. and G. Drews. 1984. **Construction of a gene bank of *Rps. capsulata* using a broad host range DNA cloning system.** Arch. Microbiol. 39: 319-325.

Klug, G., N. Kaufmann and G. Drews. 1985. **Gene expression of pigment binding-proteins of the bacterial photosynthetic apparatus: Transcription and assembly in the membrane of *Rhodopseudomonas capsulata*.** Proc. Natl. Acad. Sci. USA. 82: 6485-6489.

Klug, G., R. Liebetanz and G. Drews. 1986. **The influence of bacteriochlorophyll biosynthesis on the formation of pigment-binding proteins and assembly of pigment protein complexes in *Rhodopseudomonas capsulata*.** Arch. Microbiol. 146: 284-291.

Klug, G., C. W. Adams, J. Belasco, B. Doerge and S. N. Cohen. 1987. **Biological consequences of segmental alterations in mRNA stability: effects of deletion of the intercistronic hairpin loop region of the *Rhodobacter capsulatus puf* operon.** The EMBO Journal. 6: 3515-3520.

Kramer, H. J. M., R. van Grondelle, C. N. Hunter, W. H. J. Westerhuis and J. Amez. 1984a. **Pigment organisation of the B800-850 antenna complex of *Rhodopseudomonas sphaeroides*.** Biochim. Biophys Acta. 765: 156-165.

Kramer, H. J. M., J. D. Pennoyer, R. van Grondelle, W. H. J. Westerhuis, R. A. Neiderman and J. Amesz. 1984b. **Low temperature optical properties and pigment organisation of the B875 light-harvesting bacteriochlorophyll-protein complex of the purple photosynthetic bacteria.** Biochim. Biophys. Acta 767: 335-344.

Kranz, R. G. and R. Hazelkorn. 1986. **Anaerobic regulation of nitrogen-fixation genes in *Rhodopseudomonas capsulatus*.** Proc. Natl. Acad. Sci. USA. 83: 6805-6809.

* Krinsky, N. I. 1979. Pure Applied Chemistry 51: 649-660.

Kyte, J., and R. F. Doolittle. 1982. **A simple method for displaying the hydropathic character of a protein.** J. Mol. Biol. 157: 105-32.

Lam, E. and N.-H. Chua. 1987. **Chloroplast DNA gyrase and *in vitro* refgulation of transcription.** Plant. Mol. Biol. 8: 415-424.

Lang, F. S. and D. Oesterhelt. 1989. **Gene transfer system for *Rhodopseudomonas viridis*.** J. Bacteriol. 171: 4425-4435.

* Title not available.

Leigh, J. S. 1978. **EPR studies of primary events in bacterial photosynthesis.** In "The Photosynthetic Bacteria". (Eds. Clayton, R. K. and W. R. Sistrom) Plenum Press, New York. pp. 431-436.

Lien, S. and H. Gest. 1973. **Chlorophyll synthesis in photosynthetic bacteria.** J. Bioenergetics. 4: 423-434.

Lien, S., H. Gest and A. San Pietro. 1973. **Regulation of chlorophyll synthesis in photosynthetic bacteria.** Bioenergetics. 4: 423-434.

Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. **Molecular Cloning: A Laboratory Manual.** Cold Spring Harbour Laboratory, Box 100, Cold Spring Harbour, New York.

Marder, J. B., P. Goloubinoff, and M. Edelman. 1984. **Molecular architecture of the rapidly metabolized 32-kilodalton protein of Photosystem II.** J. Biol. Chem. 259: 3900-3908.

Marrs, B. and H. Gest. 1973. **Regulation of bacteriochlorophyll synthesis by oxygen in mutants of *Rhodopseudomonas capsulata*.** J. Bacteriol. 114: 1052-1057.

Marrs, B. 1981. **Mobilization of the genes for photosynthesis from *Rhodopseudomonas capsulata* by a promiscuous plasmid.** J. Bacteriol. 146:1003-1012.

Marrs, B. L., D. A. Young, C. E. Bauer, J.-A. C. Williams. 1989. **The regulation of bacteriochlorophyll biosynthesis in *Rhodobacter capsulatus*.** In Abstracts of: Symposium on Molecular Biology of Membrane-Bound Complexes in Phototrophic Bacteria. p119.

Mason, P. J. and J. G. Williams. 1985. **Hybridization in the analysis of recombinant DNA.** pp. 113-137. In "Nucleic Acid Hybridization: A Practical Approach". (Eds. Hames, B. D. and S. J. Higgins) IRL Press, Oxford. Washington.

Mattoo, A. K., H. Hoffman-Falk, J. B. Marder, and M. Edelman. 1984. **Regulation of protein metabolism: Coupling of photosynthetic electron transport to *in vivo* degradation of the rapidly metabolized 32kD protein of the chloroplast membranes.** Proc. Natl. Acad. Sci. USA. 81: 1380-1384.

