STUDIES ON THE GENOME STRUCTURE OF
NEUROVIRULENT AND ATTENUATED POLIOVIRUSES

by

ALAN JAMES CANN, BSc.

A thesis submitted for the
degree of Doctor of Philosophy
in the University of Leicester
Variable print quality
Statement

This thesis is based on work conducted by myself, or in collaboration where noted, in the Department of Microbiology at the University of Leicester during the period between October 1980 and September 1983.

All the work recorded in this thesis is original unless otherwise acknowledged in the text by references. None of this work has been submitted for another degree in this or any other University.

Signed...A.J.Gam........... Date...31/1/84........
Acknowledgements

I am indebted to Professor P.H.A. Sneath for permission to work in the Department of Microbiology.

I would like to thank my supervisor, Dr. Jeff Almond, for his guidance throughout the work presented here and his encouragement in matters such as attendance of conferences and the publication of results. My thanks also go to Dr. Glyn Stanway for all his help with many experiments and for proof-reading; to Lindsay Clarke, Pam Hughes and Roger Mountford for their excellent technical assistance; to Martin Ryan and particularly to Dr. Rudi Hauptmann for the many discussions we enjoyed. I would also like to thank Roger Key for his help with the computing.

Grateful thanks are also due to our collaborators at NIBSC, especially Drs. Geoffrey Schild, Philip Minor and David Evans.

I would like to acknowledge the receipt of a research studentship from the Medical Research Council which supported this work.
Abstract

The RNA genomes of neurovirulent and attenuated type 3 polioviruses have been cloned in *E. coli* using an efficient RNA-cDNA hybrid technique. The complete nucleotide sequence of the vaccine-associated neurovirulent revertant P3/119 and, in collaboration with others, the attenuated vaccine strain P3/Leon 12 a₁ b, have been determined. These have been compared with that of the neurovirulent parent strain P3/Leon/37. Ten nucleotide sequence differences were observed between the parent P3/Leon/37 and the vaccine P3/Leon 12 a₁ b, three of which resulted in amino acid substitutions. Between the vaccine and the revertant P3/119, seven nucleotide sequence differences were observed. Three of these resulted in amino acid substitutions. The possible significance of individual nucleotide sequence differences to the attenuation of and reversion to neurovirulence in poliovirus type 3 is discussed.

The nucleotide sequence of P3/Leon 12 a₁ b was the first to be determined for a type 3 poliovirus. Comparison of this sequence with published type 1 sequences has demonstrated the extent of the molecular homology between them.
Abbreviations Used

A Adenine
amp Ampicillin
AMV Avian Myeloblastosis Virus
BCIG 5-bromo-4-chloro-3-indolyl-β-D-galactoside
bp Base pairs
BSA Bovine serum albumin
C Cytosine
cDNA Copy DNA
CNS Central nervous system
cpm Counts per minute
DNA Deoxyribonucleic acid
DNase Deoxyribonuclease
dNTP Deoxyribonucleotide triphosphate
DTT D,L-dithiothreitol
E. coli Escherichia coli
EDTA Diaminoethanetetra-acetic acid
EtBr Ethidium bromide
G Guanine
IPTG Isopropyl-β-D-thio-galacto-pyranoside
IPV Inactivated poliomyelitis vaccine
kb Kilobases
kbp Kilobase pairs
L Luria medium
Mr Molecular weight
mRNA Messenger RNA
OPV Oral poliomyelitis vaccine
PEG Polyethylene glycol
RNA Ribonucleic acid
RNase Ribonuclease
RF Replicative form
RI Replicative intermediate
rNTP Ribonucleotide triphosphate
SDS Sodium dodecyl sulphate
SSC Saline sodium citrate
T Thymine
TBE Tris-borate-EDTA
TE Tris-EDTA
tet Tetracycline
Tris Tris(hydroxymethyl)methylamine
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ts</td>
<td>Temperature sensitive</td>
</tr>
<tr>
<td>u</td>
<td>Units</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>vRNA</td>
<td>Virus RNA</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume for volume</td>
</tr>
<tr>
<td>w/w</td>
<td>Weight for weight</td>
</tr>
</tbody>
</table>
CHAPTER 1 - GENERAL INTRODUCTION

1. A Introduction and Objectives
1. B The Problem of Poliomyelitis:
   1. B.1 - Historical Aspects
   1. B.2 - The Epidemiology of Poliomyelitis
   1. B.3 - The Pathology of Poliomyelitis
   1. B.4 - Vaccination Against Poliomyelitis
   1. B.5 - Recent Trends in Vaccine Design
1. C Picornavirus Classification and Structure:
   1. C.1 - Classification of Picornaviruses
   1. C.2 - Morphology of Picornaviruses
   1. C.3 - Structural Proteins of Picornaviruses
   1. C.4 - The Antigenicity of Picornaviruses
   1. C.5 - The Genome Structure of Picornaviruses
1. D Picornavirus Replication:
   1. D.1 - Intracellular Events During Picornavirus Replication
   1. D.2 - Picornavirus Genetics
1. E Attenuation
1. F Molecular Cloning of Picornaviruses

CHAPTER 2 - MOLECULAR CLONING OF THE PICORNAVIRUS GENOME

2. A Introduction
2. B Preparation of Virus Stocks for Molecular Cloning
2. C Virus RNA Extraction and Analysis:
   2. C.1 - RNA Extraction
   2. C.2 - Analysis of Virus RNA
2. D cDNA Synthesis:
   2.D.1 - Reverse Transcription of Poliovirus RNA 57
   2.D.2 - Optimisation of cDNA Synthesis 62
   2.D.3 - Purification of cDNA 66

2. E Homopolymer Addition:
   2.E.1 - The Terminal Deoxynucleotidyl Transferase Reaction 67
   2.E.2 - Annealing of Homopolymer Tailed Molecules 71

2. F Transformation of E. coli:
   2.F.1 - Preparation of Competent Cells 72
   2.F.2 - Selection of Transformed Cells 72
   2.F.3 - Transformation Efficiency of E. coli 73

2. G Conclusions 76

CHAPTER 3 - CHARACTERISATION OF cDNA CLONES

3. A Introduction 77

3. B Recombinant Colony Screening Methods:
   3.B.1 - Phenotypic Selection and Screening 78
   3.B.2 - Colony Hybridisation 79
   3.B.3 - Plasmid Isolations 85

3. C Restriction Endonuclease Mapping:
   3.C.1 - Restriction Enzymes with Hexameric Sites 89
   3.C.2 - Restriction Enzymes with Tetrameric Sites 91
   3.C.3 - Size and Genomic Distribution of cDNA Clones 93
   3.C.4 - Restriction Endonuclease Map of the Total Genomic cDNA 95

3. D Conclusions 97

CHAPTER 4 - NUCLEIC ACID SEQUENCING OF CLONED POLIOVIRUS GENOMES

4. A Introduction 98

4. B M13 Cloning 100
4.C Determination of the Complete Nucleotide Sequence of P3/Leon 12 a, b:

4.C.1 - M13 Sub-cloning
4.C.2 - Nucleotide Sequencing
4.C.3 - Data Handling
4.C.4 - The Complete Nucleotide Sequence of P3/Leon 12 a, b #411

4.D Nucleotide Sequencing of Neurovirulent Type 3 Polioviruses

4.D.2 - Nucleotide Sequencing of P3/119 #643
4.D.3 - The Molecular Basis of Attenuation and Reversion

4.E Conclusions

CHAPTER 5 - GENERAL SUMMARY

CHAPTER 6 - MATERIALS AND METHODS

6.A.1 - Reagents
6.A.2 - Solutions, Buffers and Media

6.B Methods

REFERENCES
List of Tables

Table 1.1 - Comparison of Inactivated and Oral Poliomyelitis Vaccines 13-14
Table 1.2 - Classification of the Picornaviridae 19-20
Table 1.3 - The Derivation of Sabin's Attenuated Vaccine Strains 41
Table 2.1 - Monkey Neurovirulence Test Results 54
Table 2.2 - Standard Conditions for the Reverse Transcription of Poliovirus RNA 57
Table 4.1 - Analysis of the Computer Predicted Polyproteins of P1/Mahoney and P3/Leon 12 a,b 113
Table 4.2 - Nucleotide Sequences of Neurovirulent Revertants 124
Table 6.1 - Composition of Nucleotide Sequencing Reagents $^{32}$P 151
Table 6.2 - Composition of Nucleotide Sequencing Reagents $^{35}$S 153
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Processing of the Poliovirus Polyprotein</td>
<td>33</td>
</tr>
<tr>
<td>1.2</td>
<td>Picornavirus Replication</td>
<td>36</td>
</tr>
<tr>
<td>1.3</td>
<td>Assembly of Poliovirus Particles</td>
<td>37</td>
</tr>
<tr>
<td>2.1</td>
<td>Gel Electrophoresis of Polio vRNA</td>
<td>56</td>
</tr>
<tr>
<td>2.2</td>
<td>Analysis of Reverse Transcription Products</td>
<td>59</td>
</tr>
<tr>
<td>2.3</td>
<td>Time Course of Reverse Transcription Reaction Subjected to Acrylamide Gel Electrophoresis</td>
<td>60</td>
</tr>
<tr>
<td>2.4</td>
<td>Time Course of Reverse Transcription Reaction</td>
<td>61</td>
</tr>
<tr>
<td>2.5</td>
<td>Salt Concentration in Reverse Transcription Reaction</td>
<td>62</td>
</tr>
<tr>
<td>2.6</td>
<td>Graph of cDNA Yield Against Salt Concentration</td>
<td>63</td>
</tr>
<tr>
<td>2.7</td>
<td>Enzyme Concentration in Reverse Transcription Reaction</td>
<td>64</td>
</tr>
<tr>
<td>2.8</td>
<td>Restriction Endonuclease Cleavage Sites and Functional Regions of pBR 322 and pAT 153</td>
<td>67</td>
</tr>
<tr>
<td>2.9</td>
<td>Tailing of PstI Digested pAT 153</td>
<td>68</td>
</tr>
<tr>
<td>3.1</td>
<td>Colony Hybridisation</td>
<td>83</td>
</tr>
<tr>
<td>3.2</td>
<td>Rapid Plasmid Isolates (Method 1)</td>
<td>85</td>
</tr>
<tr>
<td>3.3</td>
<td>Rapid Plasmid Isolates (Method 2)</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Mapping of a Recombinant Plasmid (pSAG 22) by Restriction Endonucleases With Hexameric Recognition Sites</td>
<td>90</td>
</tr>
<tr>
<td>3.5</td>
<td>Restriction Endonuclease Mapping of Recombinant Plasmids</td>
<td>92</td>
</tr>
<tr>
<td>3.6</td>
<td>cDNA Insert Size and Genomic Location of P3/Leon 12 aₐᵇ #411 Clones</td>
<td>93</td>
</tr>
<tr>
<td>4.1</td>
<td>Restriction Endonuclease Map of M13 mp1</td>
<td>100</td>
</tr>
<tr>
<td>4.2</td>
<td>Restriction Endonuclease Fragments Used in the Determination of the Complete Nucleotide Sequence of P3/Leon 12 aₐᵇ</td>
<td>103</td>
</tr>
<tr>
<td>4.3</td>
<td>Computer Predicted Restriction Endonuclease Map of P3/Leon 12 aₐᵇ</td>
<td>109</td>
</tr>
</tbody>
</table>
Figure 4.4 - The Complete Nucleotide Sequence of P3/Leon 12 a, b
(Parts a-e)

Figure 4.5 - Location of Amino Acid Sequence Differences Between
P3/Leon 12 a, b and P1/Mahoney

Figure 4.6 - Distribution of Sequences Obtained from Randomly
Generated M13 Sub-clones of P3/119

Figure 4.7 - Nucleotide and Amino Acid Sequence Differences Between
Neurovirulent and Attenuated Type 3 Polioviruses

Figure 4.8 - Computer Predicted Hydrophilicity Profiles of Amino Acid
Substitutions
CHAPTER 1

GENERAL INTRODUCTION
1. INTRODUCTION AND OBJECTIVES

In the economically developed areas of the world such as Western Europe, Scandinavia and North America, poliomyelitis is no longer a major threat to public health. In many other areas however, for example, Africa, Asia, Central and South America, the disease is still highly prevalent and much feared (see, Bazin, 1980). During the period 1971-1980, there were 10-20 reported cases per million people per annum world-wide, with no clear upward or downward trend. In 1980, 42,000 cases were reported to the World Health Organisation (WHO, 1981). These figures are probably a gross under-estimate (see, Melnick, 1982) since other methods of collecting data, such as rural lameness surveys, often imply that the total number of cases is much higher (Nicholas et al, 1977; WHO, 1977; Sabin, 1981). The pronounced difference in the incidence of poliomyelitis which now exists between the developed countries and the third world is a relatively recent situation. In the first half of this century, Western Europe, North America and Australasia experienced large-scale poliomyelitis epidemics, interspersed with an endemic pattern of sporadic outbreaks of the disease (Paul, 1971). These were eventually curbed and the disease controlled (though not eliminated) through the use of vaccines developed during the 1950's. The control of poliomyelitis in these few countries however, although locally effective, has had only a minor effect on the total global incidence of the disease (WHO, 1981). For this reason, poliomyelitis is one of six diseases (together with diphtheria, whooping cough, tetanus, tuberculosis and measles) currently the subject of a world-wide expanded programme of immunisation (EPI) sponsored by the WHO. It is clear that although the global problem of poliomyelitis is now being addressed, much work is still needed.

The pathological changes which occur during poliomyelitis have been well studied (section 1.8.3), but the mechanisms by which infection with poliovirus initiates the disease remain obscure. It is of great practical
importance to be able to determine whether a particular strain of poliovirus is harmless or dangerously neurovirulent. Although the Sabin vaccines (section 1.F) have been in world-wide use for 25 years and several thousand million doses have been administered, the precise basis of their attenuation has never been established. The unsegmented RNA genome of picornaviruses (section 1.C) has generally proved to be intransigent to the sort of classical genetic analysis involving recombination or re-assortment which can be performed on viruses possessing either DNA or segmented RNA genomes. With this fact in mind, the work described in this thesis was begun in an attempt to harness new recombinant DNA techniques, such as molecular cloning and nucleotide sequencing, to gain an understanding of the molecular basis of attenuation and reversion to neurovirulence in poliovirus type 3. The application of these techniques offers the possibility of defining precisely those features of the virus genome responsible for attenuation. It might then be possible to approach the production of safer, more efficient means for the prophylaxis and even treatment of poliomyelitis.

The aim of the work presented here was to establish the differences in the nucleotide sequences of neurovirulent and attenuated polioviruses. There were several reason for wishing to achieve this. First, to identify the changes responsible for the attenuation of poliovirus and perhaps gain an insight into their mechanism of action. Second, it was hoped that increased knowledge of the basis of attenuation might simplify and improve vaccine testing. The development of a reliable in vitro test is a highly desirable objective and would enable the unpleasant, cumbersome and expensive test used at present (based on neurovirulence of the virus in monkeys) to be abandoned. Third, to investigate the possibility of producing an improved type 3 vaccine, perhaps by genetic manipulation in vitro. Finally, insights into the biology of picornaviruses obtained from the molecular cloning of poliovirus might illustrate those features responsible for neurovirulence and this may be relevant to the study of
other, related viruses.

The work presented in this thesis was undertaken as part of a wider study performed in the laboratories of Dr. J. W. Almond at the University of Leicester and Dr. G. C. Schild of the National Institute for Biological Standards and Control (NIBSC), London. Some of the work described was performed in co-operation with colleagues. Where data presented were obtained from joint experiments, acknowledgements are given in the text. This first chapter provides a general introduction to the problem of poliomyelitis and to the biology of picornaviruses. Chapter 2 describes the molecular cloning of the poliovirus genome by an efficient RNA-cDNA hybrid method. Chapter 3 goes on to describe the characterisation of cDNA clones of the entire P3/Leon 12 a,b genome and the determination of a restriction endonuclease map of the genomic cDNA. The results of nucleotide sequencing experiments on the vaccine strain P3/Leon 12 a,b and the revertant P3/119 are given in chapter 4. The nucleotide sequence differences observed between the genomes of neurovirulent and attenuated strains of poliovirus and the implications for the objectives listed above are discussed.
1.3 THE PROBLEM OF POLIOMYELITIS

1.B.1 - Historical Aspects

Probably the earliest record of poliomyelitis is a clay tablet from the Egyptian 18th dynasty (1580-1350 B.C.) which depicts a figure showing a deformity highly characteristic of residual poliomyelitis paralysis (see, Paul, 1971). In 1546, the writer Fracastorius devoted a whole volume of an academic treatise on disease to foot-and-mouth disease of cattle ("De Contagione et Contagiosis Morbis et Curatione, Libre III"). Thus historically, people were familiar with picornavirus diseases, although ignorant of their true cause. During the subsequent centuries, many unlikely factors or agents were implicated as being the cause of poliomyelitis, but the true cause of the disease was revealed only when the infective agent responsible for poliomyelitis was isolated by Landsteiner and Popper in 1909 (see, Paul, 1971).

During the decades following these observations, knowledge of poliovirus progressed alongside developments in other fields, particularly in the new subject of immunology. It gradually became clear that three distinct serotypes of poliovirus were involved. It was found that these were not effectively cross-neutralised by antibodies and to be complete, immunological protection must be gained against not one but three types of virus. In the 1930's and 1940's, work tended to concentrate on factors such as the spread of poliomyelitis. The isolation of polioviruses from diverse sources diverted much attention into these investigations (see, Paul, 1971). Attention also concentrated on treatment rather than prevention of the disease. Many unconventional treatments were proposed and tested. However, these also included the work of Elizabeth Kenny, whose regime of physiotherapy for those suffering poliomyelitis paralysis helped change the belief that the treatment for paralysis should consist of immobilisation. The development of the "iron lung" enabled...
poliomyelitis victims who would previously have died of suffocation (due to paralysis of their respiratory muscles) to survive (see, Paul, 1971). Because of the limited success achieved in the treatment of poliomyelitis, it became obvious that the most effective course of action would be to prevent the disease rather than search for a cure. While most attention was being paid to the above measures, other developments were happening which were to lead eventually to the development of vaccines. The discovery by Enders et al (1949, 1980) that poliovirus could be cultivated in human embryonic epithelial tissues in vitro was a vital step which opened the way for the development of poliovirus vaccines. This discovery corrected the earlier unfortunate observation that poliovirus could be cultivated only in nervous tissue in vitro (Sabin and Olitski, 1936). This was probably because the strain which had been studied by Sabin and Olitski was a highly neurotropic isolate (see, Paul, 1971). Subsequent work during the 1950's resulted in two competing types of vaccine. The first to be developed was the formalin-inactivated poliomyelitis vaccine (IPV) developed by Dr. Jonas Salk and his colleagues in 1955 (see, Paul, 1971), which was followed in 1957 by the live-attenuated, orally-administered poliomyelitis vaccine (OPV) produced by Dr. Albert Sabin (Sabin and Boulger, 1973). The successful use of these vaccines (though not entirely without problems) in controlling poliomyelitis in developed countries is discussed in section 1.B.4.

1.B.2 - The Epidemiology of Poliomyelitis

In rural communities possessing primitive sanitation facilities and in the absence of any anti-poliomyelitis vaccination, serological surveys reveal that at least 90% of children three or more years of age possess antibodies to at least one serotype of poliovirus (Schonberger et al., 1981; Melnick, 1982). In addition, it is found that there are between 100 and 200 sub-clinical immunising infections for each single (endemic)
paralytic case of poliomyelitis. In such situations, many fatal cases of poliomyelitis are likely to go unnoticed in view of high background infant mortality rates. Since polioviruses circulate freely in the community, it is probable that infants are protected from poliomyelitis by maternal antibodies. Infants thus protected probably experience sub-clinical immunising infections and hence achieve full active immunity to poliomyelitis, which is maintained by frequent challenge with poliovirus throughout their lives. The situation described above existed in many countries up to the start of the nineteenth century. After this time, the process of industrialisation, necessarily accompanied by urbanisation of the majority of the population, upset the balanced situation which had previously existed and the occurrence of poliomyelitis began to change from an endemic to an epidemic pattern. These epidemics increased almost yearly in size and severity up to the introduction of vaccination in the 1950's. Melnick (1982) referred to this pattern of change as "poliomyelitis in transition".

The origin of these epidemics has been much discussed (see, Nathanson and Martin, 1979), but essentially it would appear that as the amount of wild poliovirus circulating in the community falls, the challenge to the adult immune system becomes intermittent rather than continual. Levels of maternal antibodies are likely to fall below those required for full protection from infection. This allows the virus to establish infections in unprotected infants. When coupled with a reduction in infant mortality rates and a decline in the popularity of breast-feeding of babies in urbanised communities for the first half of this century, these facts help to explain the emergence of the epidemic pattern of poliomyelitis observed during this period. It seems that this pattern of events is being repeated (perhaps over a rather shorter time scale than occurred in Western Europe and North America) in many third world countries at the present time (see, Melnick, 1982).

The reduction in the incidence of poliomyelitis following the
introduction of vaccination, initially with IPV and in most countries shortly afterwards with OPV, was dramatic. Melnick (1982) gives the following figures: in 1955, the USSR, 23 other European countries, the USA, Canada, Australia and New Zealand experienced more than 76,000 reported cases of poliomyelitis. In 1967, 1013 cases were reported in these countries, a reduction of 99%. Current rates of incidence in the USA are 0.02-0.002/100,000 population. Most of these cases are either imported or vaccine-associated (section 1.B.4). It is notable that other enteroviruses such as echoviruses and coxsackieviruses (see section 1.C), also capable of causing poliomyelitis-like infections and paralysis (Sabin, 1981; Grist et al, 1978), but not subject to vaccination programmes, show a similar seasonal epidemic pattern of sub-clinical infections in the UK and other countries as was previously seen with polioviruses (Grist et al, 1978).

1.B.3 - The Pathology of Poliomyelitis

Bodian (1949) referred to poliomyelitis as "notoriously variable in character as well as severity". The disease of poliomyelitis may range from asymptomatic infections, detectable only by isolation of the virus from stools or by serological examination, through a few very mild, transient symptoms such as low fever, headaches or nausea to the total flaccid paralysis of many different sets of muscles or to poliovirus encephalitis. The last of these conditions is commonly fatal due to the inflammation of the brain parenchyma. Wasting and paralysis of voluntary muscles tends to affect the legs rather than the arms and is often asymmetric. In 10-15% of patients with paralysis, there is also some bulbar paralysis resulting from lesions in the medulla oblongata and cranial nerves IX and X, which may also result in death due to cardio-respiratory abnormalities (Johnson, 1982). This clinical picture of poliomyelitis has long been established (see, Paul, 1971) and the
underlying events of the pathogenesis of poliomyelitis have been studied for the last 40 years, but even now the whole process is not fully understood.

Histologically, the effect of poliomyelitis on the body was accurately described by Bodian (1949) and has since been well studied, especially in animal experiments (e.g. Beswick and Coid, 1961; Beswick et al, 1964; Araya, 1977). Although no histological changes are detectable in non-nervous tissues from which poliovirus can be isolated, the effects of poliomyelitis on the central nervous system (CNS) are clearly visible. These effects consist of inflammation and neuronal destruction of the anterior horns of the spinal cord, the bulbar motor nuclei, reticular formation, thalamus and the motor cortex (Bodian, 1949). The posterior horns of the spinal cord may become inflamed. It is characteristic that the disease chiefly affects the grey matter of the CNS rather than the white matter. While pain is often experienced at the onset of paralysis, damage is primarily to motor rather than sensory function. Some recovery from partial paralysis is often seen, which results from the rerouting of nerve pathways from the interrupted primary paths to alternative secondary paths, and by relearning patterns of motor activity (Bodian, 1949; Johnson, 1982). Physiotherapy may assist both in this latter process and in counteracting the effects of muscular wasting in paralysed limbs.

The initial alimentary tract multiplication of the virus following infection is usually asymptomatic, but may be accompanied by low grade fever, headaches, sore throat or nausea. Excretion of the virus in stools reaches a peak 5-6 days after infection and has normally ceased after 2-3 weeks, although in some cases the virus continues to be excreted for up to 12 weeks. Virus can also be isolated from lymphatic tissues in the oropharynx and mesentery during the early stages of the infection. Patients infected in this way are subsequently found to have high levels of both serum and secretory antibodies against the virus (Bodian and Paffenbarger, 1954; Horstmann et al, 1954). In more than 99% of cases,
this is the end of the infection and the only lasting consequence is a protective antibody response against further infections by polioviruses of the same serotype. However, in the remaining 0.5-1% of cases, a systematic infection may follow, usually with some degree of neurological involvement (see above). Although the portal of entry of poliovirus into the body has now been established to be the faecal-oral route (rather than the nasal route as previously thought, see, Paul, 1971), the way in which the virus gains entry to the CNS is still not known. The initial multiplication of the virus seems to be associated with lymphatic tissues in the alimentary tract (Peyers patches in the small intestine, mesenteric lymph nodes and adenoidal tissues in the throat). Localised infections in the CNS resulting in lesions are preceded by viraemia (Bodian and Paffenbarger, 1954; Horstmann et al, 1954; Sabin, 1956). In what precise way these events are interrelated is not clear. A brief, 2-3 day remission follows the initial multiplication of the virus and then headaches, stiffness and muscle pain are followed very quickly by the rapid onset of paralysis, a process which is complete within 24 h.

Several factors are known which predispose infected patients to paralysis. Undoubtedly the major one is the degree of neurovirulence of the strain of poliovirus involved (Bodian, 1949; Sabin, 1956), but host factors involved include physical activity at the time of the onset of paralysis (Horstmann, 1950) and trauma such as tonsillectomy, dental surgery or intramuscular injections. Pregnancy and increasing age also tend to predispose patients to paralysis. For reasons which are not understood, adults experiencing primary poliovirus infection are at least ten times more likely to develop paralysis than infants. In virgin population epidemics, almost all the fatal cases are seen in those over 40 years of age (see, Johnson, 1982). Poliovirus infections occurring close to the full-term of pregnancy are more likely to result in severe paralysis with bulbar involvement (Hurley, 1983). This is perhaps due, at least in part, to partial immune suppression which occurs during pregnancy.
in order for the mother to tolerate foetal antigens. Some reports have linked infection (or immunisation) with poliovirus and other picornaviruses with neurological syndromes such as Guillian-Barré syndrome or Reye's syndrome, but the connection between these events has yet to be satisfactorily demonstrated (Johnson, 1982).

Cell-mediated immunity probably plays a more significant role in poliomyelitis than had been previously thought, but has been much less well studied than antibody production. Initial protection against infection does involve the production of secretory antibodies in the saliva and in the gut (see, Johnson, 1982; section 1.8.4). It is possible that serum antibody may also play a role in modifying the course of the infection during the viraemia phase (Sabin, 1956; Melnick, 1982). OPV administered to hypogammaglobulinaemic subjects is 100,000 times more likely to result in paralysis than with normal healthy individuals (Wyatt, 1973). Such people continue to excrete the virus in stools for far longer than normal, in some cases, up to several years. However, once CNS involvement occurs, the T-cell response is dominant in clearing virus from the CNS (Johnson, 1982), but damage progresses so rapidly (i.e. within hours) that it is impossible to halt the disease at this stage. In addition, a misdirected or over vigorous T-cell response may also contribute to the total damage to the CNS (see, Banatvala, 1983).

1.8.4 - Vaccination Against Poliomyelitis

In the 1920's, hopes rested on the maintenance of passive immunity by the administration of convalescent serum to patients in the acute phase of poliomyelitis as a means of treatment (see, Paul, 1977). These trials were an expensive and time consuming failure, due largely to difficulties involving sufficiently early, accurate diagnosis of poliomyelitis and the administration of serum at the critical time. After a lull of a few years, attention turned to the induction of active immunity against poliomyelitis
by vaccination and thus to prevention rather than treatment of the disease, a situation which still exists at the present time.

The first poliomyelitis vaccine was produced by Brodie and Park (1935) and comprised a crude 10% emulsion of infected monkey spinal cord, in which the virus had been inactivated by formalin treatment (see, Paul, 1971). This vaccine was tested on at least 3000 children with catastrophic results, which have never been fully revealed (Paul, 1971). The intervening 20 years between these ill-fated events and the successful production of a formalin-inactivated vaccine by Salk in 1955 were needed to permit developments in the fields of tissue culture, growth of polioviruses in vitro and to test methods of inactivating polioviruses while still maintaining their antigenicity. Such treatments included heat, phenol, UV light and electron bombardment. Ironically, formalin-inactivation was eventually demonstrated to be the best method of IPV production, Paul, 1971). At the same time, work was also begun on investigation of the neurovirulence of well-characterised strains of poliovirus maintained in vitro. In the USA, Sabin in particular was involved in attempts to isolate experimentally produced variant strains of poliovirus which showed reduced neurovirulence when tested in monkeys (Sabin et al, 1954). Soon afterwards (1957), live-attenuated, orally-administered vaccines against all three strains of poliovirus were produced (see, Sabin, 1965; Sabin and Boulger, 1973; section 1.E).