Mattoo, A. K., U. Pick, H. Hoffman-Falk, and M. Edelman. 1981. **The rapidly metabolised 32kD polypeptide of the chloroplast is the "proteinaceous shield" regulating photosystem II electron transport and mediating diuron herbicide sensitivity.** Proc. Natl. Acad. Sci. USA 78: 1572-1576.

McElroy, J. D., G. Feher, and D. C. Mauzerall. 1969. **On the nature of the free radical formed during the primary process of bacterial photosynthesis.** Biochim. Biophys. Acta 172: 180-183.

Meinhardt. S. W., P. J. Kiley, S. Kaplan, A. R. Crofts and S. Harayama. 1985. **Characterisation of light harvesting mutants of *Rhodobacter sphaeroides* .1. Measurement of the efficiency of energy transfer from light harvesting complexes to the reaction centre.** Arch. Biochem. Biophys. 236: 130-139.

Menzel, R. J. and M. Gellert. 1983. **Regulation of the genes for *E.coli* DNA gyrase: Homeostatic control of DNA supercoiling.** Cell. 34: 105-113.

Menzel, R. J. and M. Gellert. 1987. **Modulation of transcription by DNA supercoiling: A deletion analysis of the *E.coli gyr*A and *gyr*B promoters.** Proc. Natl. Acad Sci. USA. 84: 4185-4189.

Messing, J., B. Gronenborn, B. Muller-Hill, and P. H. Hofschneider. 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 75: 3642.

Messing, J., and J. Vieira. 1982. **A new pair of M13 vectors for selecting either strand of a double-digest restriction fragnent.** Gene 19: 269.

Messing, J. 1983. **New M13 vectors for cloning.** Methods Enzymol. 101: 20.

Michel, H. 1982. **Three dimensional crystals of a membrane protein complex: The photosynthetic reaction centre from *Rps. viridis*.** J. Mol. Biol. 158: 567-572.

Michel, H. 1983. **Crystallisation of membrane proteins.** Trends in biochemical science. 8: 56-59.

Michel, H. and J. Deisenhofer. 1986. **X-ray diffraction studies on a crystalline bacterial photosynthetic reaction centre.** In "Photosynthesis III: Photosynthetic membranes and light harvesting systems". Encyclopedia of plant physiology. New Series. Vol 19. (Eds. Staehelin, L. A. and C. J. Arntzen).

Michel, H., K. A. Weyer, H. Gruenberg, I. Dunger, D. Oesterhelt and F. Lottspeich. 1986. **The "light" and "medium" subunits of the photosynthetic reaction centre from *Rhodopseudomonas viridis*: isolation of the genes, nucleotide and amino acid sequence.** The EMBO Journal 5: 1149-1158.

Miller, K. R. 1979. **The photosynthetic membrane.** Sci. Am. 241(4): 102-113.

Miller, K. R. 1982. **Three dimensional structure of a photosythetic membrane.** Nature 300: 53-55.

Miller, K. R. and J. S. Jacob. 1983. **Two dimensional crystals formed from photosythetic reaction centres.** J. Cell Biol. 97: 1266-1270.

Mitchell, P. 1961. **Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism.** Nature 191: 144-148.

Molisch. 1907. **Die purpurbakterien nach neuen untersuchungen.** Fisher Verlag. Jena: 95.

Monger, T. G. and W. W. Parson. 1977. **Singlet-Triplet fusion in *Rps. sphaeroides* chromatophores; a probe of the organisation of the photosynthetic apparatus.** Biochim. Biophys. Acta 460: 393-407.

Murray, N. E. , and K. Murray. 1974. **Manipulation of receptor targets in phage λ to form receptor chromosomes for DNA fragments.** Nature 251: 476.

Nabedryk, E., D. M. Tiede, P.L. Dutton and J. Bretton 1982. **Conformation and orientation of the protein in the bacterial photosynthetic reaction centre.** Biochem. Biophys Acta 682: 273-280.

Niederman, R. A., D. E. Mallon and J. J. Langan. 1976. **Membranes of *Rhodopseudomonas sphaeroides*. IV. Assembly of chromatophores in low aeration cell suspensions.** Biochim. Biophys. Acta 440: 429-447.

Netzel, T. L., P. M. Rentzepis, and J. Leigh. 1973. **Picosecond kinetics of reaction centres containing BChl.** Science 182: 238-241.

van Niel, C. B. 1931. **On the morphology and physiology of the purple and green sulphur bacteria.** Arch. Mikrobiol. 3: 1-112.

van Niell, C. B. 1941. **The bacterial photosyntheses and their importance for the general problem of photosynthesis.** Advances in Enzymology 1: 263-328.

van Niel, C. B. 1944. **The culture, general physiology, morphology and classification of the non-sulphur purple and brown bacteria.** Bacteriological reviews. 8: 1-118.

van Niel C. B. 1957. In: Bergeys manual of determinative bacteriology. 7th Edition p35. Eds. Breed, R. S., E. G. D. Murray and N. R. Smith. Pub. Williams and Wilkins Baltimore Maryland.

Norrander, J., T. Kempe, and J. Messing. 1983. **Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis.** Gene 26: 101.

Norris, J. R., R. A. Uphaus, H. L. Crespi, and J. J. Katz. 1971. **Electron spin resonance of chlorophyll and the origin of signal I in photosynthesis.** Proc. Natl. Acad. Sci. USA. 68: 625-628.

Norris, J. R., M. E. Druyan and J. J. Katz. 1973. **Electron nuclear double resonance of bacteriochlorophyll free radical *in vivo* and *in vitro*.** J. Am. Chem. Soc. 95: 1680-1682.

Novak, P. D. and R. J. Maier. 1987. **Inhibition of hydrogenase synthesis by DNA gyrase inhibitors in *Bradyrhizobium japonicum*.** J. Bacteriol. 169: 2709-2712.

Oberle, B., H.-V. Tichy and G. Drews. 1988. **B800-850 specific gene expression in *Rhodobacter capsulatus*.** In Abstracts of: VI International Symposium on Photosynthetic Prokaryotes. Noordwijkerhout. The Netherlands 1988. p202.

Okamura, M. Y., L. A. Steiner, and G. Feher. 1974. **Characterisation of reaction centres from photosynthetic bacteria. I. Subunit structure of the protein mediating the primary photochemistry in *Rps. sphaeroides*. R-26. Biochemistry. 13: 1394-1402.**

Okamura , M. Y., G. Feher and N. Nelson. 1982. Reaction Centres. In "Photosynthesis volume I : Energy conversion by plants and bacteria. (Ed. Govinjee). pp195-272. New-York. Academic Press.

Olson, J. M. 1980. **Chlorophyll organisation in green photosynthetic bacteria.** Biochim. Biophys. Acta. 594: 33-51.

Otter, R. and N. R. Cozzarelli. 1983. ***E. coli* DNA gyrase.** Methods in Enz. 100: 171-180.

Parson, W. W. and B. Ke. 1982. **Primary Photochemical Reactions.** pp331-386. In Govindjee (Ed.) Photosynthesis volI. Academic Press Inc., New York.

Parson, W. W., and D. Holten. 1986. **Primary electron transfer reactions in photosynthetic bacteria: Energetics and kinetics of transient states.** In "Photosynthesis III: Photosynthetic membranes and light harvesting systems". (Eds. Staehelin, L.A. and C. J. Arntzen). Encyclopedia of plant physiology Vol. 19. Springer-Verlag.

Patzelt, C., H. S. Tager, R. J. Carroll, and D. F. Steiner. 1979. **Identification and processing of proglucagon in pancreatic islets.** Nature 282: 260-266.

Pemberton, J. M., and C. M. Harding. 1986. **Cloning of carotenoid biosynthesis genes from *Rhodopseudomonas sphaeroides*.** Curr. Microbiol. 14: 25-29.

Peters, J. and G. Drews. 1983. **The B870 pigment-protein complex of *Rhodopseudomonas capsulata* contains two different pigment-binding polypeptides.** FEMS. Microbiol. Lett. 17: 235-237.

Peter, J., J. Takemoto and G. Drews. 1983. **Spatial relationships between the photochemical reaction centre and the light harvesting complexes in the membrane of *Rps. capsulata*.** Biochemistry 22: 5660-5667.

Peters, J., W. Welte and G. Drews. 1984. **Topographical relationships of polypeptides in the photosynthetic membrane of *Rhodopseudomonas viridis* investigated by reversible chemical cross-linking.** FEBS. Lett. 171: 267-270.

Pettijohn, D. E. and O. Pfenninger. 1980. **Supercoils in prokaryotic DNA restrained *in vivo*.** Proc. Natl. Acad. Sci. USA. 77: 1331-1335.

Pfennig, N. 1967. **Photosynthetic bacteria.** Ann. Rev. Microbiol. 21: 285.

Pfennig, N. 1969. ***Rhodopseudomonas acidophila* s.p.n., A new species of the budding purple nonsulphur bacteria.** J. Bacteriol. 99: 597-602.

Pfennig, N. and H. G. Truper. 1971. **New nomenclature combinations in the phototrophic sulphur bacteria.** Int. J. Syst. Bacteriol. 21: 11-24.

Pfennig, N. and H. G. Truper. 1973. **The Rhodospirillales (Phototrophic or Photosynthetic Bacteria).** Handbook of Microbiology VolI. Organismic Microbiology. Eds. Laskin, A.I. and A. Lechevalier. Pub. CRC Press.

Pfennig, N. and H. G. Truper. 1974. **The phototrophic bacteria.** In: Bergeys manual of determinative bacteriology 8th Edition. (Eds. Buchanan, R. E. and N. E. Gibbons) pp24-64. Pub. The William and Williams Co. Baltimore.