By the late 1950's, there were two effective anti-poliomyelitis vaccines. Although most countries, with a few exceptions (most notably the Netherlands, Finland and Sweden), eventually switched from use of the Salk IPV to the Sabin OPV, there was a dramatic fall in the incidence of poliomyelitis during the 2-3 years in which IPV was used before the introduction of OPV. This trend continued downwards to the present very low levels after OPV was adopted (Paul, 1971; Melnick, 1982). However, perhaps not surprisingly, there was considerable controversy at the end of the 1950's over which type of vaccine was superior. This has still not
been resolved (Salk, 1980; Melnick 1977, 1978, 1982; Sabin, 1981). Table 1.1 gives a brief list of the major advantages and disadvantages of each type of vaccine.

While it is true that IPV has the potential to totally eradicate poliovirus (Salk, 1980), a prospect which could never be matched by OPV (Cossart, 1977), there is no doubt about the efficacy of both types of vaccine. Undoubtedly the major problem faced, especially in third world countries, is how to obtain sufficient vaccine coverage rather than the inefficiency of the vaccines (Bazin, 1980; Sabin, 1981). Assad (1981) summed up this problem by stating that the reasons for low vaccine coverage are mostly managerial and the problem hinges on how to convey the vaccine in good condition to the people who need it most. However, even in developed countries, there may be a problem in achieving sufficiently high voluntary vaccine coverage to maintain "herd immunity" and prevent the reappearance of epidemic poliomyelitis. The current take-up rate for a full course of OPV vaccination in the USA is approximately 65% (Johnson, 1982) and in the UK approximately 80% (Noah, 1983).

The live-attenuated Sabin poliomyelitis vaccine strains (see section 1.E) are amongst the safest of such vaccines, although rare instances of vaccine-associated poliomyelitis do occur (Melnick, 1982). In spite of this, the Sabin vaccine displays some worrying adverse features which conflict with the above criteria. The highly contagious nature of poliovirus and its tendency to spread to household contacts is well known (Bodian and Paffenbarger, 1954) and live vaccine strains show a similar pattern of spread to close (possibly non-immune) contacts.
Table 1.1

INACTIVATED POLIOMYELITIS VACCINE

Advantages
- Confers humoral immunity in vaccinees when properly administered.
- Can easily be incorporated into other vaccination programmes, and administered with other vaccines (diphtheria, pertussis, tetanus).
- No potential for reversion to neurovirulence (inactivated virus).
- Suitable for use in immunodeficient or immunosuppressed individuals and their households.
- Has effectively eliminated polioviruses from the countries where it has been used (Finland, the Netherlands, Sweden) (Salk, 1980).
- May be more effective in tropical areas in overcoming the problems of interference experienced with OPV (Oduntan et al., 1978; Bazin, 1980).

Disadvantages
- Early batches of vaccine were not of sufficient antigenic potential to effectively compete with OPV, but this has now been improved.
- Repeated booster doses are necessary to maintain good protection.
- Does not induce production of secretory IgA in gut and saliva, thus offers less resistance to infection than OPV.
- More costly to produce and administer (by injection rather than oral administration) and booster doses are required.
- Scarcity and expense of monkeys creates problems for production, but some vaccine is now produced in continuous-passage Vero monkey cells.
- Use of antigenically potent but neurovirulent strains of poliovirus has caused tragedy when a vaccine batch is improperly inactivated (Paul 1971) as neurovirulent strains are used as immunogens.
### Table 1.1 (continued)

### ORAL POLIOMYELITIS VACCINE

**Advantages**
- Confers high levels of both humoral and secretory immunity.
- Immunity induced is very long, perhaps life-long (?) (Melnick, 1982).
- Oral administration is quicker, cheaper and more acceptable than injection.
- Effective quickly, blocks spread of wild viruses in epidemics.
- Relatively inexpensive to produce and administer.
- Now prepared in human cell cultures, eliminating the need for so many monkeys or possible contamination with monkey viruses (Sweet and Hilleman, 1960).

**Disadvantages**
- Live virus, capable of reverting to neurovirulence on rare occasions, resulting in vaccine-associated paralysis.
- Vaccine progeny spreads to contacts, e.g. household residents, frequently not immune unless vaccinated at the same time. Tendency to become more neurovirulent with passage (Boulger et al, 1979).
- Not safe for use in immune deficient or immune suppressed individuals (Wyatt, 1973) or contacts of such people.
- Difficulties sometimes experienced in tropical countries with the "cold chain" - the problem of how to get the vaccine to the vaccinees in a viable condition.
- Large numbers of monkeys required for testing of each batch of vaccine in vivo.
Thus, while lowest estimates put the rate of vaccine-associated poliomyelitis at 1 in 11 million doses (see below), the incidence in household contacts of vaccinees is 1 in 4 million doses (and 1 in 30 million doses for community contacts) (Melnick, 1982). In a recent WHO survey of 281 cases of paralytic poliomyelitis which occurred in ten developed countries over a period of ten years (Assad and Cockburn, 1982), 52 cases occurred in vaccine recipients, 70 cases in household contacts and 159 cases in community contacts. Out of 153 cases in this last category, 116 had no history of vaccination, 10 had received one dose of live vaccine, 12 two doses and 6 three or more doses at some time in the past. In countries using the Sabin vaccine, strains of poliovirus circulating in the community are designated as vaccine-like by biochemical and serological tests. Such strains have displaced the previously occurring wild-type strains since the introduction of the live vaccine (Cossart, 1977). While this could be regarded as a possible benefit of the live vaccine, the fact that such apparently vaccine-like strains can cause paralytic poliomyelitis must be regarded as a hazard (Kew et al., 1980; Minor, 1980, 1982). Thus, although the Sabin vaccine tends to give rise to better protection against infection than inactivated vaccines (see above), this must be viewed in the light of the consequences of its use on the community as a whole. In countries where inactivated poliomyelitis vaccine is used exclusively, polioviruses do not circulate in the community and thus there is no opportunity for chance infections. Inactivated poliomyelitis vaccines thus offer the possibility of the total eradication of polioviruses (which has been achieved for smallpox), a feat which could never be performed using the present live-attenuated vaccines.

An even more vexing question than dissemination of the vaccine concerns the stability of the live-attenuated strains currently in use. Melnick's figure of 1 paralytic case per 11 million administered doses is not universally accepted. This is because this estimate takes no account of the number of doses administered to susceptible individuals and the
number of subsequent booster doses given, plus the number of doses discarded (OPV is distributed in the UK in 10 dose vials with instructions to discard any vaccine not used at the end of each clinic) thus the estimate of 1 in 11 million doses is too low. In the WHO study (Assad and Cockburn, 1982) 46/52 (89%) of cases of vaccine-associated paralysis occurred after administration of the first dose of the vaccine, 5/52 (9%) after the second dose and 1/52 (2%) after the third dose. Thus the true risk of vaccine-associated disease is higher than that given by Melnick and in reality is probably closer to 1 case per 0.5-1 million doses administered (CDR, 1982). Since each case has the potential to become the focus of a localised outbreak of poliomyelitis, this low level of vaccine-associated disease must be a cause for concern.

There is a particular problem with the stability of the type 3 Sabin vaccine strain. Although the total number of cases seen has fallen since the start of vaccination (see Melnick, 1982), the proportion of type 3 strains associated with paralytic poliomyelitis has increased (from 4.5% to 52% in one country) while the proportion of cases due to type 1 has fallen (WHO, 1969). In countries where poliomyelitis has been well controlled by the Sabin vaccine, paralysis due to type 1 poliovirus is now very rare indeed. Most if not all of these cases are due to imported wild-type viruses. However, there is a constant low level of vaccine-associated paralysis due to type 3 strains and this shows no sign of decreasing (Melnick, 1982). Of the 52 vaccine-associated cases in the WHO survey, in 21/52 (40%) cases, more than 1 serotype of virus was isolated, 2/52 (4%) were associated with type 1, 8/52 (15%) associated with type 2 and 17/52 (33%) associated with type 3 (no virus was isolated in 4/52 cases). However, both of the type 1 isolates were wild-type strains by biochemical and serological criteria and therefore probably represent chance infections, but all of the type 2 and 3 strains were vaccine-like. Possible reasons for the discrepancies in the behaviour of the different Sabin vaccine strains are discussed in section 1.E.
1.B.5 - Recent Trends in Vaccine Design

The development of chemotherapy directed against picornavirus infections has been disappointing (Butterworth et al., 1976; Herrmann, 1982). Recently, advances have been made in the production of synthetic antigens which elicit neutralising antibodies and thus afford protection against disease. The most promising findings in this field have been made with foot-and-mouth disease. Production of the antigenic VP1 polypeptide (see section 1.C) in E. coli by the use of genetic manipulation techniques (Kupper et al., 1981; Kleid et al., 1981) and the production of synthetic antigenic peptides in vitro (Bittle et al., 1982; Pfaff et al., 1982) offer hope for the development of new vaccines against poliomyelitis and other diseases. The poliovirus type 1 antigenic region has recently been produced in E. coli as a fusion protein, but has not yet been shown to be capable of protecting animals against challenge with poliovirus (van der Werf et al., 1983; Wychowski et al., 1983). Also in the future lies the possibility of other developments such as the ability to immunise against a pathogenic virus by genetic manipulation of a harmless "carrier" virus so that it contains and expresses the antigenic determinants of the pathogen. This has already been achieved by the construction of viable recombinants of vaccinia virus which express the hepatitis B virus surface antigen (Smith et al., 1983). However, such developments rely on a more complete understanding of the mechanisms of pathogenicity and antigenicity of the virus. It was in attempt to obtain such information that the work described in this thesis was undertaken.
1.C PICORNAVIRUS CLASSIFICATION AND STRUCTURE

1.C.1 - Classification of Picornaviruses

Probably the most useful general system for the classification of viruses is that proposed by Baltimore (1971). Under this scheme, the viruses are divided into six groups on the basis of the composition of their genomes and their replication strategy:

- **Class I** - Double-stranded DNA genome.
- **Class II** - Single-stranded positive sense (i.e., same sense as mRNA) DNA genome.
- **Class III** - Double-stranded RNA genome (all known examples fragmented).
- **Class IV** - Single-stranded positive sense RNA genome.
- **Class V** - Single-stranded negative sense RNA genome.
- **Class VI** - Single-stranded RNA genome, replicating via a DNA intermediate.

To this scheme could now be added a seventh group to include hepatitis B and related viruses, possessing a partially double-stranded DNA genome which is thought to replicate (using reverse transcriptase) via a single-stranded RNA intermediate (see, Marion and Robinson, 1983).

The picornaviruses (the name being derived from the prefix "pico", meaning small and the possession of an RNA genome) fall into class IV of this system, since their single-stranded positive sense RNA genome is replicated via a double-stranded RNA intermediate (see, Matthews, 1982). Table 1.2 shows the currently accepted sub-division of the picornavirus family into genera (Matthews, 1982).
TABLE 1.2 CLASSIFICATION OF THE PICORNAVIRIDAE

FAMILY - PICORNAVIRIDAE

GENUS - Enterovirus

Main Characters - Stable at acid pH, buoyant density in CsCl = 1.33-1.34 g/ml, primarily viruses of the gastro-intestinal tract, but also multiply in other tissues, e.g. nerve, muscle, etc.

Members - Human polioviruses 1-3

Human coxsackieviruses A1-22, 24 (A23 = echovirus 9), B1-6
(Dalldorf, 1949)

Swine vesicular disease virus (= coxsackievirus B5 ?)

Human echoviruses 1-9, 11-27, 29-34
(Committee on the ECHO viruses, 1955)

Human enteroviruses 68-71, 72 (hepatitis A virus)
(Rosen et al, 1970)

Murine polioviruses

Bovine enteroviruses 1-7

Simian enteroviruses 1-18

Porcine enteroviruses 1-8

GENUS - Cardiovirus

Main Characters - Unstable at pH 5-6 in presence of 0.1 M halide, buoyant density in CsCl = 1.33-1.34 g/ml, unique serotype and clinical manifestations, poly(C) tract of variable length (80-250 nucleotides) about 150 nucleotides from 5' terminus of genome.

Members - Encephalomyocarditis virus (EMCV)

Mengovirus

Murine encephalomyelitis virus (MEV)
TABLE 1.2 (continued)

GENUS - Rhinovirus

Main Characters - Unstable below pH 5-6, buoyant density in CsCl = 1.38-1.42 g/ml, unique clinical manifestations.

Members - Human rhinoviruses 1A, 2-115

(MacNaughton, 1982)

Bovine rhinoviruses 1 & 2

GENUS - Aphthovirus

Main Characters - Unstable below pH 5-6, buoyant density in CsCl = 1.43-1.45 g/ml, poly(C) tract of variable length (100-200 nucleotides) about 400 nucleotides from 5' terminus of genome, unique clinical manifestations.

Members - Foot-and-mouth disease virus (FMDV) O, A, C, SAT 1-3, Asia 1

Other possible picornaviruses:-

Cricket paralysis virus

Drosophila A, C, P viruses

Equine rhinoviruses 1 & 2

Gonometa virus

Bee X virus

Bee slow paralysis virus

Bee acute paralysis virus

Sacbrood viruses
1.C.2 - Morphology of Picornaviruses

Picornaviruses are small, icosahedral, non-enveloped, RNA-containing viruses. The dry diameter of the virus particle (which normally contains approximately 25% (w/w) water) is 27-28 nm. Purified poliovirus was first crystallised by Schaffer and Schwerdt (1955) and the architecture of the virus particle has been examined by X-ray diffraction (Finch and Klug, 1959). Their data suggested that the virus particle might be made up of 60 identical, asymmetric subunits, each 6.0-6.5 nm in diameter, forming a single shell one protein molecule thick. Using values of Mr $6.7 \times 10^6$ for the weight of the particle and Mr $2 \times 10^6$ for the RNA genome, a weight was calculated for the capsomeres of Mr $8 \times 10^4$. In light of other observations, this figure was subsequently raised to Mr $8.6 \times 10^4$ (see section 1.C.3). Models for the virus capsid were proposed which suggested a shell of 32 (Mayor, 1964), 42 (Agrawl, 1966) or 60 capsomeres (Horne and Nagington, 1959). This discrepancy was only resolved after later discoveries about the protein composition of the virus particle. It is now known that the particle contains 60 copies of each of 4 non-identical polypeptides, plus trace amounts of a fifth, larger precursor polypeptide, VPO (see, Rueckert, 1976; Scraba, 1979; section 1.C.3).

The virus capsid is a densely packed structure impermeable to electron-dense stains (see, Scraba, 1979). The particle is composed of approximately 70% protein and 30% RNA, with no detectable lipid or carbohydrate present (see, Rueckert, 1976; Scraba, 1979). The RNA genome is a continuous single-stranded positive sense molecule of approximately Mr $2.6 \times 10^6$, tightly packed into the virus capsid (see, Scraba, 1979; section 1.C.5). The whole particle has a sedimentation coefficient of approximately 150 S (Putnak and Phillips, 1981). Enteroviruses and cardioviruses have a buoyant density in CsCl of 1.34 g/ml and aphthoviruses a density of 1.43 g/ml, while rhinoviruses have an intermediate buoyant density of 1.4 g/ml. These differences can be
explained by a difference in the penetration of caesium ions into the virus particle (see, Scraba, 1979). The four genera also have different pH stabilities which correlate with their respective buoyant densities and ionic permeabilities.

1.C.3 - Structural Proteins of Picornaviruses

In the enterovirus and aphthovirus genera, the 4 structural proteins are referred to as (in order of decreasing size) VP1-4. In the cardioviruses and rhinoviruses, the equivalent polypeptides are referred to as α, β, γ, δ respectively. Although the exact size of these proteins shows slight variation from one genus to another, their approximate sizes range from VP1(α) Mr 35.5 ± 1.5 x 10^3, VP2(β) Mr 30.0 ± 1.0 x 10^3, VP3(γ) Mr 25.0 ± 2.0 x 10^3 and VP4(δ) Mr 8.0 ± 2.0 x 10^3 (Putnak and Phillips, 1981). Four distinct polypeptides were observed when whole radiolabelled poliovirus particles were dissociated with SDS and a reducing agent and subjected to electrophoresis on SDS-polyacrylamide gels (Maizel, 1963; Rueckert, 1965). The discovery of trace amounts of a larger, precursor polypeptide (VPO) in such experiments was explained by Jacobson and Baltimore (1968b) as being an immature precursor which was not fully cleaved during the maturation of a proportion of the virus population (see, Rueckert, 1976). The question of the construction of the virus capsid referred to in section 1.C.2 was eventually resolved by the thermal dissociation of MEV particles into 13-14 S subunits, which were themselves further dissociated into homogeneous 5 S subunits (Dunker and Rueckert, 1971). These findings showed that 3 of the 4 capsid polypeptides (VP1-3/ α, β, γ) are organised into a single type of protomer in the capsid. This observation rules out the possibility of 32 or 42 protomer capsids (see section 1.C.2). Thus the picornaviruses must have a 60 subunit capsid, each 5 S protomer comprising a single copy of 3 different polypeptides (see, Rueckert, 1976). Similar findings on the structure of
the capsid have since been made for poliovirus (Katagiri et al, 1971), rhinoviruses (Medappa et al, 1971) and aphthoviruses (Talbot and Brown, 1972).

Attempts to determine precise locations for the various proteins comprising the capsid have been made by surface-labelling of intact virus particles, by treatment of particles with cross-linking reagents and by reaction with specific antibodies raised against isolated capsid polypeptides. Surface-labelling studies showed that iodination of exposed tyrosine residues or labelling of free amino groups generally labelled only the VP1 polypeptide, but denaturation may occur during most surface labelling procedures and make results difficult to interpret (see, Putnak and Phillips, 1981). Cross-linking studies using chemical reagents or UV illumination in which the protein components of the capsid are covalently cross-linked to one another or to the RNA show that VP1 and VP3 of poliovirus are external and in close contact and that VP1, 2 and 4 but not VP3 could be cross-linked to the RNA (Wetz and Habermehl, 1979, 1982).

Heating of picornavirus particles at 56°C results in the production of "A" particles, which do not possess VP4, have a decreased sedimentation coefficient (140 S instead of 150 S), are unable to attach to susceptible cells and also have altered antigenic properties. (Antigenic changes are also induced in picornaviruses by heating, chemical treatments and by UV illumination, see section 1.c.4). Surface-labelling of "A" particles primarily labels VP2/3 rather than VP1 (VP4 being absent), which implies that "A" particles have undergone a profound conformational change from the state of the native virus particle (see, Scraba, 1979; Putnak and Phillips, 1981; MacNaughton, 1982). Particles with similar properties are produced when picornaviruses allowed to adsorb to susceptible cells in vitro and eluted before sufficient time has elapsed for them to be internalised. The aphthoviruses differ from the other picornaviruses in that limited proteolysis with trypsin cleaves the VP1 polypeptide into two fragments, abolishing infectivity and native antigenicity but leaving the
The sedimentation coefficient of the particle and the structure of the RNA remain unaltered (Wild and Brown, 1967; Brown, 1979).

1.C.4 - The Antigenicity of Picornaviruses

Preparations of picornaviruses contain two distinct antigenic populations. Intact, mature picornavirus particles show D-antigenicity (D = "dense" i.e. 150 S, also called N = "native"). Anti-D sera neutralise the infectivity of the virus. Virus preparations which have been heated, treated with certain chemicals or exposed to UV light show altered C-antigenicity (C = "coreless" i.e. 140 S, also called H = "heated"). Anti-C antisera are generally not neutralising (see, Rueckert, 1976; Putnak and Phillips, 1981; MacNaughton, 1982). ("A" particles and empty capsids as well as denatured virus particles show C-antigenicity.) This antigenic change is related to physical changes in the conformation of the virus capsid (section 1.C.3). Unfortunately it has not generally been possible to associate antigenic determinants with specific capsid polypeptides, except in the case of the aphtho and cardioviruses, where isolated VP1 apparently does give rise to neutralising antibodies, (Bachrach et al, 1975; Lund et al, 1977; Kleid et al, 1981), although at much lower titres than whole virus particles (see, Bittle et al, 1982).

Four independent antigenic determinants have now been identified within the FMDV VP1 polypeptide, but only one of these appears to be present on the surface of intact virions (Haresnape and McCahon, 1983). The neutralising antibody binding site of FMDV is split into 2 regions, one approximately 140-160 amino acid residues from the amino terminus of VP1 and the other approximately 200-213 residues (the carboxy terminus) (Strohmaier et al, 1982; Bittle et al, 1982). The 140-160 region seems to be dominant (Bittle et al, 1982; Haresnape et al, 1983; Rowlands et al, 1983). Other regions of the VP1 polypeptide are more strongly conserved between serotypes (Beck et al, 1983; Cheung et al, 1983).
An analogous situation is thought to exist in poliovirus type 3 (Minor et al., 1983; Evans et al., 1983; Wychowski et al., 1983; Dr. P. D. Minor, personal communication). In contrast to FMDV, antibodies directed against purified capsid proteins of poliovirus do not neutralise infectivity (Meloen et al., 1979; Dernick, 1981), although a recent report showed that a small amount of neutralising activity may be produced on inoculation with isolated VP1 (Blondel et al., 1983). These apparently conflicting observations could be resolved by the effect on the quaternary structure of the VP1 molecule of other components of the tightly-knit picornavirus capsid (section 1.C.3), especially if the poliovirus VP1 is less "rigid" in isolation from the capsid than aphtho or cardiovirus VP1's (thus the antigenic configuration is lost during purification on SDS-polyacrylamide gels containing urea). It is believed that "sequence determinants alone do not mediate virus neutralisation, which may depend on antigenic determinants specified by complex conformational arrangements of the capsid proteins" (Thorpe et al., 1983). Emini et al. (1983a) claimed to have identified a neutralisation epitope located on the VP3 polypeptide of poliovirus type 1, but surprisingly, this epitope was identified by serum raised against purified VP4. The authors also reported that they were unable to repeat this observation. Wiegers and Dernick (1983) claimed that monospecific VP3 antisera could neutralise infectivity of poliovirus type 1. However, again it is difficult to interpret the significance of these results, which were based on a very sensitive plaque-reduction assay. In Coxsackie virus B3, there is some evidence that both purified VP1/3 and VP2 can induce neutralising antibodies (Beatrice et al., 1980).

Antibody neutralisation follows first order kinetics, implying that a single type of antibody-virus interaction is responsible for neutralisation (Dulbecco et al., 1956). The mechanism of antibody neutralisation has not yet been established. It is known that neutralised virus may still bind to susceptible cells (Holland and Hoyer, 1962; Dr. P. D. Minor, personal communication) and thus virus neutralisation may not
always be due to a simple blocking of attachment (although antibodies which block attachment will of course be neutralising). It is possible that one mechanism of neutralisation results from a conformational change in the virus capsid, possibly analogous to that of the D-C transition (Emini et al., 1983c). It has recently been suggested that a minimum of 4 antibodies must bind to a single virus particle in order for it to be neutralised (Icenogle et al., 1983). Antigenic regions of the VP1 polypeptide may also be defined by the use of synthetic peptides which elicit neutralising antibodies (Bittle et al., 1982; Emini et al., 1983b; Rowlands et al., 1983; Dr. G. C. Schild, personal communication).

1. C. 5 - The Genome Structure of Picornaviruses

The picornavirus genome is a continuous, single-stranded, positive sense RNA molecule of Mr $2.6 \pm 0.2 \times 10^6$ (Fellner, 1979; Putnak and Phillips, 1981). Although picornaviruses have often been used as a convenient model to study the translation of eukaryotic mRNA's, the virus genome has three unique features which distinguish it from such molecules:

1) The 5'-terminus - Unlike all eukaryotic mRNA's so far examined, picornavirus genomes do not possess the 5' m7G(5')ppp(5')Np cap structure (Nomoto et al., 1976), but instead have a small (approximately 22 amino acids), basic, virus encoded protein (VPg) covalently linked to the 5' uridylic acid terminus of the genome (Nomoto et al., 1977a,b; Pettersson et al., 1977; Flanagan et al., 1977; Sangar et al., 1977; Gauntt, 1980). RNA isolated from the polysomes of infected cells does not contain VPg (Nomoto et al., 1977a), but the 5' uridylic acid terminus of the negative strand formed during the replication of the genome does (Pettersson et al., 1977; see section 1.D.1). Infected and uninfected cells contain an enzymatic activity which cleaves VPg from the RNA (Ambros et al., 1978, 1980; Sangar
et al, 1981). vRNA from which VPg has been removed by protease treatment is still infectious (Nomoto et al, 1977b; Sangar et al, 1977). Thus VPg is not essential for the translation or infectivity of vRNA. The roles which have been suggested for VPg include:

- To selectively inhibit the translation of cellular mRNA's and promote the translation of virus RNA (see section 1.D.1).

- As a primer for virus RNA replication. VPg is found on both positive and negative sense strands of RNA in the infected cell, but is cleaved off from (at least a proportion of) polysomal RNA prior to translation. Antibodies against VPg strongly inhibit the function of the poliovirus replicase complex in vitro (Baron and Baltimore, 1982b; see section 1.D.1).

- Involved in regulation of RNA replication other than as a primer.

- As a regulatory protein which determines which strands of RNA are to be packaged into virus capsids. Although VPg is found attached to negative sense strands of replicative-intermediates in infected cells, these are present early in the course of the infection, before many capsid components have been synthesised.

- To protect the 5' terminus of the genome from exonuclease digestion.

It has recently been discovered that aphthoviruses possess three similar but distinct VPg's, all of which appear to be used in vivo (King et al, 1980b; Forss and Schaller, 1982). The significance of this finding is not at present understood. Nucleotide sequencing of other picornaviruses such as poliovirus reveals that this is not true for all members of the family.

ii) Internal sequences - The genomes of cardioviruses and aphthoviruses contain internal poly(C) tracts which are not found in enteroviruses or rhinoviruses (Porter et al, 1974; Brown et al, 1974). In EMCV, this region is approximately 100 nucleotides long (Porter et al, 1974) and is situated approximately 150 nucleotides from the 5' end of the genome, while in the
aphthoviruses it is of variable length from approximately 100-200 nucleotides and located approximately 400 nucleotides from the 5' end of the genome (Harris and Brown, 1977; Rowlands et al, 1978; Black et al, 1979). The function of this region is not known, but it is apparently not translated \textit{in vivo}. If these poly(C) tracts were of great importance to a fundamental biological function of the virus genome such as replication, it would be logical to expect them to occur in all the picornavirus genera, which they do not. There is some speculation that these regions might be involved in the encapsidation of the genome. The 5' terminal 742 nucleotides of poliovirus are thought not to be translated (Dorner et al, 1982), although the synthesis of small peptides from this region has not been ruled out (Racaniello and Baltimore, 1981a; Kitamura et al, 1981). It is known that there is a stable hairpin structure within the 5'-terminal 85 nucleotides of poliovirus (Larsen et al, 1981; chapter 4) and the conservation between serotypes has implications both for the functional significance of this structure and for the translation of poliovirus RNA (see section 1.D.1). In aphthoviruses, the major open reading frame spanning most of the genome is located roughly the same distance from the 5' end of the genome as in poliovirus, on the 3' side of the poly(C) tract (Sangar et al, 1980). The "S" fragment of the FMDV genome corresponding to the region on the 5' side of the poly(C) tract is not translated \textit{in vitro} (Rowlands et al, 1978).