Pfennig, N. 1978. **General physiology and ecology of photosynthetic bacteria.** In: The photosynthetic Bacteria (Eds. Clayton, R. K. and W. R. Sistrom pp3-14. Plenum Press, New York.

Pfennig, N. and H. G. Truper. 1981. **Isolation of members of the families Chromatiaceae and Chlorobiaceae.** In: The Prokaryotes-A handbook of habitata, isolation and identification of bacteria. volI. (Eds. Starr, M. P., H.Stolp, H. G. Truper, A. Balows and H. G. Schlegel. pp274-279. Springer-Verlag. Berlin, Heidelberg, New York.

Picorel, R., G. Belanger, and G. Gingras. 1983. **Antenna holochrome B880 of *Rhodospirillum rubrum* S1. Pigment, phospholipid and polypeptide composition.** Biochemistry. 22: 2491-2497.

Pradel, J., J. Lavergne and I. Moya. 1978. **Formation and development of photosynthetic units in repigmenting *Rps. sphaeroides* wild type and 'Phofil' mutant strain.** Biochim. Biophys. Acta. 502: 169-182.

Prasad, R. V., R. J. Butowski, J. W. Hamilton, and K. E. Ebner. 1982. **Amino acid sequence of rat α-lactalbumin: A unique α-lactalbumin.** Biochemistry 21: 1479-1482.

Prince, R. C. and P. L. Dutton. 1978. **Protonation and reducing potential of the primary electron acceptor.** In "The Photosynthetic Bacteria". (Eds. Clayton, R. K. and W. R. Sistrom). Plenum press New York. pp. 439-450.

Racker, E. and W. Stoeckenius. 1974. **Reconstitution of purple membrane vesicles catalysing light-driven proton uptake and adenosine triphosphate formation.** J. Biol. Chem. 249: 662-663.

Racker, E. 1976. **Structure and function of ATP-driven ion pumps.** TIBS. 1: 244-247.

Raleigh, . 1987. **Host strains that alleviate underrepresentation of specific sequences: Overview.** Methods in Enzymology 152: 173-180.

Raleigh, E. A., and G. Wilson. 1986. Proc. Natl. Acad. Sci. USA 83: 9070-9074.

Rambach, A., and P. Tiollais. 1974. **Bacteriophage λ having EcoRI endonuclease sites only in the non-essential region of the genome.** Proc. Natl. Acad. Sci. USA 71: 3927.

Reed, D. W. and R. K. Clayton. 1968. **Isolation of a reaction centre fraction from *Rhodopseudomonas sphaeroides*.** Biochem. Biophys. Res. Comm. 30: 471-775.

Reitzer, L. J. and B. Magasanaik. 1986. **Transcription of *glnA* in *E.coli* is stimulated by activator bound to sites far from the promoter. Cell. 45: 785-792.**

Robertson, D. E. and D. L. Dutton. 1988. The nature and magnitude of the charge separation reactions of ubiquinone, cytochrome C_2 oxidoreductase. Biochim. Biophys. Acta 935: 273-291.

Sain, B., and N. E. Murray. 1980. **The *hsd* (host specificity) genes of *E.coli* K12.** Mol. Gen. Genet. 180: 35-46.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. **DNA sequencing with chain terminating inhibitors.** Proc. Natl. Sci. USA 74: 5463.

Sauer, K. 1978. **Photosynthetic membranes.** Accounts of chemical research. 11: 257-264.

Sauer, K. and L. A. Austin. 1978. **Bacteriochlorophyll-protein complexes from the light-harvesting antenna of photosynthetic bacteria.** Biochemistry 17: 2011-2019.

Schleif, R. F., and P. C. Wensink. 1981. In "Practical Methods in Molecular Biology" Springer-Verlag. New York, Heidelburg, Berlin.

Schmidt, K., R. Brunisholz and H. Zuber. 1987. **Unpublished. Cited by Brunisholz, R. (1987) Structural aspects of the antenna polypeptides from *Rhodopseudomonas acidophila* and *Rhodocyclus gelatinosus*.** In Abstracts of: Structure, Function and Formation of membrane bound complexes in phototrophic bacteria. pp9. Freiburg im Breisgau. Fed. Rep. Germany. 1987.

Schumacher. A. and G. Drews. 1978. **The formation of bacteriochlorophyll-protein complexes of the photosynthetic apparatus of *Rps. capsulata* during early stages of development.** Biochim Biophys. Acta. 501: 183-194.

Scolnik, P. A., M. A. Walker and B. L. Marrs. 1980. **Biosynthesis of carotenoids derived from neurosporene in *Rhodopseudomonas capsulata*.**J. Biol. Chem. 255: 2427-2432.

Scolnik, P. A. and B. L. Marrs. 1987. **Genetic research with photosynthetic bacteria.** Ann. Rev. Microbiol. 41: 703-726.

Sharp, P.A. , B. Sugden and J. Sambrook. 1973. **Detection of two restriction endonucleases activities in *Haem. parainfluenzae* using analytical agarose.** Biochem. 12: 3055.

Shine J., and L. Dalgarno 1974. **Identical 3'-terminal octanucleotide sequences in 18S ribosomal RNA from different eukaryotes.** Biochem. J. 141: 609-15.

Shiozwa, J. A., P. A. Cuendet, G. Drews and H. Zuber. 1980. **Isolation and characterisation of the polypeptide components from light harvesting pigment protein complex B800-850 of *Rhodopseudomonas capsulata*.** Eur. J. Biochem. 111: 455-460.

Shiozawa, J. A., W. Welte, N. Hodapp, and G. Drews. 1982. **Studies on the size and composition of the isolated light harvesting B800-850 pigment-protein complex of *Rhodopseudomonas capsulata*.** Arch. Biochem. Biophys. 213: 473-485.

Sinden, R. R. and D. E. Pettijohn. 1981. **Chromosomes in living *E. coli* cells are segregated into domains of supercoiling.** Proc. Natl. Acad. Sci. USA. 78: 224-228.

Sistrom, W. R. 1978. In: The photosynthetic bacteria. Clayton, R. K. and W. R. Sistrom (Eds.) Plenum, New York London. pp841-848.

Sistrom, W. R., A. Macalusa, and R. Pledger. 1984. **Mutants of *Rhodopseudomonas sphaeroides* useful in genetic analysis.** Arch. Microbiol 138:161-165.

Smith and Waterman. 1981. **Advances in Applied Mathematics** 2:482-489.

Southern, E. M. 1975. **Detection of specific sequences among DNA fragments separated by gel electrophoresis.** J. Mol. Biol. 98: 503-17

Sprague, S. G., L. A. Staehelin, M. J. DiBartolomeis and R. C. Fuller. 1981. J. Bacteriol. 147: 1021-1031.

Sprague, S. G. and R. A. Varga. 1986. **Topography, composition and assembly of photosynthetic membranes.** In: Photosynthesis III, Photosynthetic membranes and light harvesting systems, Encyclopedia of plant physiology. New series vol19 . Ed. Stahelin, L. A. and C. J. Arntzen. Pub Springer-Verlag.

Stark, W., W. Kuhlbrandt, I. Wildhaber. 1984. **The structure of the photoreceptor unit of *Rps. viridis*.** The EMBO J. 3: 777-783.

Stirdivant, S. M., L. D. Crossland and L. Bogorad. 1985. **DNA supercoiling affects *in vitro* transcription of two maize chloroplast genes differentially.** Proc. Natl. Acad. Sci. USA. 82: 4886-4890.

Staehelin, L. A., J. R. Golecki and G. Drews. 1980. Biochim. et Biophys. Acta. 589: 30-45.

Stanier, R. Y. 1970. In: The Twentieth Symposium of the society for general microbiology. p1. Cambridge University Press. Cited in "Handbook of Microbiology, volI" 1973. Ed. Lechevalier, L.) pp17 CRC. Press.

Stanier, R. Y., M. Doudoroff and E. A. Adelberg. 1975. In "General Microbiology". Pub. Macmillan Press Ltd. pp559-577.

Stanier, R. Y., E. A. Adelberg, and J. L. Ingraham (Eds) 1976. **The photosynthetic prokaryotes.** In: "General Microbiology" Fourth Edition. pp527-563. The MacMillan Press Ltd. London.

Stanier, R. Y., N. Pfennig and H. G. Truper. 1981. **Introduction to photosynthetic prokaryotes.** In: The Prokaryotes-A handbook of habitats, isolation and identification of bacteria. VolI. (Ed. Starr, M. P., H. Stolp, H. G. Truper, A. Balous and H. G. Schlegel) pp195-121. Springer-Verlag. Berlin. Heidelberg. New York.

Steitz, J.A. 1979. In " Ribosomes" (Ed. G. R. Davies) pp47-495.

Stormo, G. D. , T. D. Schneider and L. M. Gold. 1982. **Characterisation of translational initiation sites in *E.coli*.** N.A. R. Vol.10. No.9. pp2971

Straley. S.C., W. W.Parson., D.C. Mauzerall., R.K. Clayton. 1973. **Pigment content and molar extinction coefficients of photochemical reaction centres form *Rps. sphaeroides*.** Biochim. Biophys. Acta 548: 597-609.

Tadros, M. H., F. Suter, G. Drews, and H. Zuber. 1983. **The complete amino acid sequence of the large bacteriochlorophyll binding polypeptide from light-harvesting complex II (B800-850) of *R. capsulata*.** Eur. J. Biochem. 129: 533-536.

Tadros, M. H., F. Suter, H. H. Seydewitz, I. Witt, H. Zuber and G. Drews. 1984. **Isolation and complete amino acid sequence of the small polypeptide from light-harvesting pigment-protein complex I (B870) of *Rhodopseudomonas capsulata*.** Eur. J. Biochem. 138: 209-212.