\textbf{iii) The 3' terminus} - In common with some other virus and most eukaryotic messenger RNA's (Kates et al, 1970), picornavirus genomes contain a poly(A) tract of approximately 50-100 nucleotides at their 3' ends (Yogo and Wimmer, 1972). The precise length of this sequence varies from one genus to another (Ahlquist and Kaesberg, 1979; see, Fellner, 1979). The reasons which have been suggested for the presence of poly(A) at the 3' terminus of eukaryotic mRNA's include involvement in the processing of mRNA precursors, transportation from the nucleus to the cytoplasm,
promotion of efficient translation and protection of the 3' terminus from exonuclease digestion. Although the two former points do not apply to picornaviruses, whose replication occurs in the cytoplasm of infected cells, the second two points may be relevant. The presence of the poly(A) tract is necessary for the infectivity of poliovirus (Spector and Baltimore, 1974). Unlike eukaryotic mRNA's, picornavirus genomes (with the possible exception of cardioviruses) contain no 5'-AAUAAA-3' polyadenylation signal at the 3' terminus. Poly(A) is added to positive sense strand by transcription from poly(U) sequences found at the 5' terminus of negative sense strands during replication (Yogo et al., 1974; Spector and Baltimore, 1974; see, Fellner, 1979). In this respect again, picornavirus genomes are not typical of eukaryotic mRNA's. The 3' terminal region of the genome is not translated and contains termination codons in all the possible reading frames. In poliovirus, this region is 72 nucleotides long (Racaniello and Baltimore, 1981a; Stanway et al., 1983a) and is totally conserved in types 1 and 3, implying a definite biological function for this region, which is not conserved between different genera (chapter 4).
1. D PICORNAVIRUS REPLICATION

1. D. 1 - Intracellular Events During Picornavirus Replication

**Virus entry** - The precise mechanism by which picornaviruses gain entry into susceptible cells is unknown. Prior to entry, the virus attaches to a receptor on the cell surface, the exact nature of which has yet to be determined, but these receptors are thought to be specific for different groups of picornaviruses (Miller *et al.*, 1974; Amchenkova *et al.*, 1982). Monoclonal antibodies raised against surface antigens of susceptible cells block attachment/infectivity of one particular genus of picornavirus only, e.g. Coxsackie B viruses (Campbell and Cords, 1983) or poliovirus types 1, 2 and 3 (Dr. P. Minor, personal communication). Penetration of the host cell is assumed to occur by means of pinocytosis or viropexis (see, Dales, 1973). The virus does not actually enter the cell until penetration of the vesicle is complete. Uncoating of the virus particle then occurs. Very few actual details of this process are known for picornaviruses, although some of the molecular changes to the capsid which occur during this process have already been discussed (section 1. C. 3). Membrane factors modify bound virus particles *in vitro*, resulting in the loss of VP4 and also protect the virus from heat-induced degradation (De Sena *et al.*, 1983).

**Shut-off** - The timing of events during virus replication varies depending on a number of host and virus-specific factors and has been best studied for viruses with a high infectivity in tissue culture, such as poliovirus, rather than viruses with low infectivity such as the rhinoviruses (Butterworth *et al.*, 1976). In general, there is a rapid cessation of host cell macromolecular synthesis on infection, an "eclipse" (or "early") period of 2-4 h during which no new virus is detectable and then a 2-4 h exponential period of virus production and release (see, Baltimore, 1969). Host cell shut-off is known to result from a virus-induced block of
translation at the level of initiation (rather than by the destruction of host mRNA). In the case of poliovirus, this is believed to result from the inactivation of a cellular cap binding protein (CBP) which is required for the initiation of translation of capped mRNA's (see, Ehrenfeld, 1982), although this theory has not yet been conclusively proved. Host cell initiation factors 4A and 4B are not modified by poliovirus infection as was previously thought (Duncan et al, 1983). Earlier reports suggested that shut-off was due to an influx of monovalent cations (chiefly Na\(^+\)) which had the effect of suppressing host cell but not virus-directed translation (see, Jackson, 1979). However, while changes are known to occur in the membranes of virus infected cells which could account for these observations, a cause and effect relationship between this ionic flux and host cell shut off has not been demonstrated. Poliovirus infected cell extracts will not translate (artificially) uncapped mRNA's, thus some other property of the virus genome (e.g. the 5' linked protein or secondary structure at the 5' end of the RNA) may also play a role in translation of virus RNA. Although picornavirus polysomal RNA does not contain VPg (see section 1.C.5), VPg may be involved in the formation of initiation complexes, as its removal is believed to occur after translation has been initiated (Dorner et al, 1981). However, the mechanism of host cell shut-off is distinct for different groups of picornaviruses, e.g. poliovirus and EMCV (Jen et al, 1980; Detjen et al, 1981; see, Ehrenfeld, 1982). Even the host cell type may play a role, e.g. mengovirus in L-cells or HeLa cells (Otto and Lucas-Lenard, 1980). In short, a full list of factors responsible for picornavirus host cell shut-off has not yet been determined.

**Virus directed translation** - Some 2-4 h after infection, the rate of RNA and protein synthesis begins to rise and virus directed translation may reach as much as 40% or more of the initial host cell rate (Baltimore, 1969). The poliovirus genome is believed to contain only a single site for
the initiation of translation, 742 nucleotides from the 5' end of the genome (see, Kitamura et al, 1981; Dorner et al, 1982). This contravenes the favoured model for the selection of initiation sites by eukaryotic ribosomes, in which the 40 S subunit of the ribosome is believed to bind at the 5' terminus of the RNA (possibly due to interactions with the methylated cap) and scan along the message until the first AUG codon is reached, when the 60 S subunit binds and translation is initiated (Kozak, 1978). However, many viruses do not conform to this model (e.g. Rous sarcoma virus, turnip yellow mosaic virus, SV40 late mRNA and adenovirus late mRNA). In poliovirus type 1, there are 8 other methionine codons between the 5' terminus of the genome and the initiation codon (Racaniello and Baltimore, 1981a), while in type 3, there are 7 (Stanway et al, 1983a). It is possible that secondary structure (Larsen et al, 1981) or VPg (Dorner et al, 1981) play a part in the selection of the initiation codon (see, Jackson, 1979). Kozak (1981) has suggested that the nucleotide sequence flanking methionine codons may also be important in selection of the initiation site. Some reports have suggested that there may be 2 or more initiation sites used (see, Ehrenfeld, 1979).

Translation of the RNA results in the synthesis of a polyprotein of Mr 220,000 (only seen in the presence of amino acid analogues which inhibit proteolytic cleavages). This represents 90% of the coding capacity of the whole genome (Sangar et al, 1980; Putnak and Phillips, 1981; Dorner et al, 1982; Stanway et al, 1983a). This polyprotein is subsequently cleaved into virus coded structural and non-structural polypeptides (see, Rueckert, 1976). Jacobson and Baltimore (1968a,b) showed that poliovirus RNA lacks internal initiation/termination codons and is transcribed to produce a polyprotein which is subsequently processed by post-transcriptional cleavages. Although some uncertain genetic evidence existed for the separation of virus capsid and non-structural polypeptides into the 5' and 3' portions of the genome respectively (see section 1.D.2), individual polypeptides were first accurately mapped onto the
picornavirus genome by Summers and Maizel (1971). This was achieved by using the antibiotic pactamycin to inhibit the initiation of translation. By pulse-labelling with radiolabelled amino acids, the relative order of the peptides on the genome was deduced. This work has recently been confirmed by nucleotide sequencing studies (Kitamura et al., 1981; Racaniello and Baltimore, 1981a; Kupper et al., 1981; Kurz et al., 1981).

Post-translational cleavage of the polyprotein is very rapid in the absence of inhibitors such as fluorophenylalanine (FPA) (which strongly inhibits polypeptide cleavages but not synthesis) and polypeptides are cleaved in statu nascendi (see, Baltimore, 1969; Rueckert, 1976). Kinetic evidence implies that some of these cleavages are autocatalytic, while others are dependent on the action of a separate virus coded protease (Rueckert et al., 1980; Palmenberg and Rueckert, 1982; Hanecak et al., 1982). Recent nucleotide sequencing studies have elucidated the location of the virus coded protease and the specific cleavage sites in the polyprotein of poliovirus (Kitamura et al., 1981; Hanecak et al., 1982), although the situation is less clear in other groups of picornaviruses. However, some of the cleavages which occur may be due to host cell proteases. The pattern of cleavages thought to occur during processing of the poliovirus polyprotein are shown in figure 1.1 (Kitamura et al., 1981; Hanecak et al., 1982; Semler et al., 1983).

Virus directed transcription - Picornavirus directed RNA synthesis occurs in the cytoplasm of the infected cell and is not influenced by actinomycin D or alpha-amanitin (which inhibit DNA-dependent RNA polymerases). Replication occurs in enucleate cells (Follett et al., 1975; Detjen et al., 1978), indicating that there is no obligatory nuclear involvement during picornavirus infection (see, Baltimore, 1969). Early during the infection, the rate of RNA synthesis increases exponentially, but at later times levels off. This change occurs at the end of the eclipse phase when the production of new virus particles begins, but its exact
Figure 1.1
Processing of the poliovirus polyprotein

NC VP OD

P1-1a  P2-3b  P3-1b

VP0  VP3  VP1

VP2  VP4

P2-5b  P3-9  P3-2

P2-X  P3-10  P3.7c  P3-4b

P3-6a  P3-6b

Q-G

Y-G

O N-S
timing depends on the multiplicity of infection. The final rate of virus-directed transcription may exceed that of the host cell prior to infection (Baltimore, 1969). RNA synthesis is blocked by both puromycin and cycloheximide (which inhibit elongation of translation and in the case of puromycin, cause the breakdown of polysomes), indicating that the synthesis of virus-specified proteins (presumably the virus polymerase) is necessary before RNA synthesis can occur. Thus there is a tight coupling of transcription and translation in the infected cell (Baltimore, 1969). High concentrations of the amino acid analogue FPA (see above) also inhibit transcription. Guanidine hydrochloride (GHC1) specifically inhibits transcription of vRNA but not host cell macromolecular synthesis. The point of action of GHC1 is not known, but resistant mutants occur and such mutations appear to map approximately in the region encoding the polymerase protein (see, Cooper, 1969; section 1.D.2).

Three types of virus-related RNA molecules can be isolated from poliovirus-infected cells. These are, firstly, single-stranded, positive-sense RNA, which may be either vRNA or mRNA, depending on the presence or absence of VPg respectively (Nomoto et al, 1977a). Secondly, a small but reproducible amount of double-stranded RNA appears to be present in infected cells. This RNA consists of stable, hydrogen bonded positive and negative sense strands and has been called "replicative form" (RF) (see Baltimore, 1969). Like vRNA, this double-stranded RNA is infectious, but is now generally believed to have no direct part in virus replication (Perez-Bercoff, 1979). The third type of RNA present is a high molecular weight, rapidly sedimenting form known as the replicative intermediate (RI). Unlike RF RNA, the RI is partially rather than completely RNase resistant. The nature of the RI has been elucidated from the fact that short pulses of $^3$H-uridine designed to label only nascent strands of RNA are incorporated into RI and also from the degree of RNase resistance of the structure. Estimates for the number of nascent, positive sense RNA strands present on each RI complex give results of about 5.5-6.5 nascent
strands per complex (Baltimore, 1969; Nomoto et al, 1977b). VPg is found attached to these nascent RNA's, thus implicating it in RNA replication (Nomoto et al, 1977b). Antibodies against VPg inhibit the poliovirus replicase complex in vitro (Baron and Baltimore, 1982b; Semler et al, 1982; Dasgupta et al, 1982). Some evidence exists that the replication complex of which the RI forms the central part may be membrane-linked (Semler et al, 1982; see, Putnak and Phillips, 1981) and VPg precursors are found in membrane-linked fractions from infected cells (Takegami et al, 1983). The asymmetry of virus-directed transcription, i.e. the bulk of RNA found in infected cells (especially during the late exponential phase of infection) is positive-sense, has never been fully explained, but probably relies on some property of the RI-replicase complex (Perez-Bercoff, 1979).

Recently, Baltimore and his colleagues have purified from poliovirus infected cells enzymatic activities which transcribe poliovirus vRNA in vitro. A high molecular weight, soluble protein of approximately Mr 77,000 was first isolated, but further purification resulted in a lower molecular weight membrane-associated, virus-coded protein, called "p-63" from its mobility on SDS-polyacrylamide gels (Dasgupta et al, 1979). The Mr 77,000 protein was found to be able to copy poliovirus RNA in vitro without added primer, but p-63 required added oligo(U) as a primer. Progressively purer fractions of p-63 were increasingly specific for poliovirus RNA as template. Later work has shown that p-63 is actually a protein of Mr 53,000 and that a soluble host cell protein (Mr 67,000) could replace the oligo(U) primer requirement (Baron and Baltimore, 1982b,c; Dasgupta, 1983a). Antibodies against VPg inhibited host factor dependent but not oligo(U) primed transcription, further evidence for the involvement of VPg in the de novo initiation of RNA synthesis (Baron and Baltimore, 1982a). Antibodies against the host cell factor strongly inhibit the initiation of RNA synthesis but did not affect the elongation of previously initiated strands (Dasgupta et al, 1982) and precipitate
p-63 from poliovirus infected cells (Dasgupta, 1983b). Interesting though these results are, the p-63 host factor replication complex displays different kinetics of RNA synthesis in vitro from those observed in vivo and the products made in vitro seem to resemble RF rather than RI forms. Recent experiments using antisera raised against synthetic copies of virus-coded peptides (Baron and Baltimore, 1982c) or purified virus polypeptides (Semler et al, 1983) strongly suggest that P3-4b forms an essential part of the poliovirus replicase complex. Figure 1.2 is a diagram of the events believed to occur in picornavirus infected cells.

Virus assembly - Baltimore (1969) stressed the lack of temporal control over virus induced processes in the infected cell. The lack of internal initiation/termination signals in the virus genome and the tight coupling of transcription to translation offers only limited opportunities for control of the intracellular events during the growth cycle. Post-translational cleavages are undoubtedly the major control process during picornavirus growth (see above). However, at the most basic level, control is also provided by simple chemical kinetics which control the interaction of virus components in the infected cell (Baltimore, 1969). It is known that the initial multiplicity of infection affects the timing of intracellular events, but kinetic control of virus replication can perhaps best be seen during the assembly of virus particles. This complex subject has recently been reviewed by Putnak and Phillips (1981) and the events which are thought to occur are summarised in figure 1.3. It is possible that alternative mechanisms of virus assembly may occur in different picornavirus groups. The ratio of virus particles:plaque-forming units is at least 60:1 for poliovirus (Baltimore, 1969) and may be much higher, e.g. 2000:1 for rhinoviruses (Butterworth et al, 1976), indicating that the process of assembly may not be either highly efficient or infallible. Temperature sensitive mutants of poliovirus with altered structural proteins have recently been used to provide further evidence for the fact
FIGURE 1.2
PICORNAVIRUS REPLICATION

- Adsorption to susceptible cells
  - Entry
  - Uncoating
  - VPg cleaved from vRNA
  - Polysome formation, translation
    - Polyprotein processing (Non-structural Proteins?)
    - Replicative form?
    - Positive strand synthesis
      - vRNA (+VPg)
    - Replicative intermediate
      - mRNA (-VPg)
    - Negative strand synthesis
      - Replicative intermediate
  - Virion assembly
    - Translation, processing (Structural proteins?)
    - Maturation
    - Cell lysis and virus release
that the main morphogenetic pathway of poliovirus is that shown in figure 1.3 (see, Drescher-Lincoln et al, 1983).

1. D. 2 - Picornavirus Genetics

Cooper (1969) defined genetic analysis as "an attempt to describe the structure and function of the genetic material by means of its behaviour in an altered (mutant) condition". The key point of this definition is the way in which it relates the structure of the genetic material to its function. However, it has proved difficult to relate picornavirus genes to specific functions, especially in the regions of the genome which code for non-structural proteins. In recent years, nucleic acid cloning and nucleotide sequencing has contributed greatly to knowledge of the fine structural details of picornaviruses (see section 1.E.4), but understanding of the functional aspects of these viruses has progressed more slowly (Baltimore, 1982). For this reason, the conventional type of (recombinational) genetic analysis was regarded as very important, at least until recently.

Genetic recombination in picornaviruses (the first reports of genetic recombination between RNA molecules) was first observed by Hirst (1962) and Ledinko (1963). Experiments with poliovirus showed that recombination occurred at the relatively low frequency of 0.3-0.4%. These recombination frequencies are much lower than those seen in other experimental systems, e.g. influenza A virus, where frequencies of 30-40% are observed (e.g. Simpson and Hirst, 1968), but the latter observations, although commonly called "recombination", are in fact due to the reassortment of genome segments. There is no evidence to suggest that intrasegment recombination occurs in influenza virus. Similarly, in large, unsegmented negative-stranded RNA viruses such as vesicular stomatitis virus and Newcastle disease virus, there is no evidence of recombination. Cooper and his colleagues (1968, 1975) used the observation of picornavirus recombination
Figure 1.3
Assembly of poliovirus particles

[NCVP
P1-1a]

[13 S PARTICLE]

[VP1
VP3
VP0 5]

[14 S PARTICLE]

[70 S EMPTY CAPSID]

[VP1
VP3
VP0 60]

[75 S PROCAPSID]

[RNA]

[150 S PROVIRION]

[VP1
VP2
VP3
VP4 60]

[150 S MATURE VIRION]
to derive a genetic map of the poliovirus genome. Many mutants of poliovirus have now been isolated. Cooper (1969) lists 17 selectable and 23 non-selectable classes. The most commonly used selectable markers are temperature sensitivity, guanidine resistance and resistance to serum inhibitors, while non-selectable markers include a diverse range of properties such as adsorption/elution to various compounds and many modifications of cellular functions. Using many of these characters followed by physiological characterisation of the mutants, Cooper et al (1968, 1975) produced a genetic map of poliovirus in which the 5' half of the genome was shown to code for structural proteins while mutations in the 3' half resulted in defects in RNA synthesis, in agreement with the results of Summers and Maizel (1971).

Some problems were experienced in performing recombination experiments with poliovirus. One problem has been the way in which many of these mutations affect the biological properties of the virus, especially that of virulence (McCahon et al, 1981). In addition, the nature and extent of the changes induced by chemical mutagenesis is uncertain, thus making interpretation of the results of recombination experiments difficult (e.g. Bengtsson, 1968). These low recombination frequencies make such experiments difficult to carry out. Another difficulty was pleiotropism or covariation, in which variation in one character would often accompany variation in another. Reversion was a serious problem, as the most stable (single) mutants isolated still tended to revert at a frequency of about $10^{-4}$ per particle per generation. Cooper et al were unable to produce revertant free virus stocks even after several plaque purifications. Many of the mutants isolated were also "leaky", i.e. retained wild-type function to some extent under restrictive conditions. In spite of these problems, an internally consistent recombination map of the poliovirus genome was eventually produced.

Although similar studies had previously been made on FMDV (e.g. McCahon et al, 1977), fresh interest was aroused in the genetic analysis
of picornaviruses when it was shown that ts mutants of FMDV possessed biochemically detectable alterations in their structural proteins (King and Newman, 1980; King et al., 1980a). Many mutants were isolated and it was then possible to produce a genetic map of FMDV (King et al., 1982). Analysis of unselected markers in the recombinants provided definitive proof of the occurrence of recombination in picornaviruses. A recent report has repeated similar experiments to these with poliovirus, by obtaining ts/guanidine resistant recombinants, using the virus serotype as a stable non-selected marker. These experiments definitively identify the progeny of two clearly distinct parent strains and thus prove that the viruses selected are due to recombination not reversion (Tolskaya et al., 1983).
Although historically a wide variety of techniques have been successfully used to attenuate many organisms (Almond and Cann, 1984), two main methods have been applied to the production of attenuated vaccines. The first of these is host or tissue adaptation and the second the isolation of temperature sensitive strains. Host adaptation is for the most part empirical and repeated passages in alternative host animals or tissue types has been found to give rise to attenuated strains, e.g. the host adaptation of FMDV for mice or chick embryos (Skinner, 1960). Early attempts to produce a live-attenuated poliomyelitis vaccine by adaptation of neurovirulent strains of poliovirus to growth in mice or chick embryos produced unsatisfactory results (Sabin, 1956, 1965). However, there were good grounds for thinking that the isolation of temperature sensitive strains might well lead to the development of a suitable live vaccine. Lwoff (1959, 1969) has demonstrated that virus infections are modified by supra-optimal temperatures. McCahon et al (1981) demonstrated that ts mutations could reduce the pathogenicity of FMDV to the level where apparently avirulent strains were produced.

In practice, Sabin's attenuated strains were derived purely empirically after many passages of neurovirulent wild-type polioviruses of all three serotypes \textit{in vitro}. This was achieved by repeated passage of the wild-type viruses in tissue culture. Variants which displayed lower neurovirulence than the parental strains were selected for further passages. The whole process is summarised in table 1.3 (see, Sabin and Boulger, 1973). The derivation of the type 1 and type 3 vaccines was very similar, comprising approximately 50 passages \textit{in vitro} with about 20 passages in rhesus and cynomolgus monkeys \textit{in vivo}. The type 2 vaccine strain was derived by rather fewer passages (table 1.3). The most notable distinction between the 3 vaccines strains lies in the origin and nature of the parental strains used rather than in the process of attenuation.
<table>
<thead>
<tr>
<th>Type</th>
<th>Strain</th>
<th>In Vivo Passages</th>
<th>In Vitro Passages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1/Mahoney/41</td>
<td>(faecal isolate, healthy children)</td>
<td>14 in vivo</td>
<td>2 in vitro (monkey testicle)</td>
</tr>
<tr>
<td>P1/Mahoney/Monk14 T2</td>
<td></td>
<td>24 in vitro</td>
<td>18 in vitro (monkey kidney)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 alternate</td>
<td>5 in vivo (intradermally in monkeys)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>passages</td>
<td>5 in vitro (monkey kidney)</td>
</tr>
<tr>
<td>P1/Mahoney/LS-c</td>
<td></td>
<td>5 in vitro</td>
<td>3 plaque purifications (monkey kidney)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 in vitro</td>
<td>2 plaque purifications (monkey kidney)</td>
</tr>
<tr>
<td>P1/LS-c, 2ab/KP₂/56</td>
<td>Sabin vaccine strain.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2/P712/56</td>
<td>(faecal isolate, healthy children)</td>
<td>4 in vitro</td>
<td>3 plaque purifications (monkey kidney)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 in vivo</td>
<td>1 in vivo (orally in chimpanzees)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 plaque</td>
<td>3 plaque purifications (monkey kidney)</td>
</tr>
<tr>
<td>P2/P712, Ch, 2ab/KP₂/56</td>
<td>Sabin vaccine strain.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3/Leon/37</td>
<td>(isolate from fatal paralytic case)</td>
<td>21 in vivo</td>
<td>8 in vitro (monkey testicle)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39 in vitro</td>
<td>3 plaque purifications (monkey kidney)</td>
</tr>
<tr>
<td></td>
<td>Sabin (1954)</td>
<td></td>
<td>3 in vitro (preparative) (monkey kidney)</td>
</tr>
<tr>
<td>P3/Leon 12a₁b KP₃/56</td>
<td>Sabin vaccine strain.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The type 1 and 2 strains (P1/Mahoney/41 and P2/P712/56) were isolated from the faeces of healthy children and were not associated with paralytic cases, but the type 3 strain P3/Leon/37 was isolated from the brain stem and spinal cord of a fatal case of paralytic poliomyelitis (Sabin and Boulger, 1973). This difference might well be significant in the light of the relative stabilities of the three vaccine strains and the high incidence of paralytic poliomyelitis associated with the type 3 vaccine (see above). The type 3 Sabin strain also has a higher tendency to revert to neurovirulence on passage in tissue culture than the type 1 and 2 strains (Boulger et al., 1979; Marsden et al., 1980).

Some differences in the properties of neurovirulent and wild-type strains have been observed. Early studies attempted to correlate physical properties of the virus particle with attenuation or virulence (Dubes and Wenner, 1957). For example, it was observed that the so-called d characteristic (which relates to efficiency of plating in cell culture under acid overlay medium) showed some correlation with the degree of neurovirulence (Vogt et al., 1957). Other workers observed a differential affinity between attenuated and virulent strains for various compounds such as calcium phosphate (Hodes et al., 1960) cellulose resin (Woods and Robins, 1961) or aluminium hydroxide (Koza et al., 1963). Differences have also been observed in the degree to which various strains aggregate in conditions of low ionic strength (Totsuka et al., 1978). The P1/Sabin vaccine strain is more labile in the presence of SDS than the virulent P1/Mahoney strain from which it was derived (Young and Moon, 1975). Kew et al. (1980) were able to demonstrate amino acid changes in three of the four capsid proteins of P1/Sabin from those of P1/Mahoney. These results suggest that attenuated strains have altered physical and particularly surface properties, a conclusion that is further strengthened by the observation that there are small antigenic differences between P1/Sabin and P1/Mahoney (Nakano et al., 1978; van Wezel and Hazendonk, 1979). The latter finding is confirmed by the nucleotide sequence of the genomes of
these two strains, interpreted in light of recent information on the location of a major antigenic site on the virus particle (section 1.F).

For none of the above physical changes has it been possible to show a cause and effect relationship with attenuation. However, the possibility that attenuation involves physical changes to the surface of the virus particle is an attractive model. Sabin suggested that the neurovirulence of poliovirus may be determined by its specific interaction with neuronal target cells (Sabin, 1956). Mutations giving rise to a reduced affinity of the virus for nerve cells at the level of cell-receptor binding would therefore possess an attenuated phenotype. Such a model suggests that a difference may exist between virus recognition of cells of neural origin and of gut cells, since attenuated strains are perfectly capable of replication in the alimentary tract. At the present time, there is no evidence for such a difference. The possibility that Sabin's vaccine strains are attenuated simply because they have acquired random ts mutations has been considered (Lwoff, 1959, 1969; Sabin, 1961). Certainly all the vaccine strains exhibit a reduced capacity for growth at 40°C. Such ts mutations could themselves affect the surface properties of the virus and thereby influence cell-virus interactions. However, it is interesting to note that there are reports of ts mutations in P1/Sabin, affecting later stages of the replication cycle, e.g. translation of virus-specific RNA (Denniston-Thompson and Tershak, 1976) and virus assembly (Fiszman et al, 1972).

Analysis of picornavirus genomes by T1 oligonucleotide fingerprinting has already found wide application in strain differentiation, has been used to identify and follow strains of virus during outbreaks of disease (King et al, 1981; Kew et al, 1981; Nottay et al, 1981) and has provided strong molecular evidence for the occurrence of vaccine-associated poliomyelitis (Minor, 1982). Nomoto et al (1981b) used this technique to compare the genomes of the type 1 Sabin vaccine strain and its neurovirulent progenitor, P1/Mahoney. This analysis revealed that no major
changes such as large insertions or deletions had occurred during attenuation. A total of 7 point mutations were detected, making it possible to estimate that approximately 35 mutations had occurred in the whole genome. However, the changes observed did not correlate in any observable way with the attenuation of neurovirulence. Determination of the complete nucleotide sequences of these viruses was required to provide a more definitive analysis of the attenuation process.

Nomoto et al. (1982b) showed that there were 57 nucleotide changes (resulting in 21 amino acid changes) between the P1/Mahoney and P1/Sabin genomes. 12 out of the 21 amino acid changes observed by Nomoto et al. were in the capsid proteins. 5 of these were clustered within 19 amino acids in a region of the VP1 polypeptide. Such a cluster of changes might well be responsible for attenuation if it were due to the differential binding of neurovirulent and attenuated viruses to different tissues. However, an analysis of monoclonal antibody resistant mutants of P3/Leon/37 (Minor et al., 1983; Evans et al., 1983) revealed that the mutations responsible for the escape of antibody neutralisation in poliovirus type 3 were located in the analogous region of the VP1 protein to the cluster of changes observed by Nomoto et al. (1982b) (section 1.E). Since the type 1 vaccine strain, unlike the type 3 strain, was derived by intradermal passage in monkeys during its attenuation (table 1.3), it is entirely possible that this cluster of changes represents antigenic variation and may not be related to attenuation. The initial results obtained from nucleotide sequencing studies were, therefore, disappointing and the identity of the attenuating mutations in poliovirus type 1 remains obscure.
1.F MOLECULAR CLONING OF PICORNAVIRUSES

The molecular cloning and nucleotide sequencing of virus genomes offers the possibility of being able to precisely define the differences between neurovirulent and attenuated strains. Kacian and Myers (1976a,b) demonstrated that it was possible to synthesise a complete DNA copy (cDNA) of the poliovirus genome by reverse transcription of vRNA. Several groups have now used cDNA cloning to study the genomes of picornaviruses. Kupper et al (1981) used double-stranded cDNA cloning to obtain a copy of the region of the genome of FMDV encoding VP1, the immunogenic capsid protein, which was then used to direct the synthesis, in E. coli, of a polypeptide with antigenic properties. A similar approach has also been used to obtain nucleotide sequence data from the antigenic regions of several serotypes of FMDV (Kurz et al, 1981; Boothroyd et al, 1981; Makoff et al, 1982; Villaneuva et al, 1983).

Racaniello and Baltimore (1981a) determined the nucleotide sequence of the P1/Mahoney strain of poliovirus from cloned double-stranded cDNA, while Wimmer and his colleagues determined the same sequence directly from cDNA without recourse to molecular cloning (Kitamura and Wimmer, 1980; Kitamura et al, 1981). (van der Werf et al, (1981) subsequently cloned most of the P1/Mahoney genome by homopolymeric tailing of RNA-cDNA hybrids and insertion into a complementary tailed plasmid vector.) Although there were 4 nucleotide sequence differences between these two sequences, they were generally in agreement and confirmed many of the features of the poliovirus genome which had previously been determined by other means (sections 1.C, 1.D.2). As mentioned in the previous section (section 1.E), Nomoto et al (1982a, 1982b) cloned the genome of the P1/Sabin vaccine strain (using restriction endonuclease fragments of double-stranded cDNA) and determined its nucleotide sequence for the purpose of a comparison with P1/Mahoney.

A major conceptual and technical advance occurred when Racaniello and
Baltimore (1981b) demonstrated that it is possible to regain viable virus by transfection of susceptible cells in tissue culture with cloned cDNA copies of the poliovirus genome. This discovery means that it is now possible to further analyse the mutations identified by nucleotide sequencing by the construction of recombinant viruses in vitro. This technique should make it possible to overcome the difficulties experienced with in vivo recombination experiments (section 1.D.2) and has increased the importance of being able to obtain cloned cDNA copies of picornavirus genomes. Similarly, the expression of antigenic proteins from FMDV (Kupper et al., 1981; Kleid et al., 1981) and from poliovirus (van der Werf et al., 1983; Wychowski et al., 1983) in E. coli has also added to the value of molecular cloning.

Information obtained by nucleotide sequencing studies has been used to synthesise antigenic peptides in vitro. Such peptides may be used both to induce protective immunity against picornavirus infections (Bittle et al., 1982; Emini et al., 1983b; Dr. G. C. Schild, personal communication) and to investigate the role of non-structural proteins in picornavirus replication (Baron and Baltimore, 1982c; Semler et al., 1982; Morrow and Dasgupta, 1983). Nucleotide sequence data is also playing an important role in the identification of the precise epitopes involved in virus neutralisation (Minor et al., 1983; Evans et al., 1983; Haresnape and McCahon, 1983; Haresnape et al., 1983) and may soon help to elucidate the cellular receptors necessary for picornavirus infections (Campbell and Cords, 1983; Dr. P. D. Minor, personal communication).
CHAPTER 2

MOLECULAR CLONING OF THE POLIOVIRUS GENOME
2. A INTRODUCTION

Over the last decade, the rapid development of DNA manipulation techniques has facilitated the molecular analysis of virus genomes in vitro. In particular, progress in DNA sequencing (Sanger et al., 1977; Maxam and Gilbert, 1977; Biggin et al., 1983) has made it feasible to compare the nucleotide sequences of related virus genomes and to define the genetic basis of their differing biological properties. These techniques are not applicable to RNA and it is necessary to obtain a DNA copy of an RNA genome prior to such analyses. Therefore, the starting point for the analysis of an RNA virus genome is the synthesis of a DNA copy (cDNA) of the RNA template. Although restriction endonuclease mapping and nucleotide sequencing studies of picornavirus genomes have been carried out directly on cDNA (Kupper et al., 1981; Kitamura et al., 1981), it is usually more convenient to obtain a cloned cDNA copy of the virus genome in a suitable vector molecule in E. coli, (for a review of this subject, see Williams, 1981). The molecular cloning of RNA virus genomes thus provides an abundant supply of working material ideally suited for further analysis in vitro.

The basic strategy of cDNA cloning, i.e. the synthesis of a cDNA copy from an RNA template and its insertion into a vector molecule, has been adapted by different investigators and a variety of experimental procedures have been devised which vary widely in their complexity and efficiency. In most protocols, transcription of a polyadenylated RNA template is primed by hybridisation of oligo(dT) to the 3' poly(A) tract. Avian myeloblastosis virus (AMV) reverse transcriptase is then used to synthesise a single-stranded complementary DNA. Following alkaline hydrolysis of the RNA, the single-stranded cDNA is rendered double-stranded by synthesis of a second cDNA strand. E. coli DNA polymerase I, the large "Klenow" fragment of the same enzyme and reverse transcriptase are all capable of carrying out this reaction. The double-stranded cDNA
contains a covalently closed "hairpin" structure at its 5' end, resulting from the internal self-priming of second-strand synthesis. After single-strand specific nuclease treatment to remove the hairpin, the flush-ended, double-stranded cDNA is inserted into a suitable vector molecule, either with DNA ligase or by annealing of complementary homopolymer "tails" added to both cDNA and vector using terminal deoxynucleotidyl transferase (Jackson et al., 1972; see, Williams, 1981).

Various adaptations have been made to the basic strategy of double stranded cDNA cloning described above. For example, Kurtz and Nicodemus (1981) used restriction endonuclease "linkers" to clone globulin cDNA with high efficiency. Vitek et al. (1981) synthesised cDNA to mRNA's bound to oligo(dT)-cellulose, a method designed to facilitate the cloning of low abundance mRNA's such as Drosophila ecdysterone. Several methods designed to allow the cloning of the 5'-proximal sequences of RNA templates have been designed (Cooke et al., 1980; Land et al., 1981; Okayama and Berg, 1982; Heidecker and Messing, 1983). These are significant because sequences corresponding to the 5' end of the template are usually lost during the single-strand nuclease treatment described above (Shenk et al., 1975; Higuchi et al., 1976). Some of these last methods, though claimed to be efficient in terms of colony yield/µg of RNA template, e.g. 1 x 10^5 colonies/µg of reticulocyte mRNA (Okayama and Berg, 1982), are complex and require many different sequential enzymatic and manipulative steps.

An alternative to double-stranded cDNA cloning is to clone RNA-cDNA hybrid molecules, a procedure which has received much less attention. After reverse transcription, the resulting hybrid molecule is homopolymer tailed using terminal deoxynucleotidyl transferase and is then annealed to a vector molecule bearing complementary homopolymer "tails" (e.g. Rabbitts, 1976; Wood and Lee, 1976; Zain et al., 1979). Although early reports implied that the efficiency of RNA-cDNA hybrid cloning was low, e.g. 1 x 10^2 colonies/µg RNA (Zain et al., 1979), the investigators did not compare their hybrid cloning procedure with parallel double-stranded cDNA
cloning experiments. Thus the true relative efficiency of this method was unknown.

In the work described in this chapter, RNA-cDNA hybrids were inserted into the vector molecule pAT 153 (Twigg and Sherrat, 1980) by means of homopolymer tailing. The vector was cleaved with PstI and tailed using dGTP. The RNA-cDNA hybrids were tailed using dCTP. After being annealed together under favourable ionic conditions, the hydrogen-bonded DNA could be used directly to transform E. coli and was repaired in vivo to form covalently closed circular molecules (Wensink et al., 1974). This process results in the regeneration of PstI restriction endonuclease cleavage sites on either side of the cDNA insert (Otsuka, 1981) and has the advantage that it allows precise excision of the cloned cDNA insert (bearing homopolymer tails) by PstI digestion. This strategy was chosen in preference to the direct ligation of vector and cDNA, as it had the important advantage that tailed cDNA molecules would not anneal to other homologously tailed cDNA molecules, preventing the formation of concatemers of cDNA fragments. This was important as at the outset of this study, no restriction maps or sequence information from polioviruses were available and such concatemers could have caused serious problems during subsequent restriction endonuclease mapping and nucleotide sequencing experiments. Tailing was also preferred to the cloning of restriction endonuclease fragments of double-stranded cDNA (Nomoto et al., 1982a, b). Apart from the fact that there was no information available on the restriction endonuclease maps of poliovirus cDNA's, this technique is probably not suitable for the cloning of RNA-cDNA hybrids, since although these molecules can be cleaved with certain restriction endonucleases (Cann et al., 1983) (though not those with hexameric recognition sites most useful for this purpose, see Nomoto et al., 1982a, b) it was not known whether it would be possible to ligate these fragments into a vector molecule, because of the RNA strand. Enzymes which are known to catalyse RNA-DNA joining reactions, e.g. T4 DNA ligase (Nath and Hurwitz, 1974) and
T4 RNA ligase (Higgins et al., 1979) catalyse this reaction only inefficiently. This chapter describes the necessary reactions performed during the molecular cloning of the genome of P3/Leon 12 a₁b₁, i.e. the optimisation of cDNA synthesis, homopolymer tailing of cDNA and vector and transformation of competent bacteria with recombinant plasmids.
2. B PREPARATION OF VIRUS STOCKS FOR MOLECULAR CLONING

The work described in this section (2. B) was performed by Drs. P. D. Minor and P. Reeve at the National Institute for Biological Standards and Control, London, UK and is recorded here as a background to the molecular cloning studies described in this chapter. Three closely related strains of poliovirus type 3 with carefully documented passage histories which were believed to form a direct lineage were chosen for study. P3/Leon/37 was isolated from a fatal case of paralytic poliomyelitis in Los Angeles, USA, in 1937 (see, Sabin and Boulger, 1973). The strain P3/Leon/37 used in this study was obtained from Dr. A. B. Sabin via Dr. O. M. Kew of the Center for Disease Control, Atlanta, USA. This isolate is significantly different in its molecular and biological properties from the strain deposited under the same name in the American Type Culture Collection (Dr. P. D. Minor, personal communication). P3/Leon 12 a,b was derived from P3/Leon/37 by Dr. A. B. Sabin and is the present World Health Organisation (WHO) approved vaccine strain. The vaccine stock used in this work had undergone one further passage from Sabin's original (SO) stock (Sabin and Boulger, 1973), i.e. SO+1 passage level (WHO, 1969). P3/119 was isolated from a fatal case of paralytic poliomyelitis associated with the administration of the Sabin oral poliomyelitis vaccine and had previously been well characterised as part of a WHO study on vaccine-associated paralysis (Minor, 1980, 1982).

It is generally believed that errors in genome replication are of the order of $1 \times 10^6$ times more frequent for RNA than for DNA (see Holland et al., 1982). Such high mutation rates may present a potential problem when RNA virus genomes are analysed using cDNA cloning, since an individual copy of the virus genome is likely to represent a minority of the population rather than the consensus sequence. However, the commonly accepted mutation rates for RNA genomes may not be accurate (see Palese and Young, 1982) nor universally applicable to all virus-host systems.
Furthermore, the rate of genome mutation may vary among different groups of RNA viruses.

The three strains of poliovirus used in this study were characterised by careful biological and biochemical analysis prior to molecular cloning. Since the history of plaque purification of these three strains was uncertain and in order to minimise the possibility that some of the cDNA clones might be derived from a minor population with different biological properties, the parental stocks of each strain were plaque purified on Hep2c cell monolayers (Minor, 1980). Virus from a single plaque was used to grow a small primary pool of each strain. These primary pools, designated P3/Leon/37 #960, P3/Leon 12 a,b #411 and P3/119 #643, were tested by observation of their plaque morphology on Hep2c cell monolayers over a range of temperatures from 35-40°C, by SDS-polyacrylamide gel electrophoresis of the polypeptides induced in infected Hep2c cell monolayers (Minor, 1980) and by T1 oligonucleotide fingerprinting (Minor, 1982). The results of these analyses were as follows:

P3/Leon/37 - Slight alteration in plaque morphology (smaller plaques) at 39°C but no decrease in number of plaques.

P3/Leon 12 a,b - 1000 fold decrease in plaques at 39°C, alteration in plaque morphology at 38°C.

P3/119 - As P3/Leon/37.

All three strains showed a significant decrease in the number of plaques produced at 40°C. The plaque-purified primary pools were shown to have identical properties to the non-plaque-purified parental stocks and induced identical patterns of infected cell polypeptides.

The results obtained from the T1 oligonucleotide fingerprint analysis showed that the migration positions of the 55 larger, characteristic
oligonucleotides were identical for each strain (Minor et al., 1983; Stanway et al., 1983b). This was confirmed by co-electrophoresis of mixtures of RNA digests from pairs of strains, none of which showed any additional spots to the individual fingerprints. Although this method examines only approximately 15% of the total genome sequence, large RNA molecules which produce identical T1 fingerprints may be expected to be greater than 99% homologous in nucleotide sequence (Aaronson et al., 1982). Hence it was shown that the molecular changes responsible for attenuation and reversion to neurovirulence must compose 1% or less of the total genome, (i.e. 74 nucleotides) and therefore probably would not include extensive deletions, insertions or rearrangements of the genome. Since so few changes were present, any differences that might be observed between the strains are likely to be directly related to the known biological differences between them. The similarity between the T1 oligonucleotide maps, plus the fact that the three strains were believed to form a direct lineage were the major reasons why these three particular strains of poliovirus were chosen for this study.

The primary pools of virus were examined by the WHO-approved monkey neurovirulence test (WHO, 1982) and again, the plaque-purified isolates were found to be identical to the parental stocks. The results of these tests are shown in table 2.1. As expected, all the animals inoculated with the vaccine strain P3/Leon 12 a1b #411 survived free from paralysis until day 22 when they were sacrificed. This result in the WHO neurovirulence test would permit the release of this preparation as a vaccine. In contrast, P3/Leon/37 #960 and P3/119 #643 were both highly neurovirulent, causing paralysis in all the animals tested (Stanway et al., 1983b).
## TABLE 2.1
MONKEY NEUROVIRULENCE TEST RESULTS

<table>
<thead>
<tr>
<th>Virus (Clone#)</th>
<th>Number of animals tested</th>
<th>Number of animals paralysed</th>
<th>Mean time before onset of paralysis (days)</th>
<th>Mean histological lesion score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P3/Leon/37 #960</td>
<td>6</td>
<td>6/6</td>
<td>8</td>
<td>2.91</td>
</tr>
<tr>
<td>P3/Leon 12 a,b #411</td>
<td>18</td>
<td>0/18</td>
<td>22</td>
<td>0.72</td>
</tr>
<tr>
<td>P3/119 #643</td>
<td>6</td>
<td>6/6</td>
<td>3</td>
<td>3.06</td>
</tr>
</tbody>
</table>

Mean histological lesion scores were calculated by the standard WHO scoring method (WHO, 1982):

0 = No specific damage
1 = Cellular infiltration only
2 = Cellular infiltration and minor neuronal damage
3 = Extensive neuronal damage + cellular infiltration
4 = Massive neuronal destruction + cellular infiltration
2.C VIRUS RNA EXTRACTION AND ANALYSIS

2.C.1 - RNA Extraction

RNA was extracted from virus purified on sucrose gradients either by direct phenol extraction in the presence of SDS, or by phenol extraction following proteinase K digestion of the virus capsid (see chapter 6). Since RNA is relatively labile (e.g. compared with DNA), the extraction and subsequent treatment of RNA is likely to have a considerable influence on the results obtained from molecular cloning experiments. Two major factors which influence the efficiency of cDNA synthesis and cloning are the use of nuclease free enzyme preparations throughout and the initial quality of the RNA template used in the reverse transcription reaction. The quality of RNA extracted from virus stocks was therefore closely examined (see below). Considerable precautions were taken to ensure that RNA preparations used in this study were intact and had not been subjected to chemical, physical or enzymatic degradation. Since cDNA synthesis is primed from the 3' poly(A) tract using p(dT)$_{17}$, this was important if cDNA clones corresponding to the entire genome of the virus (including the 5' terminus) were to be obtained. RNA extracted from virus preparations was ethanol precipitated, dried in vacuo and redissolved in sterile distilled deionised water.

2.C.2 - Analysis of Virus RNA

Small aliquots of RNA (0.1-1 µg) were subjected to electrophoresis on 1% (w/v) agarose gels in Elfo buffer, containing 0.5 µg/ml EtdBr. Some attempts were made to examine purified RNA preparations by sedimentation on 15-30% (w/v) sucrose gradients, but the use of agarose gels, on which the RNA was visualised by EtdBr staining and UV illumination, proved to be a much more convenient method. Very small quantities of RNA (0.1 µg) could
be used for gel analysis (without radiolabelling) and the method proved to be both rapid and convenient. Non-degraded RNA preparations, such as that shown in figure 2.1, showed a single, discrete EtdBr-staining band, migrating at about the same rate as linear single-stranded M13 mp7 DNA, (approximately 7.3 kb). Preparations where lower molecular weight material (resulting from either the degradation of full-length RNA or contamination with cellular RNA) was predominant were not used for molecular cloning.

As can be seen in figure 2.1, some batches of RNA examined by agarose gel electrophoresis showed two EtdBr-staining bands. It was demonstrated that the more slowly migrating band could be removed if the RNA was loaded onto the gel in 50% (v/v) formamide after heating to 65°C for 1 min. The faster migrating band had an estimated size of between 7.0 and 7.5 kb. The composition of the slower migrating band is uncertain, but may be due to either a high degree of secondary structure possessed by the RNA or to aggregation of RNA molecules. This observation was not a constant feature of all RNA preparations examined in this way. The presence or absence of the second band in RNA preparations had no observable effect on the results of reverse transcription.
Figure 2.1

Gel electrophoresis of polio vRNA

Phenol extracted P3/Leon 12a1b #411 vRNA subjected to electrophoresis on a 1%(w/v) agarose - Elfo gel containing 0.5µg/ml EdtBr.

A - P3/Leon 12a1b #411 vRNA loaded in Elfo loading buffer.
B - P3/Leon 12a1b #411 vRNA pre-heated in 50% (v/v) formamide.
C - M13 mp7 RF DNA loaded in Elfo loading buffer.
D - M13 mp7 RF DNA pre-heated in 50% (v/v) formamide.
2. D cDNA SYNTHESIS

2. D. 1 - Reverse Transcription of Poliovirus RNA

Reaction conditions for cDNA synthesis were first determined for eukaryotic messenger RNA templates, e.g. ovalbumin cDNA from hen oviduct mRNA (Monahan et al., 1976; Higuchi et al., 1976) and globin cDNA from reticulocyte mRNA (Maniatis et al., 1976; Efstratiadis et al., 1976). These conditions proved to be close to the optimum for cDNA synthesis from most RNA templates and have been widely used (with occasional slight modifications). Reverse transcription of poliovirus RNA was carried out under similar conditions to those used by other workers (e.g. Buell et al., 1978; Retzel et al., 1980). These conditions (referred to throughout as "standard conditions") are summarised in table 2.2.

<table>
<thead>
<tr>
<th>RNA</th>
<th>50 µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(dT)\textsubscript{17}</td>
<td>5 µg/ml</td>
</tr>
<tr>
<td>Tris-HCl</td>
<td>50 mM, pH 8.3</td>
</tr>
<tr>
<td>MgCl\textsubscript{2}</td>
<td>8 mM</td>
</tr>
<tr>
<td>2-mercaptoethanol</td>
<td>20 mM</td>
</tr>
<tr>
<td>NaCl</td>
<td>50-100 mM</td>
</tr>
<tr>
<td>dCTP/dGTP/dTTP</td>
<td>0.5 mM</td>
</tr>
<tr>
<td>dATP</td>
<td>0.1/0.5 mM (Analytical/Preparative)</td>
</tr>
<tr>
<td>[\alpha\textsuperscript{32}P]dATP</td>
<td>100 µCi/ml</td>
</tr>
<tr>
<td>AMV reverse transcriptase</td>
<td>250-500 u/ml</td>
</tr>
</tbody>
</table>

\textbf{cDNA yield} - The total yield of cDNA from a given reaction was determined by measuring the incorporation of radioactivity into perchloric acid precipitable material as described in chapter 6. While the yield of cDNA
obtained under particular experimental conditions was qualitatively reproducible, it was found that the yields obtained in different experiments were quite variable. For this reason, the effects of altered experimental conditions on cDNA yield are expressed throughout this chapter not as absolute figures but as proportions of that observed in an unaltered control carried out in parallel. Probably the most important factor in using the poliovirus genome as a template for reverse transcription is its size, since the yield of full-length cDNA obtained from long RNA templates such as this is typically much lower than the yield obtained from shorter templates, i.e. less than 2 kb (Buell et al., 1978; Retzel et al., 1980). No measurement was made of what proportion of the total cDNA consisted of full-length reverse transcripts, other than visual estimation from samples subjected to gel electrophoresis (see below).

Under optimum conditions for cDNA synthesis (see table 2.2), the maximum yield of cDNA obtained was approximately 40% of the RNA template input. It is not known why the cDNA yield was not closer to 100%, but possible reasons are the presence of low levels of nuclease activity in the enzyme preparation, which results in the slow destruction of the RNA template as the reaction proceeds, or possibly the proportion of the total RNA input which is in some way "competent" to act as a template for reverse transcription (e.g. primed with p(dT)_17 and not aggregated or degraded).

Analysis of reverse transcription products - In addition to measurement of the total yield, poliovirus cDNA was was analysed qualitatively by electrophoresis on 1% (w/v) alkaline agarose gels (McDonell et al., 1977). These gels and the running buffer contain 30 mM NaOH and have a pH of 12.5. They strongly denature DNA and hydrolyse the RNA strand of RNA-cDNA hybrids, thus cDNA's migrate as single-stranded molecules. Alkaline agarose gels were used to analyse the products synthesised in the reverse
transcription reaction because of their convenience. cDNA's were subjected to electrophoresis at 90 V for 2-3 h and the gels dried onto glass plates prior to autoradiography. Figure 2.2 shows an autoradiograph of poliovirus cDNA synthesised under standard conditions, extracted once with an equal volume of phenol/chloroform (1v/1v) and subjected to electrophoresis on such a gel (1 x 10^5 – 1 x 10^6 cpm/track). Also shown in this figure are the products of a reverse transcription of influenza A/England/321/77 vRNA (Hauptmann et al, 1983), synthesised under identical conditions (track A) which range in size from 2.3 kb (genes 1 and 2) to 0.9 kb (gene 8) and ^32p-labelled DNA size markers of 5.9 and 7.9 kb (track D) derived from an EcoRI digestion of the 13.8 kbp plasmid pLG 215 (Wilkins et al, 1981).

Although the poliovirus cDNA shows a heterogeneous size distribution, dark bands are visible against the background of heterogeneous material. In view of the homogeneity of the RNA template (section 2.C) and the specific priming of reverse transcription at the 3' end of the genome (see below), the wide size distribution of the cDNA is surprising. This heterogeneity is probably due to non-specific premature termination of transcription by reverse transcriptase (Kacian and Myers, 1976a) and the darker bands may possibly be due to sequence-specific termination of transcription (e.g. due to the nucleotide composition or the secondary structure of the RNA at certain points).

A small amount of cDNA apparently larger than 7.4 kb is visible in figure 2.2. This may be due to "hairpin" self-priming of second-strand synthesis by the enzyme (see, Williams, 1981), or alternatively may be a gel artefact. As might be expected from the T1 oligonucleotide fingerprinting results (section 2.C), identical patterns of cDNA synthesis were seen with the three strains of poliovirus examined (P3/Leon/37 #960 is not shown in figure 2.2). The bulk of the cDNA synthesised under standard conditions was between 1-5 kb in size. However, the relative size distribution of the material synthesised varied slightly under different conditions from the standard reaction (see next section).
Figure 2.2
Analysis of reverse transcription products
Reverse transcription of poliovirus RNA in a standard reaction (see text). Autoradiograph of alkaline agarose gel.

A - Influenza A/England/321/77 reverse transcription.
B - P3/119 #643 reverse transcription.
C - P3/Leon 12 a1b #411 reverse transcription.
D - DNA size markers.
Figure 2.3 demonstrates an important experimental detail which was observed when identical samples from the same reverse transcription reaction shown in figure 2.4 (see below) were subjected to electrophoresis on a supposedly denaturing 3% (w/v) polyacrylamide gel containing 7 M urea and TBE1 buffer. A totally different pattern of migration was observed from that seen in figure 2.4 (an autoradiograph of an alkaline agarose gel) with a compression of material larger than about 2 kb into a tight band, which was at first assumed to represent a large yield of full-length cDNA. However, analysis of the same reaction on an alkaline agarose gel revealed that this was not the case. This artefact of the polyacrylamide gel system was not readily observable with cDNA's smaller than 2 kb, e.g. influenza virus cDNA. The alkaline agarose gel system was therefore used to analyse poliovirus cDNA.

**Kinetics of cDNA synthesis** - A time-course of cDNA synthesis in a standard reaction was monitored by electrophoresis of identical aliquots of the reaction, taken at time points up to 1 h, on an alkaline agarose gel. Figure 2.4 shows an autoradiograph of the dried gel. The pattern of incorporation seen implies that strand elongation is very rapid since the relative proportion of full-length cDNA does not increase later in the reaction. This suggests that some other factor, possibly the initiation of cDNA synthesis, is the rate limiting step in the reaction (see, Williams, 1981). The concentration of deoxynucleotide triphosphates (dNTP's) was in excess of that which would limit the rate of the reaction (see below) (Retzel et al., 1980). Incubation of standard reactions for longer than 1 h did not result in a further increase in the cDNA yield. On some occasions, a slight decrease in the amount of radiolabel incorporated into high molecular weight material was observed, possibly due to low levels of nuclease activity present in the enzyme preparation (as discussed below). Long incubation times have been shown to promote the synthesis of low molecular weight cDNA late in the reaction (Kacian and Myers, 1976b),
Figure 2.3

Time course of reverse transcription reaction subjected to acrylamide gel electrophoresis

P3/Leon 12a,b #411 RNA reverse transcribed in a standard reaction and subjected to electrophoresis on a 3% (w/v) acrylamide – TBE1 gel.

A - Size markers.
B - 0 min.
C - 1 min.
D - 5 min.
E - 15 min.
F - 30 min.
G - 60 min.
although this would not result in a decrease in the total cDNA yield. No significant difference was observed in either the total yield of cDNA or in the size of the cDNA product in standard reactions incubated at 37, 40, 42 or 46°C. For convenience, standard reverse transcription reactions were therefore routinely incubated for 1 h at 37°C. The above observations agree with those published by others (Buell et al, 1978; Retzel et al, 1980).

Pretreatment of the RNA template with 6-10 mM concentrations of methylmercury hydroxide abolishes secondary structure and/or aggregation of RNA molecules and has been reported to enhance the reverse transcription of certain RNA molecules (Payvar and Schimke, 1979). Therefore, attempts were made to determine whether the addition of denaturing agents might promote more efficient synthesis of full-length cDNA. However, pre-treatment of the RNA with 1-10 mM methylmercury hydroxide was unsuccessful, resulting in no discernable increase in the yield of full-length cDNA molecules. The inclusion of 1-10 mM concentrations of spermidine in the reaction mixture were similarly unsuccessful. Inclusion of actinomycin D in the reverse transcription reaction at a concentration of 100 pg/ml has been shown to inhibit the synthesis of short cDNA's, although decreasing the yield of cDNA (Kacian and Myers, 1976b). However, when attempting to repeat these observations, it was found that a concentration of as little as 2.5 pg/ml actinomycin D reduced the total yield of cDNA by 50% or more. For this reason, actinomycin D was not routinely included in reverse transcription reactions.

Attempts were made to increase the yield of cDNA by fresh additions of various reaction components, e.g. buffer, nucleotides, enzyme, etc., at various times throughout the incubation period. These were not successful, producing no increase in either the yield or the size of the cDNA.
**Figure 2.4**

Time course of reverse transcription reaction

P3/Leon 12a\textsubscript{1}b #411 RNA reverse transcribed in a standard reaction and subjected to electrophoresis on a 1\% (w/v) alkaline agarose gel.

A - Size markers.
B - 0 min.
C - 1 min.
D - 5 min.
E - 15 min.
F - 30 min.
G - 60 min.
2.D.2 - Optimisation of cDNA Synthesis

Primer requirement - AMV reverse transcriptase has a requirement for a single-stranded primer complementary to the RNA template. This requirement is met by the inclusion of p(dT)$_{17}$ in standard reactions at a concentration of 5 µg/ml, i.e. 1 µg p(dT)$_{17}$ to 10 µg poliovirus RNA, a molar excess of primer of approximately 40-fold. The precise molar ratio of primer to RNA did not affect the cDNA yield. A decrease in the ratio from 40:1 to 20:1 resulted in a very slight decrease in the cDNA yield but no difference in the size of the cDNA product. Any increase in the ratio, e.g. to 60:1, did not result in an increase in the cDNA yield from that achieved under standard conditions. Omission of p(dT)$_{17}$ from the reverse transcription reaction resulted in a reduction in the yield of cDNA of approximately 80%. When analysed on an alkaline agarose gel this was shown to consist of low molecular weight, heterogeneous material, which presumably results from reverse transcription primed by hydrogen bonding of RNA fragments at random points throughout the genome. In standard reactions, no special priming procedure was performed, p(dT)$_{17}$ was added directly to the reverse transcription reaction mixture with the rest of the components. The low melting temperature of poly(dA.dT) hybrids enables adequate priming of cDNA synthesis to occur at 37°C and an annealing step was not necessary.