Tadros, M. H., R. Frank, and G. Drews. 1985. **The complete amino-acid sequence of the small bacteriochlorophyll-binding polypeptide B800-850β from light harvesting complex B800-850 of *Rhodopseudomonas capsulata*.** FEBS. Lett. 183: 91-94.

Tadros, M. H., R. Cortese, and G. Drews. 1987. **The polypeptides of the light harvesting complex of *Rhodopseudomonas palustris*: DNA and amino acid sequences.** In Abstracts of: Structure, function and formation of membrane bound complexes in phototrophic bacteria pp56. Freiburg. Fed. Rep. Germany. 1987.

Tadros, H. B. and K. Waterkamp. 1989. **Multiple copies of the coding regions for the light harvesting B800-850 α- and β-polypeptides are present in the *Rhodopseudomonas palustris* genome.** The EMBO Journal 8: 1303-1308.

Tager, H. S., and D. F. Steiner. 1973. **Isolation of a glucagon-containing peptide: Primary structure of a possible fragment of proglucagon.** Proc. Natl. Acad. Sci. USA 70: 2321-2325.

Takemoto, J. 1974. **Kinetics of photosynthetic protein assembly in *Rps. sphaeroides*.** Arch. Biochim. Biophys. 163: 515-520.

Takemoto, J., J. Peters and G. Drews. 1982. **Cross linking of photosynthetic membrane polypeptides of *Rps. capsulata*.** FEBS. Letts. 142: 227-230.

Taylor, D. P., S. N. Cohen, W. G. Clark, and B. L. Marrs. 1983. **Alignment of genetic and restriction maps of the photosynthesis region ofthe *Rhodopseudomnas capsulata* chromosome by a conjugation-mediated marker rescue technique.** J. Bacteriol. 154: 580-590.

Theiler, R., R. Brunisholz, G. Frank, F. Suter and H. Zuber. 1982. IVth International Symposium Photosynthetic Prokaryotes. Bombannes, France, C-40.

Theiler, R., F. Suter and H. Zuber. 1983. **N-Terminal sequences of subunits L and M of the photosynthetic reaction centre from *Rhodospirillum rubrum* G-9-+.** Hoppe-Seylers Z. Physiol. Chem. 365: 689-701.

Theiler, R. and H. Zuber. 1984. **The light harvesting polypeptides of *Rhodopseudomonas sphaeroides* R26.1. II Conformational analysis by attenuated conformational analysis by attenuated total reflection infra-red spectroscopy and possible structure of the hydrophobic domain of the B850 complex.** Hoppe-Seylers Z. Physiol. Chem. 364: 1765-1776.

Theiler, R., F. Suter, V. Wiemken, and H. Zuber. 1984a. **The light harvesting polypeptides of *Rhodopseudomonas sphaeroides* R-26.1.** Hoppe-Seylers Z. Physiol. Chem. 365: 689-701.

Theiler, R., F. Suter, H. Zuber and R. J. Cogdell. 1984b. **A comparison of the primary structure of the two B800-850-apoproteins from the wild type *Rhodopseudomonas sphaeroides strain* 2.4.1 and a carotenoidless mutant strain R-26.1.** FEBS. Lett. 175: 231-237.

Thomas, M., J. R. Cameron, and R. W. Davis. 1974. **Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA.** Proc. Natl. Acad. Sci. USA 71: 4579.

Thompson, R. J. and G. Mossig. 1987. **Stimulation of a *Chlamydomonas* chloroplast promoter by novobiocin *in situ* and in *E. coli* implies regulation by torsional stress in the chloroplast DNA.** Cell. 48: 281-287.

Thornber, J. P. 1970. **Photochemical reactions of purple bacteria as revealed by studies of three spectrally different carotenobacteriochlorophyll-protein complexes isolated from *Chromatium* strain D.** Biochemistry. 9: 2688-2697.

Thornber, J. P., T. L. Trosper and C. E. Strouse. 1978. **Bacteriochlorophyll *in vivo*: Relationship of spectral forms to specific membrane components.** In "The photosynthetic bacteria" (Ed. Clayton, R. K. and W. R. Sistrom). pp133-160. Plenum Press. New York England.

Thornber, J. P., R. J. Cogdell, R. E. B. Seftor, B. K. Pierson and E. M. Tobin. 1983. **Comparative biochemistry of chlorophyll-protein complexes.** Advances in photosynthesis research. In Proceedings of VIth International Congress on photosynthesis 1983. Martinus Nijhoff/DR W. JUNK Publishers. The Hague. The Netherlands.

Thornber, J. P., R. J. Cogdell, B. K. Pierson and R. E. B. Seftor. 1984. **Pigment protein complexes of purple photosynthetic bacteria: An overview.** J. Cell Biochem. 23: 159-169.

Thornber, J. P. 1986. In: Photosynthesis III. **Photosynthetic membranes and light harvesting systems.** Encyclopedia of plant physiology. New Series vol.19. (Eds. Staehelin, L. A. and C. J. Arntzen). Springer-Verlag. Berlin, Heidelberg, New York, Tokyo. p98-142.