Salt concentration - The salt concentration in the reaction mixture was found to be an important factor in influencing the nature of the cDNA product. Figure 2.5 shows an autoradiograph of the products of three standard reverse transcription reactions containing 0, 50 or 100 mM NaCl respectively. In the absence of NaCl (track A), low levels of cDNA synthesis occur, approximately 25% of the level at the optimum salt concentration (50 mM) and the mean size of the cDNA is smaller than that synthesised in the presence of NaCl. Maximal cDNA synthesis occurs at a
Figure 2.5
Salt concentration in reverse transcription reaction
P3/Leon 12 a1b #411 RNA reverse transcribed in standard conditions with varying salt concentrations. Autoradiograph of alkaline agarose gel.

A - 0 mM NaCl
B - 50 mM NaCl
C - 100 mM NaCl
D - DNA size marker

-5.9 kb
concentration of 50 mM NaCl (track B). Above this concentration, the total yield of cDNA decreases. However, at a concentration of 100 mM NaCl, although the yield is only approximately 70% of that at 50 mM NaCl, the mean size of the product is slightly larger (track C). At still higher salt concentrations, e.g. 150 mM NaCl (not shown), the yield of cDNA, which shows a similar size distribution to that seen with 100 mM NaCl, is only approximately 50% of that at the optimum salt concentration. These results are summarised in figure 2.6 and are in broad agreement with those obtained by Buell et al (1978) and Retzel et al (1980), but conflict with those obtained by Monahan et al, (1976), who suggested that low salt concentrations promoted the synthesis of high molecular weight cDNA from an ovalbumin mRNA template. The inhibitory effect of high salt concentrations, i.e. more than 100 mM, on the synthesis of cDNA using AMV RNA (7.9 kb) as template noted by these authors was similar (i.e. 20-30% of maximum yield) to that observed above using poliovirus RNA. These authors went on to note that the substitution of KCl for NaCl partially overcame the inhibitory effect of high salt concentrations on the cDNA yield. However, for poliovirus RNA, no difference in the results obtained with KCl were observed from those achieved with NaCl. Preparative reverse transcription reactions for cDNA cloning were therefore performed in the presence of 100 mM NaCl. This concentration was chosen as the "break-even" point between the synthesis of longer cDNA molecules and unacceptably low total cDNA yield.

pH optimum - Although the pH optimum for the reverse transcription of poliovirus RNA was not fully investigated, no significant difference was observed between the cDNA synthesised in 50 mM Tris-HCl buffer at pH 8.1 and that synthesised at pH 8.3. Retzel et al (1980) observed the pH optimum for both the maximum yield and for the synthesis of full-length cDNA was achieved using a Tris-HCl buffer system at pH 8.1 and that the use of alternative buffer systems (phosphate or cacodylate) did not
Figure 2.6
Graph of cDNA yield against salt concentration
improve the efficiency of the reaction.

**Enzyme concentration** - The concentration of reverse transcriptase in the reaction mixture was found to have a considerable influence on both the yield and size distribution of the cDNA synthesised (figure 2.7). As would be expected, low levels of reverse transcriptase (7 u/µg) resulted in a low yield of cDNA, i.e. approximately 25% of maximum yield (track A). The yield rose to a maximum point (approximately 40% of the RNA template input). In figure 2.7 this is at a ratio of 14 u reverse transcriptase/µg RNA (track B). The precise ratio of cDNA yield to product size was found to vary slightly from one batch of reverse transcriptase to another. A representative pattern of synthesis shown in figure 2.7. It is significant that when the ratio of reverse transcriptase to RNA is increased above the optimum level, not only does the size distribution of the product alter, but the total yield of cDNA decreases under these conditions, with less full-length cDNA synthesised (track C). The average size of the cDNA synthesised at sub-optimal enzyme concentrations appears to be greater than that synthesised at supra-optimal concentrations. This observation is presumably due to low levels of nuclease activity present in the reverse transcriptase preparation and is in agreement with the report published by Buell et al (1978). Alterations in the concentration of 2-mercaptoethanol of between 10-30 mM (standard reverse transcription conditions contained 20 mM 2-mercaptoethanol) had no effect on the synthesis of cDNA with any of the batches of enzyme tested.

**Additions to the standard reaction** - Another component of the reaction mixture which could be expected to influence the synthesis of cDNA is the concentration of dNTP's in the reaction mixture. Kacian and Myers (1976a) reported that 1 mM concentrations of dNTP's promoted the synthesis of full-length poliovirus cDNA. Buell et al (1978) determined the rate limiting dNTP concentration (for cDNA synthesis from hen oviduct mRNA's)
P3 Leon 12a, b #411 RNA reverse transcribed in a standard reaction and subjected to electrophoresis on a 1% (w/v) alkaline agarose gel.

A - 7u reverse transcriptase /µg RNA, (140 u/ml).
B - 14u reverse transcriptase /µg RNA, (280 u/ml).
C - 28u reverse transcriptase /µg RNA, (560 u/ml).
to be 50 μM and found that 1 mM concentrations of dNTP's resulted in increased yield of cDNA and did not affect the size of the product. Retzel et al. (1980) obtained optimal synthesis of full-length AMV cDNA at nucleotide concentrations of 75 μM and obtained similar results to those of Buell et al. with 1 mM dNTP's. Analytical scale reverse transcription reactions of poliovirus RNA described above contained dCTP, dGTP and dTTP at 0.5 mM and dATP at 0.1 mM, all above the rate limiting concentration determined by Buell et al. (1978) and Retzel et al. (1980). dATP was used at lower concentrations than the other nucleotides primarily for reasons of economy, to enable the synthesis of cDNA with a relatively high specific activity (0.1 μCi/μg), which could be readily analysed by gel electrophoresis and autoradiography. Production of high specific activity cDNA was undesirable for cloning experiments. Preparative reverse transcription reactions contained all four dNTP's at 0.5 mM. cDNA produced in these reactions had a specific activity of 0.01 μCi/μg. No differences were observed between preparative and analytical reactions.

Attempts were made to improve the yield of full-length cDNA from the reverse transcription reaction by inhibiting the apparent nuclease activity present in the reverse transcriptase preparations. The inclusion of sodium pyrophosphate in the reverse transcription reaction has been reported to increase the length of the cDNA product. This is believed to be due to the inhibition of nucleases (Myers and Spiegelman, 1978). However, no marked effect was observable when sodium pyrophosphate was included in standard reverse transcription reactions of poliovirus RNA at concentrations of 1, 2 or 4 mM. Inclusion of a ribonuclease inhibitor vanadyl-ribonucleoside complex (Berger and Birkenmeier, 1978) was not successful. Strong inhibition of transcription occurred at concentrations as low as 1 μM. However, use of the recently available human placental ribonuclease inhibitor, (BRL Ltd.), in the standard reverse transcription reaction appears to result in an increase in the cDNA yield of up to 30-40%, although this effect has not yet been fully investigated (Dr. G.
2. D. 3 - Purification of cDNA

Analytical reverse transcription reactions were terminated by phenol extraction before the incorporation of radiolabel was determined and gel electrophoresis carried out. Preparative reactions were stopped by placing on ice, the addition of EDTA to 20 mM and a single extraction with phenol/chloroform/isoamylalcohol/8-hydroxyquinoline (100v/100v/4v/0.8w). The cDNA was then purified by gel filtration through a small (8 cm) Sephadex G100 column formed in a pasteur pipette. This step was a rapid and efficient purification procedure, resulting in the separation of high molecular weight cDNA from the other reaction components and contaminating phenol. The cDNA was purified further by ethanol precipitation of the appropriate column fractions (identified by Cerenkov counting of the $^{32}$P-radiolabelled material in a Hewlett Packard "Prias" scintillation counter) washed with cold 70% (v/v) ethanol and redissolved in an appropriate buffer for the next enzymatic reaction.
2. E HOMOPOLYMER ADDITION

2. E. 1 - The Terminal Deoxynucleotidyl Transferase Reaction

The conditions necessary for homopolymer tailing of DNA using terminal deoxynucleotidyl transferase (terminal transferase) have been thoroughly studied (Roychoudhury et al., 1976; Deng and Wu, 1981). Tailing of RNA-cDNA hybrids however, is much less well understood (Wood and Lee, 1976; Zain et al., 1979). The terminal transferase reaction was found to be of uncertain reproducibility. For this reason, an effort was made to determine the exact conditions required for the addition of homopolymer tracts. Initially, reaction conditions were determined on an analytical scale (25 µl), using vector DNA with protruding 3' ends. PstI digested pAT 153 was used for this purpose. (Figure 2.8 shows the restriction endonuclease map of this vector in relation to pBR 322, the plasmid from which it was derived.) Tailing reactions using either dCTP and dGTP as substrates were examined in this way. It was found to be essential that DNA to be tailed was free from contamination with protein, phenol or salt, as this seemed to disrupt the terminal transferase reaction. For this reason, all DNA to be tailed (vector and cDNA) was carefully purified by phenol extraction, gel filtration through Sephadex G100, ethanol precipitation (at least once) and a wash in cold 70% (v/v) ethanol before redissolving in the reaction buffer.

Tailing of vector DNA - Conditions used for the tailing of pAT 153 are described below. The DNA was prepared by a complete PstI digestion of the plasmid vector pAT 153 with 2 u PstI/µg DNA at 37°C for 2 h. The reaction was monitored to ensure complete digestion by electrophoresis of small aliquots of the mixture on a 1% (w/v) agarose gel in Elfo buffer. Terminal transferase reactions were carried out under similar conditions to those published by other authors (Roychoudhury et al., 1976; Deng and Wu, 1981).
Figure 2.8
Restriction endonuclease cleavage sites and functional regions of pBR 322 and pAT 153

Restriction endonuclease map of pBR 322 (4362 bp). Shaded section is not present in pAT 153 (3657 bp).
However, the buffer system described by Roychoudhury et al (1976), (1.4 M potassium cacodylate, 0.3 M Tris, 1 mM DTT, 10 mM CoCl₂, adjusted to pH 7.6, diluted 10-fold in the reaction mixture), does not give the stated pH value (pH 6.9) on 10-fold dilution, but a higher value of pH 7.4 (R. Hauptmann, personal communication). This results in very poor polymerisation of dCTP residues by terminal transferase. The buffer system used in the experiments described below was therefore slightly modified from the published specification, the 10 x concentrated buffer being adjusted to approximately pH 7.1 and tested to ensure the diluted buffer had a pH of 6.9. This resulted in a much faster and more efficient polymerisation of dCTP residues onto the 3' ends of the DNA. This buffer system is slightly different from that described in a later report (1 M potassium cacodylate, 2 mM DTT, 1 mM CoCl₂) (Deng and Wu, 1981) which was not tested.

Under the above conditions, the addition of dC residues to the 3' ends of the DNA is dependent on factors such as the concentration of enzyme (Nelson and Brutlag, 1979) and most importantly, the molar ratio of dCTP to DNA ends (Roychoudhury et al, 1976). Under typical reaction conditions of 500 u terminal transferase/ml and a ratio of dCTP:DNA ends of 200:1, 20-30 dC residues were added after incubation at 37°C for 10-15 min. The addition of dC residues continued for at least 30 min. Under identical conditions, the addition of dG residues had slowed considerably after 30 min, when approximately 25 dG residues/DNA end had been added, see figure 2.9.A. This observation is in accord with the results of others and is thought to be due to the fact that stretches of dG residues assume a secondary structure which cannot be "processed" by DNA polymerising enzymes (Gellert et al, 1962; Otsuka, 1981). However, it was observed that when certain batches of PstI digested pAT 153 were tailed using dGTP, the calculated number of residues/DNA end added appeared to be 100-200. This is probably due to some batches of DNA containing a large number of single-strand "nicks" induced during PstI
Figure 2.9

Tailing of Pst I digested pAT 153

2.9.A  
- dCTP  
- dGTP

2.9.B  
- 1 mM CoCl₂
- 4 mM MgCl₂

Time (min)
digestion or by freeze-thawing of the DNA. These internal nicks may be
tailed by the enzyme in the presence of CoCl₂ (Roychoudhury et al., 1976).
This was clearly an undesirable event when tailing a vector molecule for
cloning purposes. For this reason, PstI digested pAT 153 was tailed using
dGTP in a buffer containing 4 mM MgCl₂ in place of 1 mM CoCl₂. Under these
conditions, internal nicks in the DNA are less favourable substrates for
the enzyme (Deng and Wu, 1981). Slightly less dG residues/DNA end are
added under these conditions (Deng and Wu, 1981), but internal nicks are
not tailed (Roychoudhury et al., 1976), see figure 2.9.B. Calculations
showed an estimated 15-20 dG residues/DNA end were added to the vector
after 30 min incubation at 37°C. This number of residues has been shown to
be efficient for dC·dG hybridisation (Nelson and Brutlag, 1979). The
terminal transferase reaction was stopped by the addition of EDTA to 20 mM
and a single extraction with phenol/chloroform/isoamylalcohol/8-hydroxy-
quinoline.

The tailed vector was purified by gel filtration through Sephadex
G100, ethanol precipitation and washing in 70% (v/v) ethanol, then
redissolved in sterile distilled deionised water, divided into aliquots
and stored frozen at -20°C. Aliquots were thawed on ice when required and
to avoid damage to the DNA were not refrozen. A single batch of this
material was used to obtain complete cDNA clones of the genomes of
P3 Leon/12 a-b #411, P3 Leon/37 #960, P3/119 #643, P1/LS-c, 2ab and
entrovirus 70 (Dr. G. Stanway, personal communication). Recently, a method
has been published which claims to improve greatly the cloning efficiency
obtained using tailed plasmid vector (Leriche et al., 1983). This relies on
the selection of tailed, linearised vector DNA by hybridisation to an
oligo(dC)-cellulose column, thus reducing the background of transformants
resulting from uncut vector DNA. However, this technique was not tested.

Tailing of RNA·cDNA hybrids - Tailing of RNA·cDNA hybrid molecules under
the conditions used above for the vector (i.e. 4 mM MgCl₂), resulted in
the addition of very few (10 or fewer) dC residues/estimated 3' end. Molecules with 3' recessed or flush ends are poor substrates for terminal transferase except in the presence of CoCl$_2$ (Deng and Wu, 1981). It was therefore necessary to tail hybrid molecules in a buffer containing 1 mM CoCl$_2$. It was not possible to calculate the exact number of dC residues added/DNA end for two reasons. First, the precise concentration of DNA ends present was not known, due to uncertainty about the exact molecular weight of the cDNA. Secondly, unlike the tailing of double-stranded DNA, where each molecule has two ends which act as equal substrates for the enzyme, RNA/cDNA hybrid molecules have one DNA 3' end and one RNA 3' end, for which terminal transferase might have different affinities. It had previously been shown that terminal transferase will polymerise rNTP residues onto a 3' DNA end in the presence of CoCl$_2$ (Roychoudhury et al., 1976), but it was not known whether dNTP residues were added to a 3' RNA end under the same conditions. $^{32}$P-radiolabelled RNA/cDNA hybrids were tailed with [5-$^3$H]dCTP, or more commonly with [$^3$H-$^{32}$P]dCTP under the conditions described above. The reaction was monitored by the increase in acid-precipitable radiolabel, which was used to estimate the addition of 15-20 dC residues/end, (based on a visual estimate of the mean size of the cDNA, obtained by electrophoresis of a small aliquot on a 1% alkaline agarose gel, assuming that both ends of the hybrid molecules were tailed equally). No biochemical tests were carried out in order to determine whether the two different ends of the hybrid molecules were tailed differently, but the transformation results obtained using hybrid molecules tailed in this way (section 2.F), suggested that both ends were efficiently tailed. This was later confirmed by nucleotide sequencing of the ends of the recombinant molecules (see chapter 4). The addition of this number of dC residues usually took 15-30 min. Tailed hybrid molecules were purified as described above.

While this work was being performed, the results of similar experiments were published (van der Werf et al., 1981). This group cloned
RNA-cDNA hybrids of the poliovirus strain P1/Mahoney in the plasmid vector pBR 322 (Bolivar et al., 1977), by homopolymer tailing and annealing the tailed hybrid molecules to complementary tailed PstI cleaved vector. For reasons which are not clear, this report describes the "trimming" of the RNA-cDNA hybrid molecules with a mixture of RNase A and RNase T1 before tailing with terminal transferase. Since the overall cloning efficiency obtained by these authors was less than that obtained for P3/Leon 12 a,b #411 hybrids cloned without trimming, it seems likely that trimming the RNA-cDNA hybrids with nucleases is at best unnecessary and may reduce their transformation efficiency.

2.2.2 - Annealing of Homopolymer Tailed Molecules

PstI cleaved plasmid and RNA-cDNA hybrids bearing complementary homopolymer tails, purified as described above, were mixed together at a ratio of one vector molecule to one RNA-cDNA hybrid molecule and co-precipitated with ethanol. The ethanol precipitated mixture was redissolved in 100 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.2 mM EDTA at a final concentration of 1 µg pAT 153/ml, in a sealed, sterile 1.5 ml Eppendorf tube (Roewekamp and Firtel, 1980). The ends of the molecules were partially denatured by heating to 65°C for 10 min. Annealing occurred during incubation for 2 h at 45°C, followed by slow cooling to room temperature in a large covered waterbath (24-48 h). Usually, the annealed mixtures were incubated for a further 24-48 h on ice as this was found to slightly increase the potential transformation efficiency of the recombinant molecules. Annealed mixtures were stored on ice or at 4°C but to avoid possible freeze-thaw damage, were never frozen.
2.F TRANSFORMATION OF E. COLI

2.F.1 - Preparation of Competent Cells

Preparations of E. coli competent for transformation with DNA were prepared as described in chapter 6 and transformed with the annealed vector-RNA-cDNA hybrid mixtures described above. The preparation of highly competent bacterial cells was found to be one of the most crucial factors to the efficiency of the RNA-cDNA hybrid cloning method, especially since RNA-cDNA hybrids seem to be slightly less efficient at transforming E. coli than double-stranded cDNA's (see below). The efficiency with which a given strain of E. coli can be transformed with DNA is directly related to the rate at which the cells have been grown (Jones et al., 1981). Thus cells grown in Luria broth are more highly competent than cells grown under slightly less favourable conditions, e.g. in nutrient broth (Oxoid Ltd). Bacteria were made competent by a modification of the original method (Mandel and Higa, 1970; Cohen et al., 1972), then held on ice for 1–6 h before use. After this time, the transformation efficiency was found to decrease, despite a published report to the contrary (Dagert and Ehrlich, 1979). Similarly, the preservation of competent cells by freezing was found to lead to a severe loss of transformation efficiency, despite a report describing this technique (Morrison, 1977), so all competent cells were freshly prepared on the day they were used. After transformation, the mixtures were incubated in ten times their original volume of Luria broth for 1 h at 37°C in order to allow phenotypic expression of the newly acquired DNA.

2.F.2 - Selection of Transformed Cells

Cells transformed to a tetracycline resistant phenotype (tet^R) were selected by plating on Luria agar plates containing tetracycline and
incubated for 18-24 h at 37°C. It was found that the exact concentration of tetracycline used had some influence on the apparent transformation efficiency observed. The highest apparent efficiencies were obtained using low concentrations of tetracycline, e.g. 5 µg/ml, but at this low level, a very high background of non-transformed cells also grew up, making true transformed colonies difficult to detect. At higher concentrations of tetracycline (up to 50 µg/ml) progressively fewer colonies were seen, but on closer examination of the agar plates, very small micro-colonies were visible. On further incubation (e.g. 48 h), these grew up into full-sized colonies. It was subsequently discovered that these micro-colonies were genuine transformants and contained recombinant plasmids similar to those isolated from the large, rapidly growing colonies. The explanation for this observation is not clear. It was therefore necessary to use a concentration of tetracycline which resulted in the highest possible number of transformed colonies, without permitting the growth of a dense background of untransformed bacteria. Luria agar plates containing 10 µg/ml tetracycline (L-tet plates) were therefore routinely used.

2. F. 3 - Transformation Efficiency of E. coli

Transformation efficiency varies considerably from one strain of E. coli to another, but the best results achieved using the bacterial strains and methods described in chapter 6 were 5 x 10^6 transformants/µg of supercoiled pAT 153 and 5 x 10^2/µg PstI digested pAT 153. Two major factors were taken into account when deciding which strains of E. coli to transform with poliovirus cDNA's. Firstly, the biological containment level of the strain in accord with the Genetic Manipulation Advisory Group (GMAG) requirements and secondly, the transformation efficiency of the strain. E. coli MRC8 (reca, F-, hadR-, dap101/103, glmS33, metB1, upp1, supE), (derived from E. coli K12 by Dr. S. Brenner, MRC LMB, Cambridge), was used in initial attempts to clone double-stranded cDNA from
This severely weakened containment strain has exacting growth requirements and grows relatively slowly even in Luria broth supplemented with 50 µg/ml diaminopimelic acid and 200 µg/ml N-acetylglucosamine. The maximum transformation efficiency achieved with this strain was $1 \times 10^5$ transformants/µg of supercoiled pAT 153.

Fortunately, relaxation in the GMAG guidelines while this work was being performed allowed the cDNA clones described in chapter 3 to be obtained using other, more robust strains of *E. coli*. The strains chosen were *E. coli* JA221 (hsdM<sup>+</sup>, hsdR<sup>-</sup>, lac<sup>Y</sup>, leuB<sup>6</sup>, trpE<sup>5</sup>, recA<sup>1/F</sup>) (Clarke and Carbon, 1978) and *E. coli* JA226 (rk<sup>-</sup>, mk<sup>+</sup>, lop<sup>11</sup>, thi<sup>-</sup>, leu<sup>-</sup>, strB, RecBC) (Dr. J. Windass, personal communication).

The transformation efficiencies of both of the above strains were examined to determine whether the different DNA recombination pathways they possessed would affect the efficiency of cloning RNA-cDNA hybrids, which probably require extensive repair in vivo after transformation. It is known that strains of *E. coli* bearing the recA mutation (e.g. JA221), contain elevated levels of exonuclease Υ, the recBC product (see, Benzinger, 1978). Strains bearing recBC mutations carry out DNA repair by the recA pathway and do not contain exonuclease Υ (e.g. JA226). It was thought that this might influence the intracellular events following transformation with an RNA-cDNA hybrid. However, in spite of the different genotypes of these two strains, the cDNA clones prepared from each were indistinguishable. The only difference observed between the strains was that JA226 grew slightly more slowly in Luria broth than JA221 and showed a slightly lower transformation efficiency ($7 \times 10^5$ transformants/µg supercoiled pAT 153 c.f. $1-5 \times 10^6$ /µg from JA221). For this reason, the great majority of the cDNA clones described in chapter 3 were isolated from JA221, (as were all the cDNA clones of P3/Leon/37 #960 and P3/119 #643 prepared by Dr. G. Stanway). This finding agrees with the results obtained by Humphreys *et al* (1979), who found that intracellular nucleases did not affect transformability with plasmid DNA. Criticisms that RNA-cDNA
hybrid cloning results in "extensive deletions in the cDNA inserts" (Okayama and Berg, 1982) are probably due to nuclease contamination of enzyme preparations or the use of recombination-proficient strains of *E. coli* as hosts for the cloned cDNA, e.g. C600 (Wood and Lee, 1976), X1776 (Zain *et al.*, 1979), since failure to use such recombination-deficient host strains (most commonly strains of *E. coli* bearing the recA mutation) can also result in deletions when double-stranded cDNA is cloned. Cloned cDNA containing internal repeats is also subject to deletions by a process independent of the recA gene product (Morrow, 1979). Under suitable conditions, the cloned cDNA does not show deletions, insertions or other re-arrangements (Cann *et al.*, 1983).

The transformation efficiency obtained from RNA-cDNA hybrids was $5 \times 10^3$ transformed colonies/µg of vector DNA transformed, i.e. $5 \times 10^3$/ml of the annealed mixture. This value was not substantially less than the transformation efficiency observed with double-stranded cDNA clones prepared from the same batch of cDNA, approximately $5 \times 10^3 - 1 \times 10^4$/µg of vector DNA (Dr. G. Stanway, personal communication). This can be compared with a background figure of $1 \times 10^3$ transformants/µg of vector DNA mock annealed in the absence of cDNA. These figures represent overall cloning efficiencies of approximately $5 \times 10^3$ transformants/pmol of RNA for RNA-cDNA hybrids and approximately $8 \times 10^3$ transformants/pmol for double-stranded cDNA. The overall cloning efficiency of the RNA-cDNA hybrid method as described in this chapter is superior to previously published figures for the method (Wood and Lee, 1976; Zain *et al.*, 1979; van der Werf *et al.*, 1981) and compares favourably in terms of overall efficiency and convenience with the much more commonly used double-stranded cDNA cloning. Overlapping cDNA clones together spanning the entire genomes of P3/Leon/37 460, P3/119 643, P1/LS-c 2ab and enterovirus 70 have also been obtained by cloning RNA-cDNA hybrids under the same conditions described here (Dr. G. Stanway and M. Ryan, personal communication).
2. G CONCLUSIONS

The overall conclusion which emerges from these experiments is that maximal synthesis of full-length cDNA is achieved by a careful balance of conditions, which are not identical to the conditions under which the maximum yield of cDNA is obtained. 100% cDNA yields are apparently not possible using the type of experimental system described above. The maximum yields obtained were of the order of 40% of the RNA template input. Much of the published literature on cDNA synthesis is concerned primarily with the synthesis of short cDNA's, (e.g. less than 2 kb). The synthesis of full-length cDNA from a poliovirus RNA template is not only considerably more difficult, but was of great importance to the success of later stages of the project. For these reasons, it was necessary to thoroughly examine and adjust the conditions used for reverse transcription so as to achieve the optimum balance between maximum yield of cDNA and maximum length of product. The proportion of full-length cDNA synthesised was not determined, but gel analysis of the reverse transcripts showed that high molecular weight cDNA of comparable size to the poliovirus genome was synthesised. In these circumstances, it was decided to use RNA-cDNA hybrids synthesised under the above conditions for molecular cloning and to further analyse the reverse transcription products by characterisation of the clones obtained. The results of these experiments are described in chapter 3.
CHAPTER 3

CHARACTERISATION OF cDNA CLONES
In any project involving cDNA cloning, the isolation and characterisation of the recombinant plasmids is necessary in order to isolate clones containing large cDNA inserts from a mixture of native plasmids and plasmids containing very small cDNA inserts. Characterisation of the transformed bacterial colonies tends to be a laborious process. However, analytical techniques have been developed to examine large numbers of recombinant colonies rapidly and economically. Thus it is now possible to characterise recombinant molecules from many hundreds or even thousands of transformed colonies in a relatively short period. The efficient application of colony screening and selection techniques was essential to the overall success of this project.

This chapter describes the selection and characterisation of recombinant cDNA clones of P3/Leon 12 a,b 411. The initial screening of recombinant colonies is performed by phenotypic examination of the transformed bacterial colonies, in order to identify colonies which contain a DNA insert in the PstI site of the pAT 153 vector. This can be accomplished by making use of the fact that such recombinant molecules confer on the host cell a tetracycline resistant, ampicillin sensitive (tet\(^r\) amp\(^s\)) phenotype, due to the insertional inactivation of the ampicillin resistance gene of the vector molecule. After this initial screening process, the clones may be examined by means of colony hybridisation experiments, using several different radiolabelled hybridisation probes. Further analysis of the size and relative genomic locations of the cDNA inserts of recombinant molecules can then be obtained by restriction endonuclease mapping. This chapter describes the use of these methods to produce a restriction endonuclease map of overlapping cDNA clones together corresponding to the entire genome of the virus. Apart from its intrinsic interest, this map was an essential prerequisite for the later nucleotide sequencing studies (chapter 4).
3. B RECOMBINANT COLONY SCREENING METHODS

3. B. 1 - Phenotypic Selection and Screening

The initial selection of transformed colonies was performed during the transformation procedure, when the transformed bacteria were plated out onto L-tet plates in order to select cells transformed by plasmid DNA. This procedure only selects cells which have been transformed and now contain a plasmid and is not specific for cells which have been transformed by recombinant plasmids. Since it was desirable to limit the number of colonies examined by hybridisation, it was considered important to ensure that all the colonies selected contained a recombinant rather than a native plasmid. Although approximately 80% of the transformed cells contained a recombinant plasmid (see below) a further phenotypic selection step was performed in order to exclude the 20% background level of native plasmids. This was achieved by testing the tetracycline resistant (\text{tet}^R) colonies for their ability to grow on L-amp plates. After incubation for 18-24 h at 37°C, recombinant colonies were identified by their sensitivity to ampicillin (\text{amp}^S). The colonies were also streaked out on L-tet master plates using sterile wooden toothpicks to ensure definite identification during subsequent analysis. Clones with a \text{tet}^R \text{amp}^S phenotype were taken from the L-tet master plates and further examined by hybridisation (see section 3. B. 2). The background level of \text{tet}^R \text{amp}^R resistant transformants (the average figure for several cloning experiments using the same batch of dGTP tailed pAT 153 was approximately 20% of the total number of transformed colonies) was probably due to a very small amount of vector DNA which escaped PstI digestion and subsequent homopolymer tailing. Since supercoiled plasmid DNA is much more efficient at transforming competent cells than the linearised, tailed and annealed vector, only a very small amount of uncut DNA, much less than could be detected on EtBr-stained agarose gels, would need to be present to give the observed transformation
background of 20% amp\textsuperscript{R} colonies.

About 1000 colonies transformed with RNA-cDNA hybrid annealed mixtures were examined in this way (and at least 1000 colonies transformed with double-stranded cDNA annealed mixtures, Dr. G. Stanway, personal communication). This process of streaking out was time consuming and laborious. Recently, a much more rapid and convenient indicator-dye method of testing bacterial colonies for ampicillin sensitivity has become commercially available ("Ampscreen Disks", BRL Ltd). This method promises to make the process of phenotypic analysis less laborious and very much more rapid. However, an even better alternative would be to reduce the background of colonies transformed with native vector molecules by purification of the dGTP tailed vector preparation. A recently published method has demonstrated the hybridisation of dGTP tailed vector preparations to oligo(dC)-cellulose, thus removing uncut and poorly tailed molecules and claims to give less than 1% of background transformants (Leriche et al, 1983). This level of efficiency would eliminate the need for any phenotypic testing and greatly increase the efficiency with which the initial screening of recombinant clones could be carried out.

3.8.2 - Colony Hybridisation

Recombinant clones, selected by virtue of their tet\textsuperscript{R} amp\textsuperscript{S} phenotype, were further examined for the presence of poliovirus cDNA by colony hybridisation experiments. This method of analysis also provided some information on the approximate size and genomic location of the cDNA inserts. The recombinant colonies from the L-tet master plates were streaked in grid formation onto the surface of nitrocellulose membrane filters resting on L-tet plates. To ensure accurate subsequent identification of the colonies, a second L-tet master plate was usually made. The plates were incubated in an inverted position for 6-12 h at 37\textdegree C. Very little material was necessary to obtain a strong hybridisation
response (Grunstein and Hogness, 1975). It was found that colonies incubated for longer than 6-12 h tended to become too large, especially in the case of E. coli JA221, which produced rapidly growing, mucoid colonies. Very large colonies did not lyse fully, sometimes resulting in a "halo" effect which made interpretation of the results difficult. Large colonies also tended to spread, merging into one another, thus decreasing the definition of the method. A total of 773 RNA-cDNA hybrid clones (and 836 double-stranded cDNA clones, Dr. G. Stanway, personal communication) were examined with at least one of the four different hybridisation probes described below.

1) Hybridisation to high molecular weight cDNA - Initial hybridisation experiments, performed as described in chapter 6, used $^{32}$P-radiolabelled cDNA of low specific activity (0.1-0.05 $\mu$Ci/pg cDNA), synthesised in a standard analytical scale reverse transcription reaction primed with p(dT)$_{17}$ (see chapter 2). To obtain unambiguous colony hybridisation results, it was essential to remove low molecular weight radiolabelled material and un-incorporated radioactivity from the hybridisation probe. This was achieved by gel filtration through Sephadex G100. For accurate interpretation of the results obtained, it was found to be important that both negative and, whenever possible, positive control colonies were included in the experiment. The controls used were E. coli JA221 containing pAT 153 as a negative control and a previously characterised clone as a positive control. (In initial hybridisation experiments, vRNA or unlabelled cDNA were used as positive controls.) Under the hybridisation conditions used, 95-99% of the recombinant colonies hybridised to this high molecular weight cDNA probe. A few of the colonies which did not hybridise or hybridised only very weakly were tested by other means and found to contain DNA with the same restriction endonuclease map as the vector, or much more commonly, very small inserts less than 50 bp), which were presumably too short to stably hybridise
under the conditions used.

Inclusion of dextran sulphate (a high molecular weight anionic polymer) in such experiments has been reported to increase the hybridisation response by more than 100-fold (Wahl et al., 1979; Jeffreys et al., 1980). However, when various concentrations of dextran sulphate were included during these colony hybridisation experiments, no increase in the sensitivity or strength of the hybridisation response was apparent. This finding probably reflects the low-stringency hybridisation conditions used (3 x SSC, 0.1% (w/v) SDS; see, Grunstein and Wallis, 1979; Jeffreys et al., 1980), or the high degree of homology between the hybridisation probe and the cDNA clones. As a result of this finding, dextran sulphate was not routinely used during colony hybridisation experiments.

Recombinant colonies hybridised equally well to cDNA synthesised from P3/Leon 12 a\(_1\) b #411 (the strain from which they themselves were derived), or from P3/Leon/37 #960 or P3/119 #643. This finding was not surprising in view of the extent of the similarity known to exist between the three strains (Minor, 1982) and with regard to the low-stringency hybridisation and washing conditions used.

In addition to results provided by this hybridisation probe, more detailed analysis of the cDNA content of the clones was achieved by further colony hybridisations performed under the same conditions described above but using three additional hybridisation probes:

11) 3' enriched hybridisation probe - In order to obtain some initial information on the genomic location of the clones and to orientate the restriction endonuclease map derived from them, a 3' enriched hybridisation probe was used. This was synthesised by reverse transcription of P3/Leon 12 a\(_1\) b #411 RNA in a reverse transcription reaction primed with p(dT)\(_{17}\), similar to the standard reaction conditions except that it contained 200-500 μCi/ml [α-\(^{32}\)P]dATP with a specific activity of 2000-3000 Ci/mmol (Cann et al., 1983). Because reverse
transcriptase is unable to synthesise long cDNA's extending to the 5' regions of the genome in reactions containing sub-optimal concentrations of dATP (the rate limiting reagent in the reaction - see chapter 2), a hybridisation probe synthesised under these conditions was enriched for short cDNA's corresponding to the 3' region of the genome. The specific activity of cDNA synthesised in this type of reaction was high, approximately 5 μCi/μg cDNA. In contrast to results obtained from the high molecular weight cDNA probe, only 25% colonies examined (approximately 200 RNA-cDNA hybrid clones) hybridised to this 3' enriched probe. pSGA 31, the clone containing the 3' terminus of the genome (section 3.D.1) was identified by its positive hybridisation response to this probe.

iii) Randomly primed hybridisation probe - After hybridisation and autoradiography of the nitrocellulose filters, the bound hybridisation probe was removed by washing the nitrocellulose filters in 0.1 x SSC, 0.1% (w/v) SDS at 80°C (3 x 20 ml changes over a 3 h period). This process was found to remove all detectable radioactivity from the filters, which were then re-hybridised to a second, randomly primed cDNA probe. P3/Leon 12 a, b # 411 vRNA was reverse transcribed in a standard analytical scale reaction mixture which contained random, low molecular weight salmon sperm DNA oligonucleotides (Taylor et al., 1976) in place of p(dT)17 to prime cDNA synthesis. cDNA synthesised in this reaction is believed to be primed throughout the genome by fortuitous hydrogen bonding of the oligonucleotides and thus presumably represents an even distribution of material from the entire genome. In the case of the dATP-starved hybridisation probe described above, or even cDNA synthesised under standard conditions, the 5' regions of the genome tended to be under-represented. The specific activity of the cDNA synthesised in this randomly primed reaction was much lower than that of the 3' enriched probe, approximately 0.25 μCi/μg cDNA.

The autoradiographs of the two hybridisation experiments performed on
the same filters were then compared. Figure 3.1 shows one such pair of autoradiographs, 3.1.A being the result of hybridisation to the high specific activity 3' enriched probe and 3.1.B the result of hybridisation of the same filter, after washing, to the randomly primed probe. The results obtained using the two hybridisation probes are clearly different. Although 95% of the colonies tested (approximately 200 RNA-cDNA hybrid clones) hybridised to the randomly primed probe to some extent, the strength of the response varied considerably. Clones which hybridised strongly to the randomly primed probe, when further examined by restriction endonuclease digestions (see next section), were found to contain relatively larger cDNA inserts than weakly hybridising clones. This suggests that the oligonucleotides prime at random points along the genome and that the strength of the hybridisation response obtained using cDNA synthesised in this way as a probe could be used as an indication of the likely size of the cDNA insert, independent of its genomic location. In contrast, many of the clones which strongly hybridised to the high specific activity 3' enriched probe were found to contain very short cDNA inserts (e.g. less than 100 bp). For this reason, clones which hybridised strongly to both of the above probes were selected for further analysis. Thus the randomly primed probe and the 3' enriched probe were used in combination to select large clones likely to contain sequences from the 3' region of the poliovirus genome.

iv) Nick translated hybridisation probes - While the process of clone selection was continuing, a restriction endonuclease map of the P3/Leon 12a b genome was gradually being determined from overlapping clones identified in early experiments, orientated with respect to their hybridisation response to the 3' enriched probe. This was important, as no other data was then available on the nucleotide sequence of any poliovirus, except for very limited information totalling less than 50 nucleotides from the extreme 5' and 3' ends of the genome (Black et al,
Figure 3.1

Colony Hybridisation

Tetracycline resistant ampicillin sensitive transformants were selected as described in section 3.A and grown up in grid formation on a nitrocellulose membrane filter.

3.1.A - Autoradiograph of filter after hybridisation to the high specific activity 3' enriched probe (see text).

3.1.B - Autoradiograph of the same filter after washing and hybridisation to the randomly primed probe (see text).
1978; Porter et al., 1978; Nomoto et al., 1981a). Without this information, orientation of the restriction endonuclease map would have been impossible. As the restriction endonuclease map was gradually extended towards the 5' end of the genome, only clones which hybridised strongly to the randomly primed probe and not to the 3' enriched probe, i.e. those likely to contain large cDNA inserts not including the sequence at the 3' end of the genome, were further analysed. However, when the restriction map extended approximately 5 kb from the 3' end of the genome, it was observed that clones which extended further in the 5' direction were present at much lower frequency. This was probably due to the size distribution of the cDNA product synthesised in the reverse transcription reaction, in which p(dT)$_{17}$ was used to prime cDNA synthesis from the 3' end of the genome. Long cDNAs containing material corresponding to the 5' end of the virus genome were less abundant than shorter species (see chapter 2).

For this reason, a fourth type of hybridisation probe was prepared to identify clones from the 5' region of the genome. These probes consisted of restriction endonuclease fragments from previously characterised clones containing the 5'-most end of the established restriction map. DNA fragments were isolated from agarose gels and radiolabelled to high specific activity (1-2 μCi/μg DNA) by nick translation (Rigby et al., 1977). The restriction endonuclease fragments used for this purpose were the 5' terminal PstI fragment of pSAG 16 (1 kbp) which resulted in the isolation of clones pSAG 21 and 22 (see figure 3.6) and the whole cDNA insert of pSAG 22 (1.25 kbp, wholly contained within pSAG 21), which resulted in the isolation of pSAG 23 (not shown in figure 3.6) and pSAG 24, which contains the 5' terminus of the P3/Leon 12 a$_1$b #411 genome. By the sequential use of these nick translated fragments, as well as the other types of hybridisation probe described above, it was possible to isolate overlapping RNA-cDNA hybrid and double-stranded cDNA clones together corresponding to the entire genome of P3/Leon 12 a$_1$b #411. This
fact was subsequently confirmed by nucleotide sequencing (chapter 4).

3.B.3 - Plasmid Isolations

Rapid isolation method 1 - Clones which showed favourable responses to the various hybridisation probes described above were grown up as 1 ml cultures in L-tet broth, incubated for 12-18 h at 37°C. Although the vector pAT 153 is a high copy number plasmid and therefore not likely to be subject to rapid loss from the host cells, a constant selection pressure was maintained on all bacteria containing recombinant plasmids by the inclusion of 10 µg/ml tetracycline in all media. Plasmid DNA from the 1 ml cultures was partially purified by the first rapid isolation method described in chapter 6 (a modification of the method of Holmes and Quigley, 1981). The major advantages of this method were its speed and simplicity, allowing up to 48 different clones to be processed in a single batch. On completion, the whole batch could then be examined by electrophoresis on a single, large agarose gel (20 cm x 20 cm x 0.5 cm thick). However, the quality of the plasmid DNA isolated by this method was found to vary quite widely. A low yield of DNA and the presence of large amounts of contaminating bacterial nucleic acids, (mostly low molecular weight ribosomal RNA plus some chromosomal DNA), often made it difficult to visualise the plasmids. Figure 3.2 shows restriction endonuclease digests of 12 rapid isolates prepared by this method. The enzymes used were (left to right for each of the 12 isolates) EcoRI, HindIII and PstI and the digestions subjected to electrophoresis on an EtBr-stained agarose gel. Unfortunately, this method of DNA isolation, although rapid and convenient, did not reliably give such high quality DNA preparations as those shown in figure 3.2. Some contaminating substance, possibly protein or RNA, not removed from the DNA during this procedure, often made it impossible to digest the DNA isolated using this method with restriction endonucleases. Frequently, in spite of using excess quantities
DNA from 12 rapid plasmid isolates prepared by the first method described in section 3.B.3 cut with restriction endonucleases EcoRI, HindIII and PstI (left to right). Note the low yield of plasmid DNA and contaminating bacterial nucleic acids.
of enzymes, partial digestions were obtained which were difficult to interpret. In addition, the presence of large quantities of RNA often obscured the presence of low molecular weight restriction endonuclease fragments in the gel. For these reasons, this method was used primarily for rapid and convenient initial screening, in which large numbers of recombinant plasmids could be examined. A total of 114 RNA-cDNA hybrid clones (and 205 double-stranded cDNA clones, Dr. G. Stanway, personal communication) were examined using this method. Migration in 0.5-1% (w/v) agarose gels of supercoiled plasmid DNA was compared with that of supercoiled plasmids of known sizes. It was therefore possible to make an estimate of the size of the cDNA inserts in the recombinant plasmids. Those with the largest inserts were then selected for more detailed restriction endonuclease analysis. Alternatively, plasmid DNA prepared by this method was digested with a specific restriction endonuclease to attempt to identify clones from a particular location on the genome. In spite of its limitations, this method proved to be valuable for the initial screening of large numbers of recombinant plasmids, previously identified by colony hybridisation. However, it proved necessary to resort to a second rapid plasmid isolation method to provide plasmid DNA of sufficient quality for accurate restriction endonuclease mapping.

Rapid isolation method 2 - A more precise estimation of the size of cDNA inserts in recombinant plasmids and a more detailed restriction map was obtained for plasmids containing large cDNA inserts (longer than 1 kbp) using plasmid DNA isolated by the second rapid method described in chapter 6. This second method allowed 5-15 μg of plasmid DNA to be isolated from an overnight culture grown in 10 ml L-tet broth. It was nearly always possible to digest this DNA, purified by RNase treatment and phenol extraction, with restriction endonucleases. A single preparation provided an ample quantity of DNA for a thorough restriction endonuclease analysis. Thus the three main problems of low yield, low purity and RNA
contamination experienced with the first method were overcome. This second method was used to examine the most promising cDNA clones identified by the previous technique. Approximately one third of the clones analysed using method one were subsequently examined using this second method. Figure 3.3 shows 18 plasmids purified in this way examined on an agarose gel after digestion with EcoRI and PstI. The much higher quality of the DNA preparation is apparent when this figure is compared with figure 3.2. This method was also found to give much higher quality DNA than a rapid alkaline extraction method (Birnboim and Doly, 1979).

Large scale isolation method - In early experiments, plasmids with large cDNA inserts (longer than 2 kbp) whose genomic locations had been identified by restriction endonuclease mapping, were grown on a larger scale in 0.5 or 1 l cultures in L-tet broth and highly purified DNA prepared by CsCl-EtdBr centrifugation (see chapter 6). DNA from 24 RNA-cDNA hybrid clones (pSAG 1-24) characterised in the above experiments was prepared using this method. DNA from 29 double-stranded cDNA clones (pSGA 1-28, pSGA 31) was also prepared in this way (Dr. G. Stanway, personal communication). This DNA was used for detailed restriction endonuclease mapping and for sub-cloning into M13 prior to nucleotide sequencing. This method overcame the problem of contamination with bacterial chromosomal DNA, which was initially thought likely to result in many M13 subclones containing E. coli DNA. As the work progressed however, plasmid DNA prepared by the second rapid isolation method described above was increasingly used for more applications. These included sub-cloning into M13, with which few problems were experienced. This was probably because the relatively small quantities of chromosomal DNA present in high quality rapid plasmid isolations were effectively diluted out with respect to any particular size of fragment, especially when DNA was prepared for sub-cloning by two agarose gel electrophoresis-elution steps. It was found that plasmid DNA purified by the second rapid method could also be used
Figure 3.3
Rapid Plasmid Isolates (Method 2)
DNA from 18 rapid plasmid isolates prepared by the second method described in section 3.B.3 cut with restriction endonucleases EcoRI (top row) and PstI (bottom row).
for Smith-Birnsteil restriction endonuclease mapping, (see section 3.C). The reconstruction of the genome of P3/Leon/37 #960, referred to in chapter 4, was done exclusively from this type of rapid isolate DNA, as was most of that of P3/119 #643 (Dr. G. Stanway, personal communication). The isolation of plasmid DNA by rapid methods was therefore an important contribution in terms of both time and economy to the success of the project as a whole.
3.C RESTRICTION ENDONUCLEASE MAPPING

3.C.1 Restriction Enzymes With Hexameric Recognition Sites

The genomic locations of recombinant plasmids with large cDNA inserts were identified by restriction endonuclease mapping. Plasmids were mapped with the restriction endonucleases BanHI, EcoRI, HindIII, PstI, SphI, XbaI and XhoI (see, Roberts, 1980). 0.5-2 μg of DNA was digested with 1 u/μg of restriction enzyme at 37°C for 1 h in the appropriate buffer (see chapter 6). The samples were then heated for 10 min at 65°C to terminate the reaction and an equal volume of Elfo loading buffer was added before loading directly into 0.5 cm slots in 1% (w/v) 10 cm x 10 cm x 0.5 cm thick agarose gels. The migration of fragments produced by the digestion were compared with the migration of fragments of λ⁺c DNA doubly digested with EcoRI and HindIII. This system gave a visible range from 5.2 to 0.56 kbp (the smaller and larger fragments not being resolved in 1% agarose gels). Linearised pAT 153 (3657 bp) was used as an additional size marker. The DNA was subjected to electrophoresis in gels containing Elfo buffer and 0.5 μg/ml EtdBr at 10 V/cm until the fragments were clearly separated, then placed on a UV transilluminator box and a photograph taken. The mobilities of the size marker fragments (distance migrated) were determined from the photograph of the gel and plotted on semi-log graph paper against the size of the fragment (in kbp, on the log axis). The sizes of the unknown fragments were determined from this graph. In some later experiments, a computer program rather than the graphical method was used to determine the sizes of the unknown fragments, but no significant increase in accuracy was noted, other factors such as the conditions of electrophoresis and accuracy of measurement from the photograph being much more important in the overall accuracy of the method. On average, cleavage positions of enzymes determined in this way were found to be accurate within 50 bp, as later confirmed by nucleotide
sequencing.

Restriction endonuclease fragments were identified using the known restriction endonuclease map of the pAT 153 vector (Sutcliffe, 1978; Twigg and Sherrat, 1980) and the location of restriction endonuclease cleavage sites in the cDNA insert determined by multiple enzyme digests when necessary. At this point, the PstI site reconstruction strategy using homopolymer tailing, as described in chapter 2, became useful. Of all the plasmids examined (see section 3.D), 95% possessed two PstI cleavage sites flanking the cDNA insert and all possessed at least one flanking PstI cleavage site. Thus by a single PstI digestion, an estimate of the size of the cDNA insert in a recombinant plasmid could be obtained. A linearised pAT 153 size marker was routinely used to allow unambiguous identification of the vector fragment.

Figure 3.4 shows a typical plasmid, pSAG 22 digested with restriction endonucleases. This plasmid was selected on the basis of positive hybridisation to a probe consisting of a nick translated restriction endonuclease fragment, the 1 kbp 5' terminal PstI fragment of plasmid pSAG 16 (see section 3.B.2; figure 3.6). The plasmid DNA was prepared by the second rapid isolate method described in section 3.B and contains some low molecular weight RNA plus a faint trace of E. coli chromosomal DNA (not visible in figure 3.4). Clear restriction endonuclease mapping results were obtained with this DNA. In the case of pSAG 22 (figure 3.4), single restriction endonuclease digestions indicated unique cleavage sites (known to be in the vector DNA) for the enzymes HindIII and SphI, giving a total size for the linearised plasmid of 4.9 kbp. This was confirmed by the PstI digestion, which produced two fragments, the vector (3.65 kbp) and the cDNA insert of 1.25 kbp. The EcoRI digestion indicated the presence of two cleavage sites, one known to be in the vector and one in the cDNA insert, resulting in fragments of 3.4 and 1.5 kbp. The BamHI digestion similarly produced two visible fragments of 3.0 and 1.7 kbp, implying the presence of a further 0.2 kbp BamHI fragment, not visible on this gel (3.4.A). The
Figure 3.4
Mapping of a recombinant plasmid (pSAG 22) by restriction endonucleases with hexameric recognition sites

3.4.A - Single digests  
3.4.B - Multiple digests  
3.4.C - Restriction endonuclease map of pSAG 22

pSAG 22

4.9kbp

Insert = 1.25kbp
orientation of the cDNA insert with respect to the vector molecule was determined by a triple EcoRI, HindIII, BamHI digestion of the plasmid (3.4.B), which produced two visible fragments of 2.82 and 1.29 kbp. This result indicated that the orientation of the cDNA insert is that shown in 3.4.C. The cleavage positions of other restriction endonucleases with hexameric recognition sites was determined in a similar way to that described for pSAG 22.

3.C.2 Restriction Enzymes With Tetrameric Recognition Sites

The cleavage positions of restriction enzymes with tetrameric recognition sites could not be mapped accurately using the technique described above. This was because of the small average size of the restriction fragments produced and thus the complexity of the pattern. For this reason, cleavage positions of the restriction endonucleases AluI, HaeIII and Sau3A were mapped using the method of Smith and Birnstiel (1976). The method was used as described by the authors with only slight modifications to the reaction conditions and in the method of end-labelling the DNA (the fill-in reaction of the Klenow fragment of E. coli DNA polymerase I was used in place of T4 polynucleotide kinase).

Partial digests were performed on the asymmetrically radiolabelled DNA to be mapped, by means of adjusting the concentrations of restriction endonuclease and unlabelled carrier DNA in the reaction mixture. This results in a complete "ladder" of partial digestion products, the difference in size from one fragment to the next being equivalent to the size of the contiguous restriction endonuclease fragments. The method also results in an effective increase in the size of the often very small restriction endonuclease fragments being mapped to one more accurately resolvable in agarose gels. Since the DNA is radiolabelled at one end only (see chapter 6) and the restriction endonuclease map of the vector is known, the cleavage positions in the cDNA insert for a given enzyme can be
determined. Typically, 0.5 μ of restriction endonuclease was used in a 20 μl reaction mixture containing 0.5 μg of unlabelled high molecular weight calf thymus carrier DNA, plus 0.5-1 μg of asymmetrically labelled plasmid DNA, with a specific activity of 0.1-0.5 μCi/μg. In order to ensure a sample containing only a partial digestion of the labelled DNA was obtained, aliquots of the digestion mixture were taken at time points from 1-30 min. The endonuclease reaction was terminated by placing the sample on ice, immediately followed by phenol extraction and the addition of an equal volume of Elfo loading buffer. Electrophoresis was performed at 5 V/cm in 1% (w/v) 20 cm x 20 cm x 0.5 cm thick agarose gels containing Elfo buffer until the bromophenol blue dye marker had migrated to 2-3 cm from the bottom of the gel. The gel was then dried onto a glass plate and subjected to autoradiography. Using this gel system with a radiolabelled EcoRI and HindIII double digest of λ+c DNA as a size marker (plus the vector derived restriction fragments as low molecular weight size markers), the method was accurate to within approximately 25 bp up to a maximum cDNA insert size of about 2.5 kbp and usable with less accuracy up to a total insert size of 5 kbp. The method was very convenient and a single experiment determined unequivocally the cleavage positions of the enzyme being mapped.

Figure 3.5 shows a contact print of an autoradiograph of a typical restriction endonuclease mapping experiment, showing two plasmids (pSAG 21 and pSAG 22) digested with Sau3A. Sizes of EcoRI and HindIII digested λ+c DNA fragments are shown on the right and vector and insert derived fragments (of pSAG 21) indicated on the left. pSAG 22 has a cDNA insert of 1.25 kbp, which maps completely within that of pSAG 21 (1.7 kbp), as can be seen from the pattern of Sau3A fragments produced. In addition to fixing accurately the location of the cDNA insert relative to the existing restriction map of the virus genome, this method provided a refinement of the size estimate of the cDNA insert previously obtained by PstI excision of the cDNA, allowing very accurate measurement of insert size.
Figure 3.5

Restriction Endonuclease Mapping of Recombinant Plasmids

Cleavage positions of restriction endonucleases with tetrameric recognition sites were determined by electrophoresis of the products of partial restriction digests of asymmetrically labelled plasmids on agarose gels. (Direct contact print of autoradiograph.)
3.C.3 Size and Genomic Distribution of cDNA Clones

The restriction endonuclease mapping data from overlapping cDNA clones was used to build up a restriction endonuclease map corresponding to the entire genome of P3/Leon 12 a,b #411. This restriction endonuclease map was approximately 7.4 kbp in length and was orientated with respect to the 5' and 3' ends of the vRNA by the location of clones which hybridised strongly to the 3' enriched hybridisation probe. This orientation was later confirmed by the nucleotide sequencing experiments described in chapter 4.

Figure 3.6 shows the size and genomic location of representative clones made by the RNA-cDNA hybrid method (Cann et al, 1983; chapter 2). These clones are indicated by the notation pSAG. In order to compare the hybrid cloning method with the classical double-stranded cDNA cloning method, a parallel series of experiments was performed on an aliquot of the same RNA-cDNA hybrid material (Dr. G. Stanway, personal communication). The size and genomic location of these clones (indicated by the notation pSGA) is also shown in figure 3.6. pSAG clones marked with an asterisk were isolated from a second, separate RNA-cDNA cloning experiment. The position of the clone number in figure 3.6 indicates the end of the cDNA insert nearest the EcoRI site of the vector molecule (with the exception of pSAG 24, where this was not determined). Also shown is the restriction endonuclease map of cDNA corresponding to the total genome of P3/Leon 12 a,b #411.

The largest cDNA clones obtained using both cloning methods are shown in figure 3.6. These are, for the RNA-cDNA hybrid method, pSAG 16 and pSAG 20, both with a cDNA insert of 3.8 kbp and for the double-stranded cDNA method, pSGA 21 and pSGA 19, 2.05 and 1.65 kbp respectively. The relative distribution of cDNA clones made by the two methods is also of interest. One double-stranded cDNA clone, pSGA 31 (with a cDNA insert of
Figure 3.6

CDNA insert size and genomic location of P3/Leon 12a1b #411 clones

Restriction endonuclease map of genomic cDNA.

RNA-cDNA hybrid clones

Double-stranded cDNA clones
0.8 kbp), was isolated which corresponded to the extreme 3' end of the virus genome, a fact which was later confirmed by nucleotide sequencing. No RNA−cDNA hybrid clones were isolated which contained the extreme 3' 175 bp of the genome, but because pSGA 31 had been previously isolated, no intense effort was made to identify such a clone. However, no clones made by the double-stranded cDNA cloning method were isolated which extended further than 4.9 kbp (pSGA 28) from the 3' terminus. In contrast, overlapping cDNA clones made by the RNA−cDNA hybrid method were isolated which corresponded to almost the entire genomic cDNA (minus the 175 bp at the 3' terminus) of P3/Leon 12 a,b #411, including the extreme 5' terminus of the genome. This fact was later confirmed by nucleotide sequencing. (In subsequent RNA−cDNA hybrid cloning experiments with P3/Leon/37 #960 and P3/119 #643, overlapping cDNA clones corresponding to the entire virus genome, including the extreme 5' and 3' termini, were isolated from a single experiment. These clones contained cDNA inserts of up to 5.2 kpb, Dr. G. Stanway, personal communication.)