Thorne, H. V. 1966. **Electrophoretic separation of polyoma virus DNA from host cell DNA.** Virology 29: 234.

Tichy, H.-V., B. Oberle and G. Drews. 1988. **Structure and function of DNA sequences upstream and downstream from the B800-850 structural genes of *Rhodobacter capsulata*.** In Abstracts of: VI International Symposium on Photosynthetic Prokaryotes. Noordwijkerhout. The Netherlands. 1988.

Truper, H. G. 1976. **Higher taxa of the phototrophic bacteria Chloroflexaceae fam. nov. a family for the gliding filamentous phototrophic bacteria.** Int. J. Syst. Bacteriol. 26: 74.

Truper, H. G. and N. N. Pfennig. 1978. **Taxonomy of the Rhodospirillales. In: The phototrophic bacteria.** Eds. Clayton, R. K. and W. R. Sistrom pp19-27. Plenum Press New York London.

Trebst. A. 1974. **Energy conservation in photosynthetic electron transport of chloroplasts.** Ann. Rev. Plant Physiol., 25: 423-458.

Tsao, Y.-P., H.-C. Wu and L. F. Liu. 1989. **Transcription driven supercoiling of DNA: direct biochemical evidence from *in vitro* studies.** Cell 56: 111-118.

Ueda, T., Y. Morimoto, M. Sato, T. Kakuno, J. Yamashita and T. Horio. 1985. **Isolation, characterisation, and comparison of a ubiquitous pigment-protein complex.** J. Biochem. 98: 1487-1498

Varga, A. R. and L. A. Staehelin. 1985. **Pigment-protein complexes from *Rps. palustris*: isolation, characterisation and reconstitution into liposomes.** J. Bacteriol. 161: 921-927.

Vos, M., R. van Grondelle, and F. W. van der Kooij. 1986. **Singlet-singlet anhiliation at low temperatures in the antenna of purple bacteria.** Biochim., Biophys. Acta 850: 501-512.

Vogelstein, B. and D. Gillespie. 1979. **Preparative and analytical purification of DNA from agarose.** Proc. Natl. Acad. Sci. USA 76: 615-619

Vredenberg, W. J. and L. N. M. Duysens. 1963. **Transfer of energy from bacteriochlorophyll to reaction centre during bacterial photosynthesis.** Nature 197: 355-357.

Wahl, G. M., M. Stern, and G. R. Stark. 1979. **Efficient transfer of large DNA fragmentsd from agarose gels to diazobenzloxymethal-paper and rapid hybridisation by using dextran sulphate.** Proc. Natl. Acad. Sci. USA 76: 3683.

Walter, J. E., A. F. Carne, and H. W. Schmitt. 1979. **Topography of the purple membrane.** Nature 278: 653-654.

Wang, J. 1985. **DNA topoisomerases.** Ann. Rev. Biochem. 54: 665-697.

Westerhuis, W. H. J., J. Amesz and R. A. Neiderman. 1987. **Structural organisation of photosynthetic units in phospholipid-enriched**

membranes of *Rb. sphaeroides* as determined by singlet-singlet annihilation. In "Workshop-structure, function and formation of membrane bound complexes in phototrophic bacteria". pp35. Freiburg. FRG.

Whittaker, P. A., A. J. B. Campbell, E. M. Southern and N. E. Murray. 1989. **Enhanced recovery and restriction mapping of DNA fragments cloned in a new λ vector.** Nucleic Acids Research

Wickner, W. 1979. **Assembly of proteins into biological membranes.** Ann . Rev. Biochem. 48:23-45

Wickner, W. 1980. **Assembly of proteins into membranes.** Science 210: 861-868.

Williams, J. C., L. A. Steiner, R. C. Ogden, M. I. Simon and G. Feher. 1983. **Primary structure of the M subunit of the reaction center from *R. sphaeroides*.** Proc. Natl. Acad. Sci. USA. 80: 6505-6509.

Williams, J. C., L. A. Steiner, G. Feher. 1984. **Primary structure of the L-subunit of the reaction centre of *Rps. sphaeroides*.** Proc. Nat. Acad. Sci. USA. 81: 7303-7308.

Williams, J. C., L. A. Steiner and G. Feher. 1986. **Primary structure of the reaction centre from *Rhodopseudomonas sphaeroides*.** Proteins 1: 312-325.

Williams, J. A. 1989. In Abstracts of the: Symposium on Molecular Biology of Membrane-Bound Complexes in Phototrophic Bacteria. p9. Freiburg im Breisgau 1989.

Worcel. A. and E. Bugi. 1972. **On the structure of the folded chromosome of *Escherichia coli*.** J. Mol. Biol. 71: 127-147.

Wu, H.-Y., S. Shyy, J. C. Wang and L. F. Liu. 1988. **Transcription generates positively and negatively supercoiled domains in the template.** Cell. 53: 433-440.

Yamamoto, K. and M. L. Droffner. 1985. **Mechanisms determining aerobic or anaerobic growth in the facultative anaerobe *Salmonella typhimurium*.** Proc. Natl. Acad. Sci. USA. 82: 2077-2081.

Yanisch-Perron, C., J. Viera, and J. Messing. 1985. **New M13 host strains and the complete sequences of M13mp and pUC vectors.** Gene 33: 103.

Yeates, T. O., H. Korniya, D. C. Rees, J. P. Allen and G. Feher. 1987. **Structure of the reaction centre from *Rb. sphaeroides* R-26: membrane-protein interactions.** Proc. Natl. Acad. Sci. USA. 84: 6438-6442.

Yen, H. C. and B. Marrs. 1976. **Maps of Genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*.** J. Bacteriol. 126:619-629.

Youvan, D. C. and B. L. Marrs. 1984. **Molecular Genetics and the light reactions of photosynthesis.** Cell. 39: 1-3.

Youvan, D. C., E. J. Bylina, M. Albertie, H. Begusch and J. E. Hearst. 1984a. **Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-centre, B870 antenna and flanking polypeptides from *R. capsulata*.** Cell. 37:949-957.

Youvan, D. C., M. Alberti, H. Belgusch, E. J. Bylina and J. E. Hearst. 1984b. **Reaction centre and LHI genes from *Rhodopseudomonas capsulata*.** Proc. Natl. Acad. Sci. USA 81:189-192.

Youvan, D. C. and S. Ismail. 1985. **Light harvesting II (B800-850 complex) structural genes from *Rhodopseudomonas capsulata*.** Proc. Natl. Acad. Sci. USA 82:58-62.

Youvan, D. C., S. Ismail and E. J. Bylina. 1985. **Chromosomal deletion and plasmid complementation of the photosynthetic reaction centre and light-harvesting genes from *Rhodopseudomonas capsulata*.** Gene 38:19-30.

Zhu, Y. S., and S. Kaplan. 1985. **Effects of light, oxygen and substrates on steady state levels of mRNA coding for ribulose 1,5-bisphosphate carboxylase and light harvesting and reaction centre polypeptides in *Rhodopseudomonas sphaeroides*.** J. Bacteriol. 162: 925-932.

Zhu, Y. S. and J. Hearst. 1986. **Regulation of expression of genes for light-harvesting antenna proteins LHI and LHII; reaction centre polypeptides RC-L, RC-M and RC-H; and enzymes of bacteriochlorophyll and carotenoid biosynthesis in *Rhodobacter capsulatus* by light and oxygen.** Proc. Natl. Acad. Sci. USA 83:7613-7616.

Zhu, Y. S., D. W. Cook, F. Leach, G. A. Armstrong, M. Alberti, and J. E. Hearst. 1986a. **Oxygen-regulated mRNAs for light harvesting and reaction centre complexes and for bacteriochlorophyll and carotenoid biosynthesis in *Rhodobacter capsulatus* during the shift from anaerobic to aerobic growth.** J. Bacteriol. 168: 1180-1188.

Zhu, Y. S. and P. J. Kiley, T.J. Donohue, and S. Kaplan. 1986b. **Origin of the mRNA stoichiometry of the *puf* operon in *Rhodobacter sphaeroides*.** J. Biol. Chem. 261: 10366-10374.

Zhu, Y. S. and J. E. Hearst. 1988. **Transcription of oxygen-regulated photosynthetic genes requires DNA gyrase in *Rhodobacter capsulatus*.** Proc. Natl. Acad. Sci. USA 85: 4209-4213.

Zsebo, K. M. and J. E. Hearst. 1984. **Genetic-physical mapping of a photosynthetic gene cluster from *Rhodopseudomonas capsulata*.** Cell. 37: 937-947.

Zsebo, K. M., F. Wu. and J. E. Hearst. 1984. **Physical mapping and Tn5.7 mutagenisis of pRPS404 containing photosynthesis genes from *Rhodopseudomonas capsulata*.** Plasmid 11: 182-184.

Zuber, H., R. Brunisholz, G. Frank, P. Fuglistaller, W. Sidler and R. Theiler. 1983. In: Proceedings of the workshop on molecular structure and function of the light-harvesting pigment-protein complexes and photosynthetic reaction centres. Zurich, Switzerland. pp56-58.

Zuber, H. 1985. **Structure and function of light-harvesting complexes and their polypeptides: yearly review.** Photochem. Photobiol. 42: 821-844.

Zuber, H. 1986. **Structure of light-harvesting antenna complexes of photosynthetic bacteria., cyanobacteria and red algae.** Trends in Biochemical Sciences. 11: 414-419.

Zucconi, A. and J. T. Beatty. 1988. **Posttranscriptional regulation by light of the steady-state levels of mature B800-850 light-harvesting complexes in *Rhodobacter capsulatus*.** J. Bacteriol. 170: 887-882.