In order to isolate the clones shown in figure 3.6, plus many other smaller clones not shown, a total of 836 recombinant plasmids made by double-stranded cDNA cloning were examined by colony hybridisation. 205 of these further examined by gel analysis of rapid plasmid preparations and restriction endonuclease mapping (Dr. G. Stanway, personal communication). In comparison, 773 colonies were examined by hybridisation and 114 further examined to isolate the RNA−cDNA hybrid clones shown. The contrast between the RNA−cDNA hybrid and double-stranded cDNA cloning methods is apparent from these results. Not only was the maximum cDNA insert size smaller for double-stranded cDNA clones than for RNA−cDNA hybrid clones made in the same reverse transcription mixture, but the frequency with which clones corresponding to the 5' half of the virus genome could be isolated was much lower for double-stranded cDNA than for RNA−cDNA hybrid clones. In addition, the design of the double-stranded cDNA cloning method precludes the possibility of cloning the extreme 5' terminus of the genome, because
of the "hairpin" priming of second-strand cDNA synthesis and the subsequent S1 nuclease treatment. RNA-cDNA hybrids may be cloned by the direct terminal transferase tailing of the first-strand reverse transcription product, without the need for further manipulative steps such as alkaline hydrolysis of the RNA template, second-strand synthesis and single-strand specific nuclease treatment before tailing. These further manipulations are time-consuming and result in losses of material at each step, an important consideration when the amount of RNA starting material is limited. Apart from its speed and simplicity, the ability to obtain clones corresponding to the entire virus genome in a single experiment without the need for further manipulations to clone the 5' terminus (such as those carried out by Land et al., 1981; Racaniello and Baltimore, 1981a; Nomoto et al., 1982a; Okayama and Berg, 1982) makes the RNA-cDNA hybrid cloning method more attractive for the cloning of RNA molecules such as picornavirus genomes than the "classical" double-stranded cDNA cloning method normally used.

3.C.4 Restriction Endonuclease Map of the Total Genomic cDNA

Figure 3.6 shows a restriction endonuclease map for the enzymes AluI, BamHI, EcoRI, HaeIII, HindIII, PstI, Sau3A, Sphi, XbaI and XhoI of the total genomic cDNA synthesised from P3/Leon 12 a1b #411. The restriction endonuclease map was built up from overlapping cDNA clones which together represented a cDNA of approximately 7.4 kbp. Both the total length of the genomic cDNA and the accuracy of individual restriction endonuclease cleavage sites were later confirmed by nucleotide sequencing (see chapter 4). The restriction endonuclease sites show an apparently random (though not regular) distribution, broadly complying with the expected frequency of one hexanucleotide target every 4² bp, (i.e. 4096 bp, 1-2 sites/7.4 kbp) and one tetranucleotide target every 4⁴ bp, (i.e. 256 bp, 28-29 sites /7.4 kbp).
Apart from the intrinsic interest of such a restriction endonuclease map, three other important points arise from it. The first is this map was used to determine the strategy used for subsequent sub-cloning into phage M13 prior to nucleotide sequencing studies (see chapter 4). Secondly, the restriction endonuclease map provided reassurance of the fidelity of the RNA-cDNA hybrid cloning method. Restriction endonuclease mapping data obtained from overlapping RNA-cDNA hybrid clones was internally consistent and moreover, corresponded exactly to that obtained from clones made using the double-stranded cDNA method. This ruled out the possibility that cDNA cloned by the RNA-cDNA hybrid method might be subject to extensive insertions, deletions or other re-arrangements (c.f. Okayama and Berg, 1982). Finally, it was possible to compare the restriction endonuclease map obtained with that of other strains of poliovirus. No differences in the cleavage positions of these 10 restriction endonucleases was observed between P3/Leon 12 a,b #411 and the total genomic cDNA synthesised from the related viruses P3/Leon/37 #960 and P3/119 #643 (Dr. G. Stanway, personal communication). This concurs with previous information on the similarity of these three strains obtained by T1 oligonucleotide fingerprinting experiments (see chapter 2). In contrast, when the restriction endonuclease map of P3/Leon 12 a,b #411 was compared with a computer generated restriction endonuclease map of the serotype 1 strain of poliovirus P1/Mahoney (derived from the nucleotide sequence of this strain determined by Racaniello and Baltimore, 1981a), the type 1 and type 3 restriction endonuclease maps were very different. This not unexpected result agrees with other information on the nucleotide sequence homology between type 1 and type 3 strains of poliovirus, as determined by RNA-cDNA hybridisation experiments, which suggested that 75% nucleotide sequence homology exists between these two serotypes (Tracey and Smith, 1981).
3.D CONCLUSIONS

For the efficient exploitation of the molecular cloning techniques used in this project, it was essential to be able to characterise large numbers of bacterial colonies bearing recombinant plasmids. This requirement has been met by the use of colony hybridisation techniques and rapid, micro-scale plasmid isolation methods such as have been described in this chapter. Initial characterisation of cDNA clones synthesised from P3/Leon 12 a₁ b #411 vRNA was performed using these techniques. More detailed information was then obtained by restriction endonuclease mapping of the cDNA inserts contained in recombinant plasmids. Such studies yield important information on the size and complexity of the virus genome and are intrinsically interesting, but are also important prior to experiments such as nucleotide sequencing or the construction of a full length cDNA copy of the virus genome in a bacterial plasmid vector (see chapter 4).

Overlapping cDNA clones corresponding to the entire genome of P3/Leon 12 a₁ b #411 have been synthesised. This result confirms the quality of the cDNA synthesised under the conditions described in chapter 2. The clones have been identified and characterised using the methods described in this chapter. The RNA-cDNA hybrid cloning method devised compares favourably in terms of overall efficiency and convenience with the much more commonly used double-stranded cDNA cloning method (chapter 2). In addition, the size and genomic distribution of cDNA clones made by the former method is superior to that seen with the double-stranded method. cDNA clones extending from the 3' to the 5' end of the genome can be made without recourse to primer extension or other techniques. Thus the RNA-cDNA hybrid method has proved to be ideally suited to the cloning of picornavirus genomes.
CHAPTER 4

NUCLEIC ACID SEQUENCING OF CLONED POLIOVIRUS GENOMES
4. A INTRODUCTION

Nucleotide sequencing provides a powerful tool for genetic analysis. The first rapid nucleotide sequencing method to be developed was the "plus-minus" method of Sanger and Coulson (1975). This involved primed DNA synthesis \textit{in vitro} and was far more powerful than previous degradative methods of sequence determination. Two further rapid methods of DNA sequence determination have since been devised. The method of Maxam and Gilbert (1977) relies on base-specific partial cleavages of an end-radiolabelled DNA fragment to determine its nucleotide sequence. Sanger et al (1977) published a further method in which primed DNA synthesis \textit{in vitro} is performed in the presence of chain-terminating inhibitors (dideoxynucleotides). The relative merits of these methods have been much discussed. All rely on the use of high resolution polyacrylamide gel electrophoresis to fractionate DNA oligonucleotides differing in size by only a single nucleotide. Technical aspects of the Sanger dideoxynucleotide method have been improved since its original publication. In particular, the development in M13 sub-cloning methods, specifically designed to produce single-stranded DNA templates (section 4.B) has increased to the efficiency of this method (Sanger et al, 1980). The cost of and time taken by the two methods is now similar, with the reservation that the dideoxynucleotide method may be faster for long DNA sequences (e.g. more than 2kbp). When a carefully devised sequencing strategy is followed (i.e. determination of the sequence of both strands of the DNA from independent sub-clones), the accuracy of the results obtained from both methods is very good indeed.

The three strains of poliovirus used in this study were so closely related as to be indistinguishable by T1 oligonucleotide mapping (chapter 2) or restriction endonuclease mapping of cloned cDNA (chapter 3). It was therefore believed that any differences seen in their respective nucleotide sequences would be likely to be related to the differences in
pathogenicity between them. (As described in chapter 1, nucleotide sequencing of neurovirulent and attenuated type 1 strains of poliovirus produced results of uncertain significance (Kitamura et al, 1981; Racaniello and Baltimore, 1981a; Nomoto et al, 1982b) With the aim of determining precisely the features of the genome responsible for the attenuation of and reversion to neurovirulence in poliovirus type 3, the nucleotide sequences of cloned cDNA from the three strains of poliovirus described earlier were determined using the dideoxynucleotide method.
4. B M13 CLONING

M13 was first characterised by Hofschneider (1963) and is a rod-shaped phage of about 900 nm in length, 6 nm in diameter, containing no lipid or carbohydrate. The genome is a circular single-stranded DNA molecule of approximately 6.5 kb (11% of the weight of the particle). The M13 genome contains 8 genes, which comprise about 90% of the total coding capacity. Amber mutations in most of these genes are lethal, but there is an intergenic region of approximately 500 bp which is non-essential. M13 has the following advantages as a cloning vector:

- Its structure, biology and genetics have been well studied (see Lewin, 1977).
- Non-lytic infection means that large amounts of phage are produced (up to 100 mg of phage, 10 mg of DNA/litre). RF is present at about 300 copies/cell and is easily isolated in a similar way to plasmid DNA (Marvin and Hohn, 1969). Therefore, it is possible to isolate relatively large quantities of both double-stranded (RF) and single-stranded (virus) DNA from small-scale cultures.
- Even at high multiplicities of infection, only 1 or 2 phage genomes enter each cell, therefore recombination is unlikely to occur (Marvin and Hohn, 1969).
- The filamentous shape of the particle and the fact that assembly occurs as the particle is extruded from the cell means that up to approximately 6 genome equivalents may be packaged in the particle, making the M13 genome an ideal insertion vector (Messing et al, 1977).

Messing et al (1977) adapted wild-type M13 to function as a cloning vector for dideoxynucleotide sequencing. A restriction endonuclease fragment from the lac operon of E. coli, coding for 145 amino acids of the β-galactosidase alpha-peptide (lacZ), was inserted into the non-essential intergenic region. The resulting phage (M13 mp1, figure 4.1) expressed the lacZ amino acids and could complement lac deletions in host cells,
Figure 4.1
Restriction endonuclease map of M13 mp1

M13 Cloning Sites

Bgl II 6930

Eco B 7185

Hgi E II 6453

Avu I 6400

Bgl I 6426

Sau I 6503

Cvn I

Nar I, Aos II 6001

Ava II 5914

Hgi I II 5643

Nae I 5613

Bgl I

BstX I 2368

5000

4000

3000

2000

1000

6000

Ori

6400

V

Vi

III

IV

VIII

X

X

X

X

X
conferring a lac\textsuperscript{+} phenotype. Selection of M13 mp1 infected cells was thus easy to achieve, infected cells giving rise to turbid blue plaques in the presence of iso-propyl-\(\beta\)-d-thio-galacto-pyranoside (IPTG), a lac inducer and 5-bromo-4-chloro-3-indolyl-\(\beta\)-galactoside (BCIG), a chromogenic substrate. A unique EcoRI cleavage site was created in the lac fragment resulting in a vector (M13 mp2) in which the lac\textsuperscript{+} phenotype was destroyed by insertional inactivation of the lacZ alpha-peptide (Gronenborn and Messing, 1978). To improve the cloning potential of the vector, a multi-purpose cloning sequence containing 6 unique restriction endonuclease cleavage sites was inserted into the lac fragment (M13 mp7) (Messing et al, 1981). This artificially synthesised sequence coded for 14 extra amino acids, but was inserted in phase with lac transcription and did not destroy the lac\textsuperscript{+} phenotype. Thus recombinants could still be detected by insertional inactivation and it was possible to clone DNA fragments resulting from a wide variety of restriction endonuclease digests. Recent improvements to M13 vectors include two phages (M13 mp8 and mp9) containing a more sophisticated multi-purpose cloning sequence with 9 restriction endonuclease cleavage sites (Messing and Vieira, 1982). This sequence is present in opposite orientations in the two phages, making possible "forced cloning" of particular orientations of fragments derived from multiple restriction endonuclease digestions. The 9 site multi-purpose cloning sequence and the lac fragment have also been inserted into a derivative of plasmid pBR 322 (Bolivar et al, 1977) to create a pair of plasmid cloning vectors (pUC 8 and 9) with equivalent attributes to M13 mp8 and mp9 (Vieira and Messing, 1982).
Special strains of *E. coli* have also been constructed by Messing et al to act as hosts for these vectors. Three of these strains are:

- **JM101** - lacproA, thi, supE, F′traD36, proAB, lacM15
- **JM103** - lacproA, thi, strA, supE, endA, sbcB15, F′traD36, proAB, lacM15
- **JM105** - lacproA, thi, strA, supE, endA, sbcB15, F′traD36, proAB, lacM15

*E. coli* JM103 (Messing et al, 1981) was used exclusively throughout the work described in this chapter.

4.C.1 - M13 Sub-cloning

When this study was initiated, determination of the complete nucleotide sequence of a picornavirus genome was beyond the capacity of a single person. Consequently, the work described in this section (4.C) was a joint effort between Drs. J. W. Almond, R. Hauptmann, G. Stanway and myself. The sequence of P3/Leon 12 a,b #411 was determined from restriction endonuclease fragments of previously characterised cDNA clones (chapter 3) which were sub-cloned into the M13 vectors mp8 and mp9 (see section 4.B). Methods used for restriction endonuclease digestions, isolation of DNA from agarose gels, preparation of double-stranded and single-stranded M13 DNA's and sub-cloning of restriction endonuclease fragments are described in chapter 6.

For each individual plasmid, the restriction endonuclease fragments sub-cloned were chosen with reference to the previously determined restriction map of the P3/Leon 12 a,b #411 genome (chapter 3). Recombinant (lac\(^{-}\)) phage were isolated as "white" plaques in the presence of BCIG and IPTG. The most frequently used sub-cloning method was "shotgun" cloning of all the fragments of a restriction digest (most commonly AluI, HaeIII or Sau3A). Forced cloning in specially prepared vectors was also used to isolate certain restriction fragments. Sequencing was begun before the complete restriction endonuclease map of the genome had been established. The first clones sequenced were those from the 3' end (pSGA 31 and pSAG 6). As the sequence was built up in the 3'-5' direction, other clones extending further towards the 5' end of the genome were used. The names and genomic locations of the clones used are shown in figure 4.2 and included clones made by both the double-stranded cDNA and RNA-cDNA hybrid methods. Several clones were often used in parallel to provide overlapping sequence data from a particular region of the genome. This was useful, as
Restriction endonuclease fragments used in the determination of the complete nucleotide sequence of the P3/Leon 12, 13, #411 genome.
it not only provided confirmation of the sequence, but also assessed the relative fidelities of the two cloning methods. As stated in chapter 3, no differences were observed between double-stranded cDNA and RNA-cDNA hybrid clones. In all, 11 different clones were used to obtain the complete sequence. 85% of the sequence was determined from at least duplicate sequencing reactions and 63% was determined in both orientations. In addition, the sequence was checked by comparison with the sequences of P3/Leon/37 and P3/119 later obtained (see next section) and with the published sequence of P1/Mahoney (Racaniello and Baltimore, 1981a) to provide confirmation of the continuity of the long open reading frame (see below).

4.C.2 – Nucleotide Sequencing

Initial screening - In early experiments it was found to be necessary to screen phage templates prior to nucleotide sequencing to distinguish between recombinant phage and white plaques arising from phage deletion mutants. The number of deletion mutants isolated varied widely between different cloning experiments, e.g. from 5-100%. Screening was therefore necessary to remove these mutants, as the total number of templates which could be sequenced was a limiting factor. An additional advantage of performing this screening was that information was obtained on the approximate size, genomic location and orientation of the M13 sub-clones which avoided multiple sequencing of a single region. The screening method used was the single-tracking procedure described in chapter 6, in which a reaction containing ddTTP was performed to obtain a characteristic "T-track" of the cDNA insert. Up to 26 templates were examined in a single experiment and analysed on a 40 cm x 20 cm x 0.37 mm 6% (w/v) 8 M urea polyacrylamide gel containing TBE1 buffer (chapter 6). Templates selected by this procedure were then sequenced as described below.
Dideoxynucleotide sequencing - The method of Sangar et al (1977) was used to obtain the results described in this section (4.C) (chapter 6, method 1). Alterations to the published method included the use of Eppendorf tubes rather than glass capillaries for both the priming and synthetic reactions and slight adjustments to the dideoxynucleotide concentrations used. These are given in table 6.1 (chapter 6). Working strength solutions of dideoxynucleotides were stored at -20°C, but were found to deteriorate after 10-14 days or 5-10 freeze-thaw cycles. For this reason, fresh dideoxynucleotide solutions were made up from 4 mM stocks (stable at -20°C) when necessary. In general, it was possible to read about 200 nucleotides from a 2 h gel run, progressing to 300 from a 4 h gel and a maximum of approximately 450 from a 6-7 h gel. The gel system used was the same as that for single-tracking (above). Four phage templates (16 tracks) were analysed on each gel. Drying the gels prior to autoradiography resulted in only a slight increase in the resolution and the length of the sequence which could be determined. Therefore gels were not routinely dried, but fixed, blotted and covered with Saran-wrap for autoradiography (chapter 6). The gels were exposed for between 12 h and 5 days to Kodak x-omat S or DuPont Cronex X-ray film. Intensifier screens were not generally used since these tended to reduce resolution on the autoradiographs.

Because of the production of well-known artefacts (Sangar et al, 1977), some experience was necessary before accurate sequence data could be obtained from the autoradiographs:-

A-track - The second A of a doublet is often more intense than the first. In runs of A residues, the first band is most intense and successive bands progressively weaker.

G-track - The second G of a doublet is often more intense than the first, always if preceded by a T residue. Runs of G residues are often compressed with abnormal spacing.

C-track - The second C of a doublet is always more intense than the first.
Single C bands are often comparatively weak. The second band of a run of C residues is more intense than the first and successive residues, except when the run is preceded by a G residue. Runs of C residues are often compressed.

**T-track** - Single bands are intense. The first residue of a run is the most intense, the others progressively weaker.

In general, these artefacts were easily recognisable and could be overcome by careful inspection of the autoradiograph, especially with respect to the spacing of the bands. The exception to this were G or C compressions. These are believed to occur as a result of secondary structure in the newly synthesised DNA. Two specific problems exist. Firstly, runs of G (and possibly C) residues form helicities, altering their migration rate and resulting in a "pile-up" of bands (Gellert et al., 1962). Secondly, when 3-4 consecutive G/C residues are present with a complementary sequence within 10-15 nucleotides, a hairpin loop may form. These loops are stable when they are present at the 3' end of the synthesised chain and migrate faster than expected for their length (Dr. A. T. Bankier, personal communication). Several methods have been devised to overcome these problems, e.g. the substitution of dITP for dGTP (resulting in weaker base-pairing) (Mills and Kramer, 1979), modification of cytosine residues to 5,6-dihydro-6-sulpho-methoxycytosine (Ambartsunyan and Mazo, 1980), or inclusion of 25-50% formamide in the gel-buffer system (Dr. A. T. Bankier, personal communication). Fortunately, relatively few problems of this type were experienced with P3/Leon 12 a b #411. When such difficulties were encountered they were resolved by running the gels at 45 W (rather than 40 W) so that they were hotter and more denaturing, by sequencing such regions from both strands of the cloned DNA to read into the compression from both ends and finally by sequencing the region several times from independent M13 sub-clones.

Very slight alterations in the dideoxynucleotide concentration (e.g. 40.01 mM) had a pronounced effect on the length of sequence which could be
determined, reducing this from 300-100 nucleotides in some instances. This meant that each new batch of reagents had to be individually optimised. Poor incorporation of radioactivity or background darkening of the autoradiograph due to radiolysis of labelled DNA also decreased the length of sequence which could be determined. Most troublesome of all was the occurrence of multiple bands, either as a "complete ladder" in one or all of the tracks, or as several bands in the same position on the gel. While this problem was partly sequence-specific, the most frequent causes were due to the polymerase, which was found to deteriorate very rapidly at room temperature, or to decomposition of the dilute dideoxynucleotides. This problem was usually solved by use of either a new batch of enzyme or a fresh dilution of the nucleotide stocks.

4.C.3 - Data Handling

Nucleotide sequencing of a molecule such as the P3/Leon 12 a₁b #411 genome rapidly generates a large amount of data. A computer based system is the most practical means of recording, storing, assembling and analysing such results. Nucleotide sequence data from P3/Leon 12 a₁b #411 was stored and processed using a PDP 11/44 computer and a software package (written in FORTRAN) obtained from R. Staden, LMB, Cambridge, UK (Staden, 1980). A more recent, improved version of the package used during these studies is now available, (Staden, 1982). The following is a summary of the programs used:-

**BATIN** Gel data is entered into the computer and adjusted to a format of 80 characters per line. Such data forms the input processed by subsequent programs. A file of gel-reading names is created and single-character ambiguity codes are recognised.

**DBCMP** Gel readings are compared (in both orientations) with stored data and with each other to detect overlaps of specified length with the existing consensus sequence (see below). DBCOMP was also routinely used to
compare new data with the sequence of P1/Mahoney (Racaniello and Baltimore, 1981a). Any gels not matching with either of these sequences were compared with the vector sequences of pAT 153 and M13 mp7. The sequences are printed out and matches are indicated by "*". Data from DBCOMP was used to re-examine autoradiographs and edit gel readings. 

**DBUTIL** This system is an assembly of subroutines selected from a 6 option menu. Options available are:

- **OPTION NUMBER = 0 STOP** (exit DBUTIL).

- **OPTION NUMBER = 1 ENTER** Gel readings located by DBCOMP may be entered into the database. The new gel is positioned relative to the existing sequences ("contigs") in the database with which it overlaps and both gel and contig may then be edited to align the new gel before entry is completed. Mismatches between the new gel and the contig are displayed as "*".

- **OPTION NUMBER = 2 PRINT** Lists gels in specified contigs in the database in the form of "sorted data" consisting of the following parameters, gel name, gel number in database, starting position of gel reading, length of gel (nucleotides from left end of contig), number of gel which overlaps on left and number of gel which overlaps on right.

- **OPTION NUMBER = 3 DISPLAY** Displays all gel readings and consensus sequence derived from them for a specified region of a contig. The consensus sequence calculated by DBUTIL requires at least 75% agreement between overlapping gel readings at each position, or else "-" is displayed.

- **OPTION NUMBER = 4 JOIN** Overlapping but separate contigs may be joined to form longer contigs, resulting eventually in a single unbroken contig. Contigs may be complemented if necessary before joining.

- **OPTION NUMBER = 5 COMPLEMENT** Contigs or the whole database may be complemented if required.

- **OPTION NUMBER = 6 EDIT** The database may be edited by inserting, deleting or changing a specified number of characters at a specified location if a mistake is discovered, to remove ambiguity codes or during the entry of new gels.
Using the above programs, the complete nucleotide sequence of P3/Leon 12\textsubscript{a,b} #411 was built up. Subsequent analysis of the consensus sequence produced by the program CONSEN was performed using the following programs:

**TRANMT** Translates (in 1-3 possible reading frames) a specified region of a consensus sequence, displaying both the nucleotide and amino acid sequences (using the single letter amino acid code).

**HAIRPN** Searches a consensus sequence for palindromes, inverted repeats and potential hairpin loops. Output is sorted and displayed digitally in the form of 3 parameters, length of palindromic region, length of intervening sequence, position of start of loop.

**RUDMAP** Unlike the Staden package, this is a BASIC program run on a Cyber 73 computer. The program draws a computer predicted restriction endonuclease map from a nucleotide sequence. The predicted restriction pattern for 64 enzymes on P3/Leon 12\textsubscript{a,b} #411 is shown in figure 4.3. The cleavage positions of the 10 restriction enzymes with which the cDNA was physically mapped (chapter 3) are in agreement with the map in chapter 3.

4.C.4 - The Complete Nucleotide Sequence of P3/Leon 12\textsubscript{a,b} #411

While the work described in this chapter was in progress, the complete nucleotide sequences of two other strains of poliovirus were determined in independent laboratories. These strains were P1/Mahoney (Kitamura et al., 1981; Ranaciello and Baltimore, 1981a) and P1/Sabin (Nomoto et al., 1982b). Together with earlier biochemical studies (chapter 1), this work allowed an accurate and detailed functional map of the poliovirus genome to be compiled (Kitamura et al., 1981). The nucleotide sequence presented here was the first to be determined for a type 3 strain of poliovirus (Cann et al., 1983; Stanway et al., 1983a). No direct information on the genetic map of poliovirus type 3 was obtained during these studies. However, the similarity in the structure, biology,
Figure 4.3
Computer predicted restriction endonuclease map of P3/Leon 12 a, b #411
pathogenicity and predicted amino acid sequences of the two strains strongly implies that the functional regions which have been identified for P1/Mahoney (Kitamura et al, 1981) are directly applicable to P3/Leon 12 a,b #411. The genome of P3/Leon 12 a,b is 7432 nucleotides long (excluding the poly(A) tract), similar in length to P1/Mahoney (7440 nucleotides). There is 77.4% nucleotide sequence homology between the two serotypes. This figure is slightly higher than a previous estimate of the similarity between the two serotypes (Young, 1973) but in agreement with a more recent measurement based on cDNA hybridisation (Tracey and Smith, 1981). Comparison of the P3/Leon 12 a,b #411 with the P1/Mahoney sequence was valuable as it was possible to align the two sequences very closely. The sequence determined by Racaniello and Baltimore was arbitrarily chosen for the purpose of comparison throughout the rest of this chapter. Both the type 1 and type 3 sequence contain a single long open reading frame large enough to contain the virus coded polyprotein (chapter 1). P3/Leon 12 a,b #411 contains 3 fewer codons than P1/Mahoney (2206 c.f. 2209). The complete nucleotide sequence of P3/Leon 12 a,b #411 together with its predicted amino acid sequence and the location of the polyprotein cleavage sites used in P1/Mahoney (see below) are shown in figure 4.4.

Non-coding regions - Sequencing experiments have demonstrated that the RNA-cDNA hybrid method allows the 5' terminus of the genome to be cloned directly (see chapter 3). This has now been confirmed for three type 3 strains of poliovirus (section 4.D) and indicates that the presence of VPg attached to the 5' end of the genome does not interfere with reverse transcription of the RNA (see Racaniello and Baltimore, 1981a). The lengths of the non-coding regions of P1/Mahoney, P1/Sabin and P3/Leon 12 a,b #411 are identical in each case, 742 nucleotides at the 5' end and 72 nucleotides at the 3' end of the genome. Although the P1/Mahoney sequence determined by Racaniello and Baltimore (1981a) has only a 71 nucleotide non-coding region at the 3' end, the 3' terminal G residue being missing, this residue is present in the other two type 1
The complete nucleotide sequence of P3 Leon 12 a, b

Figure 4.4a

The complete nucleotide sequence of P3 Leon 12 a, b

Figure 4.4b

The complete nucleotide sequence of P3 Leon 12 a, b

Figure 4.4c

The complete nucleotide sequence of P3 Leon 12 a, b

Figure 4.4d

The complete nucleotide sequence of P3 Leon 12 a, b

Figure 4.4e

The complete nucleotide sequence of P3 Leon 12 a, b
Figure 4.4c
sequences (Kitamura et al., 1981; Nomoto et al., 1982b) The 3' terminal 48 nucleotides of the coding region are also conserved between P1/Mahoney and P3/Leon 12 a,b #411, making a total stretch of 119 conserved nucleotides at the 3' end. Such conservation implies a sequence specific functional role for this region, although it is diverged extensively in other picornaviruses (Black et al., 1978; Fellner, 1979; M. Ryan, personal communication). The 742 nucleotide 5' non-coding region shows homology between P1/Mahoney and P3/Leon 12 a,b #411. Although the overall homology of this region is only approximately 80%, some sections are more strongly conserved than others. For example, a 60 nucleotide sequence (positions 513-573) (510-570 in P1/Mahoney) is totally conserved. Perhaps significantly, the 5' terminal 21 nucleotides are conserved between all the strains which have been sequenced. (P1/Mahoney and P3/Leon 12 a,b #411 have only a single difference, A-T at position 22, in the first 54 nucleotides.) This suggests that this sequence is necessary for some aspect of virus replication. The stem of the 10 base-pair stem and loop structure in P1/Mahoney (positions 9-36) identified by Larsen et al. (1981) is totally conserved in P3/Leon 12 a,b #411, although there is a change in the loop (A-T at position 22). The conservation of this stable structure argues that it is of functional significance.

In the 5' non-coding region before the start of the long open reading frame, there are 7 other methionine codons (8 in P1/Mahoney) which could be used to initiate translation, but most of the ensuing reading frames are short and are not conserved between types 1 and 3 and are thus of dubious significance. One of the potential reading frames (position 239-325, 28 codons) remains free of termination signals in both P1/Mahoney and P3/Leon 12 a,b #411, but there is no evidence that such a peptide is made in P3/Leon 12 a,b infected cells (Dr. P. D. Minor, personal communication). The mechanism by which ribosomes select the apparently unique start codon at position 743 remains unknown, but initiation of poliovirus translation may be different from the normal mechanism which is
thought to act in eukaryotic cells (see chapter 1). The 40 nucleotides on the 5' side of the initiation codon are poorly conserved between the types 1 and 3, but interestingly, the 5'-A/CNNAUGG-3' sequence suggested by Kozak (1981) as being important in the selection of initiation codons is conserved between the two strains (see chapter 1).

**Coding region** - There is 89.6% predicted amino acid homology between the computer predicted polyproteins of P1/Mahoney and P3/Leon 12 a,b. This similarity is reflected in table 4.1 which shows the codon usage and amino acid compositions of the two polyproteins (see Racaniello and Baltimore, 1981a). The degree of amino acid conservation is not uniform along the whole length of the polyprotein. The region of the genome coding for structural proteins shows significantly less conservation than the non-structural region. This is illustrated in figure 4.5, where each vertical line represents an amino acid insertion, deletion or substitution. This pattern may be due to a greater selection pressure on the structural proteins, or simply reflect lower amino acid sequence flexibility in non-structural proteins whose role is mainly enzymatic. Most of the amino acid changes result from point mutations, but in a few instances, a whole codon has been inserted or deleted (see Stanway et al., 1983b). The similarity of the VP1 proteins of types 1 and 3 (Stanway et al., 1983b) points to a similar 8 amino acid sequence in P1/Mahoney being involved in virus neutralisation to that in type 3 (Minor et al., 1983; Evans et al., 1983; chapter 1).

The overall similarity of the two polyproteins and the high degree of homology between the sequences of the putative proteinase P3-7c (Hanecak et al., 1982) (97% amino acid homology, Cann et al., 1983) implies that an identical polyprotein processing mechanism occurs in the two strains. This is also suggested by the observation that all the glutamine-glycine amino acid pairs (14) in the P1/Mahoney polyprotein are conserved in P3/Leon 12 a,b, as are all the tyrosine-glycine pairs (9) and the asparagine-serine pair forming the cleavage site between VP4 and VP2.
Figure 4.5
Location of amino acid sequence differences between P3/Leon 12 a,b #411 and P1/Mahoney

Each vertical line represents a single amino acid substitution, insertion or deletion.
### Table 4.1

**Codon Usage in Open Reading Frame:**

(P3/Leon 12 a, b #411 / P1/Mahoney)

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40/38</td>
<td>18/20</td>
<td>36/40</td>
<td>23/24</td>
</tr>
<tr>
<td></td>
<td>43/45</td>
<td>38/32</td>
<td>66/58</td>
<td>18/18</td>
</tr>
<tr>
<td>L</td>
<td>21/24</td>
<td>44/48</td>
<td>STOP</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td>45/37</td>
<td>7/10</td>
<td>STOP</td>
<td>0/0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>W 27/28</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17/25</td>
<td>22/30</td>
<td>17/18</td>
<td>8/6</td>
</tr>
<tr>
<td></td>
<td>27/26</td>
<td>21/20</td>
<td>31/33</td>
<td>13/6</td>
</tr>
<tr>
<td></td>
<td>32/33</td>
<td>59/55</td>
<td>54/40</td>
<td>6/3</td>
</tr>
<tr>
<td></td>
<td>34/33</td>
<td>16/13</td>
<td>37/43</td>
<td>3/7</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65/51</td>
<td>55/53</td>
<td>49/50</td>
<td>25/20</td>
</tr>
<tr>
<td></td>
<td>43/49</td>
<td>43/54</td>
<td>61/68</td>
<td>22/21</td>
</tr>
<tr>
<td></td>
<td>23/34</td>
<td>47/52</td>
<td>69/63</td>
<td>49/47</td>
</tr>
<tr>
<td></td>
<td>65/68</td>
<td>10/13</td>
<td>55/61</td>
<td>18/24</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28/21</td>
<td>49/49</td>
<td>59/54</td>
<td>43/37</td>
</tr>
<tr>
<td></td>
<td>24/31</td>
<td>44/34</td>
<td>60/64</td>
<td>25/19</td>
</tr>
<tr>
<td></td>
<td>26/18</td>
<td>59/59</td>
<td>46/63</td>
<td>36/47</td>
</tr>
<tr>
<td></td>
<td>67/59</td>
<td>16/18</td>
<td>64/46</td>
<td>38/39</td>
</tr>
</tbody>
</table>

**Base Composition (Whole Genome):**

<table>
<thead>
<tr>
<th>T</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.2/23.9</td>
<td>23.3/23.2</td>
<td>28.9/29.9</td>
<td>23.4/22.9</td>
</tr>
</tbody>
</table>

**Amino Acid Composition of Polyprotein:**

**P1/Mahoney:**

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>L</th>
<th>I</th>
<th>M</th>
<th>V</th>
<th>S</th>
<th>P</th>
<th>T</th>
<th>A</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>84</td>
<td>180</td>
<td>132</td>
<td>68</td>
<td>138</td>
<td>151</td>
<td>118</td>
<td>172</td>
<td>160</td>
<td>98</td>
</tr>
<tr>
<td>%  (w/w)</td>
<td>5.01</td>
<td>8.26</td>
<td>6.06</td>
<td>3.62</td>
<td>5.55</td>
<td>5.33</td>
<td>4.65</td>
<td>7.05</td>
<td>4.61</td>
<td>6.48</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>82</td>
<td>117</td>
<td>124</td>
<td>118</td>
<td>109</td>
<td>41</td>
<td>29</td>
<td>95</td>
<td>142</td>
</tr>
<tr>
<td>%  (w/w)</td>
<td>2.84</td>
<td>4.26</td>
<td>5.41</td>
<td>6.44</td>
<td>5.51</td>
<td>5.71</td>
<td>1.71</td>
<td>2.19</td>
<td>6.02</td>
<td>3.29</td>
</tr>
<tr>
<td>Mr  = 246,336</td>
<td>Polarity index = +46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**P3/Leon 12 a, b #411:**

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>L</th>
<th>I</th>
<th>M</th>
<th>V</th>
<th>S</th>
<th>P</th>
<th>T</th>
<th>A</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>83</td>
<td>178</td>
<td>133</td>
<td>65</td>
<td>144</td>
<td>152</td>
<td>118</td>
<td>154</td>
<td>168</td>
<td>102</td>
</tr>
<tr>
<td>%  (w/w)</td>
<td>4.96</td>
<td>8.18</td>
<td>6.11</td>
<td>3.46</td>
<td>5.80</td>
<td>5.38</td>
<td>4.66</td>
<td>6.33</td>
<td>4.85</td>
<td>6.76</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>90</td>
<td>110</td>
<td>124</td>
<td>118</td>
<td>110</td>
<td>41</td>
<td>27</td>
<td>98</td>
<td>142</td>
</tr>
<tr>
<td>%  (w/w)</td>
<td>2.68</td>
<td>4.69</td>
<td>5.10</td>
<td>6.46</td>
<td>5.52</td>
<td>5.77</td>
<td>1.72</td>
<td>2.04</td>
<td>6.22</td>
<td>3.29</td>
</tr>
<tr>
<td>Mr  = 245,806</td>
<td>Polarity index = +46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The pattern of processing likely to occur in P3/Leon 12 a₁b #411 is shown in figure 1.1 (see chapter 1). Alternative cleavage pathways occur at tyrosine-glycine pairs in P1/Mahoney (Hanecak et al., 1982; Semler et al., 1983) (equivalent to positions 6284 and 6422 in P3/Leon 12 a₁b #411). These alternative cleavage products are not found in cells infected with P3/Leon 12 a₁b or related strains (Dr. P. Minor, personal communication) although the amino acid sequences through these regions are conserved.
4. D NUCLEOTIDE SEQUENCING OF NEUROVIRULENT TYPE 3 POLIOVIRUSES


The cloning experiments described in this section (4. D. 1) were performed by Dr. G. Stanway and are described here as a background to the following sections. RNA-cDNA hybrid clones of the neurovirulent type 3 strains P3/Leon/37 #960 and P3/119 #643 were obtained by methods similar to those described in chapters 2 and 3. The size of the longest clone obtained was 5.2 kbp. No artefacts such as insertions, deletions or re-arrangements of the cDNA were observed, once again confirming the suitability of the RNA-cDNA hybrid cloning method. The restriction endonuclease maps of the clones of both strains were identical to that of P3/Leon 12 a, b # 411. Restriction endonuclease fragments of subgenomic clones were used to construct a complete copy of the genomes of P3/Leon/37 #960 and P3/119 #643 in pAT 153 (each within a single plasmid) referred to as plasmids pOLIO Leon and pOLIO 119 respectively. The sequence of P3/Leon/37 #960 was determined by Dr. G. Stanway using similar methods to those described below.

4. D. 2 - Nucleotide Sequencing of P3/119 #643

M13 Sub-cloning - 50 μg of pOLIO 119 DNA was digested with PstI and subjected to electrophoresis on a 1% (w/v) agarose-Elfo gel. PstI digestion resulted in 3 fragments, 3.6 kbp (vector), and 4.9 and 2.6 kbp (insert). The two insert fragments were isolated from the gel and processed separately. Randomly generated M13 sub-clones were prepared using the method of Deininger (1983). The cohesive PstI termini were self-ligated and the resulting circular DNA molecules fragmented by sonication (chapter 6). Both of these reactions were monitored by electrophoresis of small aliquots of the DNA on agarose gels. The sheared
DNA was repaired to flush-ends using the "fill-in" reaction of the Klenow fragment of E. coli DNA polymerase I and fractionated by electrophoresis on a 1% (w/v) agarose-Elfo gel using EcoRI, HindIII digested \( \lambda^+c \) as a size marker (chapter 6). An estimated 0.5-1.5 kbp size fraction (see below) was isolated from the gel and redissolved in 20 \( \mu l \) of sterile distilled water. 2 \( \mu l \) aliquots of this DNA was ligated with 20 ng of CIP-treated, SmaI digested M13 mp8 RF and used to transfect E. coli JM103.

Nucleotide sequencing - Initial screening of the P3/119 #643 sub-clones was not necessary as the use of phosphatase-treated vector DNA reduced the background level of white plaques arising from M13 deletion mutants to less than 1%. This was a major contribution to the increased efficiency of the nucleotide sequencing strategy used for P3/119 #643. Recombinant phage templates were sequenced using the method of Biggin et al (1983) (chapter 6, method 2). The major differences between this method and that described for P3/Leon 12 a,b were the use of \( [\alpha^35S]dATP \) in place of \( [\alpha^32P]dATP \) and the use of buffer gradient gels to analyse the products of the sequencing reactions. Reaction conditions used differ slightly from those used in section 4.C and are described in chapter 6.

Electrophoresis was performed on 50 cm x 20 cm x 0.25 mm 6% (w/v) polyacrylamide 8 M urea gels, which contained (top-bottom) a 1-5 x TBE2 buffer gradient (Biggin et al, 1983). 6 templates (24 tracks) were analysed on each gel. Electrophoresis was carried out at a constant 40 W until the xylene-cyanol dye-marker was 10-15 cm from the bottom of the gel (approximately 10 nucleotides from the SmaI cloning site of M13 mp8). Routinely, each reaction was analysed on a single gel. The buffer gradient resulted in equal spacing of bands over the lower half of the gel. In combination with the increased resolution in the higher part of the gel resulting from the use of \(^{35}S\) radiolabel, the total amount of sequence which could be determined from each gel was increased to about 300 nucleotides/template. (On some occasions, more than 400 nucleotides were
determined from such gels.) Approximately \(300 \times 6 = 1800\) nucleotides of sequence could be determined from a single gel, compared with \(200 \times 4 = 800\) nucleotides from the gel system described in section 4.C. (For a direct comparison of the two systems, see Biggin et al, 1983.) As gel electrophoresis is often the limiting step in nucleotide sequencing studies, this represents an effective doubling in the amount of sequence which could be determined in a given time. The use of \(^{35}\)S-labelled radio-nucleotides has other advantages over \(^{32}\)P in that it gives a lower incident radiation dose during handling and that the labelled products are more stable than when labelled with \(^{32}\)P. One disadvantage of \(^{35}\)S is that since the emission energy of the \(^{35}\)S \(\beta\)-particle is much lower than that of \(^{32}\)P (0.049 and 0.695 MeV respectively) and thus more easily attenuated, gels must be dried to bring the X-ray film into close contact with the labelled material for autoradiography. Initially, gels were fixed by immersion in 10% (v/v) acetic acid for 5-10 min before transfer to Whatman 3M paper and drying. However, it was found that total immersion in acetic acid often resulted in the gel becoming detached from the glass plate. Since 0.25 mm thick 6% poly-acrylamide gels were somewhat fragile, gels were sometimes destroyed in this way. In later experiments, gels were overlain with Whatman 3M paper soaked in 10% acetic acid, allowed to fix for 5-10 min, then a stainless steel mesh clamped on top, trapping the gel, which was then washed under a slowly running tap. This resulted in improved removal of urea from the gel, without the risk of the gel being broken or distorted. Failure to remove urea from the gel resulted in distortion on drying and a tendency to stick to the X-ray film. Dried gels were entirely stable and the weaker emission energy of \(^{35}\)S could be compensated for by increasing the exposure time of the autoradiograph up to several weeks if necessary (though routine exposure times were 24-48 h, not significantly longer than for \(^{32}\)P).

The improved resolution obtained with \(^{35}\)S also resulted in less background darkening of autoradiographs and fewer superfluous bands due to
radiolysis of the DNA. Although the same problems with sequence artefacts arising from poor batches of enzyme (as described in section 4.C) were still experienced, autoradiographs of $^{35}$S-labelled DNA were generally easier to interpret than $^{32}$P-labelled material. One artefact not experienced with $^{32}$P was the presence of faint bands in the position beneath the first A-residue of a run in all four tracks. However, since when these were present they were fainter than the A-band, this artefact was easily recognised.

Because the M13 clones had been generated randomly, as the sequence neared completion increasing amounts of the data produced was found to cover previously sequenced regions. Remaining gaps in the sequence were therefore closed by repeat sequencing reactions to extend sequence obtained from the sub-clones located on either side of the gap. These reactions were subjected to electrophoresis for 3-6 h, depending on the length of sequence required, on non-gradient 40 x 20 x 0.25 mm 4% (w/v) polyacrylamide 8 M urea gels containing 1 x TBE2. Sequences of up to about 450 nucleotides from the start of the template could be determined from these gels. This extension strategy justifies the selection of a 0.5-1.5 kbp DNA size fraction during the sub-cloning. A high proportion of short clones would prevent sequence gaps being closed in this way (c.f. Deininger, 1983).

A particular problem was experienced with one gap in the sequence (position 5560-5580) where none of the M13 clones on either side were in the correct orientation to allow extension across the gap. This problem was overcome by selecting M13 sub-clones from the correct region of the genome in the following experiment. 1 µl of template DNA prepared from each of 44 unidentified sub-clones was spotted onto a nitrocellulose filter together with template DNA from positive and negative controls (a clone from the edge of the gap and one from the 5' end of the genome respectively). After pretreatment of the filter (chapter 6), the template DNA's were analysed by hybridisation to a radiolabelled probe known to
span the gap in the sequence. The probe used was the nick-translated insert of plasmid pSGA 27 (genomic location 5000-5750) derived from P3/Leon 12 #411 (Dr. G. Stanway, personal communication; chapter 3). 12 of the 44 templates tested hybridised to the probe. Nucleotide sequencing of these templates revealed that the cDNA inserts were located at least partly between sequence positions 5000-6000 and the gap in the sequence was filled in both orientations.

A total of 128 independent M13 sub-clones were used to determine the complete sequence. In all, a total of more than 30 kb of sequence was determined, an average of 240 nucleotides from each M13 clone. The genomic distribution of the sequences obtained from these clones is shown in figure 4.6 The apparently non-random distribution of the clones is believed to be due to the pOLIO 119 DNA being sheared by a single burst of sonication rather than by multiple shorter bursts (see, Deininger, 1983; chapter 6). As sequencing progressed, it was discovered that the plasmid pOLIO 119 lacked the 80 5' terminal nucleotides of the genome. This problem was resolved by cloning of the 5' terminal Sau3A fragment of the genome from the plasmid pOAG 12 (Dr. G. Stanway, personal communication) into PstI, BamHI doubly digested M13 mp9. These clones were then sequenced using the same conditions as for the randomly generated sub-clones of P3/119 #643.

Data handling - Nucleotide sequence data generated in the above reactions was processed in a slightly different way to that from P3/Leon 12 #411 (section 4.C). Sequence data from autoradiographs was entered into a PDP 11/44 computer using BATIN as before. DBCOMP was then used to compare the new data with the sequence of P3/Leon 12 #411. Because of the similarity of the two sequences, DBUTIL was not used to build a database on the computer. Instead, the output from DBCOMP was checked against the autoradiograph and any mistakes corrected and the genomic locations of the gel readings and positions of any differences from P3/Leon 12 #411
Figure 4.6
Distribution of sequences obtained from randomly generated M13 sub-clones of P3/119

Orientation of clones is shown by the circles, which represent the start of each sequence.
were recorded manually.

4.D.3 - The Molecular Basis of Attenuation and Reversion

Excluding the poly(A) tract, the genome of P3/119 #643 is 7429 nucleotides in length (c.f. 7431 for P3/Leon/37 #960 and 7432 for P3/Leon 12 a₁b #411). 66% of the sequence was determined in both orientations. The entire genome was sequenced at least twice (from independent sub-clones) with the exception of 160 nucleotides from the VP1 region, previously sequenced from subgenomic clones (Stanway et al., 1983b). The complete nucleotide sequence of P3/Leon/37 #960 discussed in this section was determined from the full-length copy of the virus genome in plasmid pOLIO Leon by Dr. G. Stanway and P. Hughes, using the same methods described in section 4.D.2 for P3/119 #643. Because of their similarity, the nucleotide sequences of the three strains are not presented here (that of P3/Leon 12 a₁b #411 is shown in figure 4.4) but the differences between them are summarised in figure 4.7. A total of ten nucleotide sequence changes were observed between P3/Leon/37 #960 and P3/Leon 12 a₁b #411, three of which give rise to predicted amino acid changes (figure 4.7). Four of the ten changes are purine-purine, four pyrimidine-pyrimidine, one purine-pyrimidine and one is a deletion (at the 3' terminus). Seven changes were observed between P3/Leon 12 a₁b #411 and P3/119 #643, two purine-purine, three pyrimidine-pyrimidine, one pyrimidine-purine and one a deletion at the 3' terminus). Three of these result in amino acid substitutions (figure 4.7). The similarity of the sequences allowed them to be cross-checked against one another in order to verify the changes seen. In the case of P3/119 #643, all of the changes have been sequenced in both orientations. The nucleotide sequence of P3/119 not only illustrates the basis of reversion to neurovirulence, but also helps in the interpretation of the process of attenuation. For this reason, the sequences of both P3/119 and P3/Leon/37 and their relation to
Figure 4.7
Nucleotide and amino acid sequence differences between neurovirulent and attenuated type 3 polioviruses
that of P3/Leon 12 a₁b are discussed below. The fact that eight of the nucleotide sequence changes seen between P3/Leon/37 and P3/Leon 12 a₁b are maintained in P3/119 (and also in other revertants detailed in table 4.2) provides strong evidence that P3/119 is a bona fide neurovirulent revertant of P3/Leon 12 a₁b and is not derived from a virulent sub-population of the vaccine strain (which was in any case subjected to plaque purification at the outset of this study, see chapter 2) or a naturally occurring independent derivative of P3/Leon/37.

Non-coding changes - Two changes were observed in the 5' non-coding regions of P3/Leon/37 and P3/Leon 12 a₁b. These were G-T at position 220 and C-T at 472. The latter change has been verified by direct sequencing of the primary pool of virus from which P3/Leon 12 a₁b #411 was derived (Dr. D. Evans, personal communication; see chapter 2) and by sequencing a different isolate of the Sabin type 3 vaccine strain (Dr. A. Nomoto, personal communication). These changes are presumed to be silent, since there is no evidence that this region of the genome is translated (see chapter 1). Neither of these positions seem to be involved in any easily predicted stable secondary structure of the RNA. In the coding region, there are four silent changes between these strains, G-A at position 871, T-C at 4064, T-C at 6127 and G-A at 7165. Three out of four of these changes are in third codon positions, the other (6127) is a conservative change in the first position of a codon. The remaining silent change is at position 7432 of P3/Leon 12 a₁b (the final residue of the 3' non-coding region), which is deleted in P3/Leon/37. This nucleotide was also deleted in one isolate of P1/Mahoney which has been sequenced (Racaniello and Baltimore, 1981a; see chapter 1).

One mutation in the 5' non-coding region occurs between P3/Leon 12 a₁b and P3/119, T-C at position 472. Although no obvious structural significance can be placed on this change, it is interesting that this is an exact reversal of one of the silent mutations observed between
P3/Leon/37 and P3/Leon 12 a_{1b}. This is the only direct back mutation of the changes between P3/Leon/37 and P3/Leon 12 a_{1b} which was observed. It may be significant that other vaccine-associated neurovirulent revertants also have C rather than T at this position (see below).

There are two silent changes in the coding region, A-G at position 1405 and C-T at 6036 (both third codon positions). In the 3' non-coding region, the two 3'-terminal residues of P3/Leon 12 a_{1b} (GG) are deleted (see above). Although the possibility that some of these changes may be involved in attenuation and subsequent reversion to neurovirulence cannot be ruled out, their minor nature suggests that they are not.

**Amino acid changes** - The three changes which result in amino acid substitutions between P3/Leon/37 and P3/Leon 12 a_{1b} are C-T at position 2034 (serine-phenylalanine, VP3), A-G at 3333 (lysine-arginine, VP1) and A-G at 3464 (threonine-alanine, P2-3b). The changes which result in amino acid substitutions between P3/Leon 12 a_{1b} and P3/119 are G-A at position 1548 (arginine-lysine, VP2), T-A at 1592 (leucine-methionine, VP2) and C-T at 2637 (alanine-valine, VP1). The possible significance of each of these changes is discussed below.

**The attenuation of neurovirulence** - Two of the amino acid changes observed between P3/Leon/37 and P3/Leon 12 a_{1b} appear to be conservative in their nature and effects. The lysine-arginine change (3333) occurs in a region of random-coil as predicted by the method of Chou and Fasman (1978) and does not result in a hydrophilicity change. Similarly, the threonine-alanine change, although predicted to occur in a region of $\beta$-pleated sheet, appears to be of minor structural importance and causes a minimal hydrophilicity change of -0.4 (threonine) to -0.5 (alanine) = -0.1 (Hopp and Woods, 1981). The serine-phenylalanine change is particularly interesting as it occurs in a region of predicted $\beta$-pleated sheet and destabilises a likely $\beta$-turn region in P3/Leon 12 a_{1b} (according to Chou...
and Fasman's criteria). Moreover, this substitution results in a hydrophilicity change from +0.3 (serine) to -2.5 (phenylalanine) = -2.8, which affects the hydrophilicity of the P-turn region (figure 4.8.A - calculated by the method of Hopp and Woods (1981) using a frame size of 6 amino acids).

Reversion to neurovirulence - The arginine-lysine change between P3/Leon 12 a1b and P3/119 occurs in a region of probable \( \beta \)-pleated sheet, but appears to be of little structural significance and causes no hydrophilicity change. The leucine-methionine change is also located in a \( \beta \)-pleated sheet region and is probably structurally unimportant, causing a small hydrophilicity change of -1.8 (leucine) to -1.3 (methionine) = +0.5 (figure 4.8.B). The third change, alanine-valine, occurs in a region of predicted random coil and is thus of uncertain significance, but results in a more significant hydrophilicity change of -0.5 (alanine) to -1.5 (valine) = -1.0 (figure 4.8.C).

It is not possible to draw firm conclusions about the precise significance of these individual changes to attenuation and reversion, since (with the exception of the silent change at position 472), there are no direct back mutations in P3/119 of the attenuating changes which occurred between P3/Leon/37 and P3/Leon 12 a1b. However, some tentative observations on their likely significance can be made. Two of the changes from P3/Leon/37 to P3/Leon 12 a1b (at positions 3333 and 3464) appear to have structurally and chemically minor consequences and it would seem unlikely that they are individually responsible for attenuation (although the possibility that they might contribute in some way towards the total effect cannot be ruled out). It is reasonable to expect that if the mutation(s) responsible for reversion to neurovirulence is located in the structural proteins (i.e. one or more of the three amino acid changes between P3/Leon 12 a1b and P3/119) that the attenuating change is located
Figure 4.8
Computer predicted hydrophilicity profiles of amino acid substitutions

4.8.A

2034

4.8.B

1592

4.8.C

2637
in these proteins also, thus making the threonine-alanine change in P2-3b (3464) an unlikely candidate. However, the serine-phenylalanine change at position 2034 appears to be more significant and must be regarded as the most likely candidate for the attenuation of P3/Leon/37 (see above). Similarly, the P3/Leon 12 a1, b to P3/119 changes at positions 1548 and 1592 are considered to be of minor structural and chemical significance. While the structural consequences of the alanine-valine change at position 2637 cannot be accurately predicted, it is responsible for a significant hydrophilicity change. Two other factors strengthen interest in the serine-phenylalanine change as a candidate for attenuation and alanine-valine for reversion to neurovirulence. The first is that these changes occur in VP3 and VP1 respectively and these two polypeptides are known to interact closely in the virus capsid (see discussion in chapter 1). Secondly, partial nucleotide sequence data has now been obtained for 5 other neurovirulent vaccine revertants (see, Minor, 1982). This data is summarised in table 4.2 (Dr. D. Evans, personal communication).

**TABLE 4.2 - Nucleotide Sequences of Neurovirulent Revertants**

<table>
<thead>
<tr>
<th>Position</th>
<th>472</th>
<th>1548</th>
<th>1592</th>
<th>1991</th>
<th>2034</th>
<th>2637</th>
<th>3333</th>
<th>3464</th>
</tr>
</thead>
<tbody>
<tr>
<td>P3/Leon/37</td>
<td>(C)</td>
<td>arg</td>
<td>leu</td>
<td>asp</td>
<td>ser</td>
<td>ala</td>
<td>lys</td>
<td>thr</td>
</tr>
<tr>
<td>P3/Leon 12 a1, b</td>
<td>(T)</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>phe</td>
<td>&quot;</td>
<td>arg</td>
<td>ala</td>
</tr>
<tr>
<td>P3/119</td>
<td>(C)</td>
<td>lys</td>
<td>met</td>
<td>&quot;</td>
<td>&quot;</td>
<td>val</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>P3/106</td>
<td>&quot;</td>
<td>arg</td>
<td>leu</td>
<td>&quot;</td>
<td>&quot;</td>
<td>ala</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>P3/115</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>P3/116</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>ser</td>
<td>thr</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>P3/122</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>phe</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>P3/131</td>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>asn</td>
<td>&quot;</td>
<td>val</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
No changes have yet been identified in three of these strains (P3/106, P3/115 and P3/122). However, P3/116 reverts from phenylalanine-serine at position 2034 (i.e. is P3/Leon/37-like at this point) and has an additional alanine-threonine mutation at 2637. P3/131 has the same alanine-valine change at position 2637 as P3/119 (with an additional G-A change at position 1991 (VP3), aspartic acid-asparagine).

Analysis of these other revertants suggests that there at least three possible ways by which the Sabin type 3 vaccine strain is able to revert to neurovirulence:—

1) By direct back mutation of the attenuating change(s). If the attenuating change is in fact the serine-phenylalanine mutation at 2034, then P3/116 would represent a revertant in this class.

2) By a P3/119-like mechanism, i.e. a suppressor mutation at a site other than that of the attenuating mutation, possibly the alanine-valine change at 2637. P3/131 would be a revertant in this class and the alanine-threonine mutation at 2637 in P3/116 may also act in a similar fashion.

3) By indirect suppression of the attenuating mutations due to changes at sites other than 2637 (not yet identified) e.g. P3/106, P3/115, and P3/122.
4.E CONCLUSIONS

Poliovirus cDNA from sub-genomic clones of P3/Leon 12 a₁b #411 was digested with restriction endonucleases and sub-cloned into M13 vectors. The single-stranded template DNA produced was sequenced by the dideoxynucleotide method. The complete nucleotide sequence was determined from overlapping restriction fragments. The nucleotide sequence of P3/119 #643 was determined from a library of M13 sub-clones containing randomly generated fragments of cDNA. (The sequence of P3/Leon/37 #960 was determined by Dr. G. Stanway using the same technique.) Comparison between type 3 and published type 1 sequences revealed that there was 89.6% homology in the predicted amino acid sequences of the two strains, and 77.4% nucleotide sequence homology. Nucleotide sequencing of conserved regions may eventually help to reveal the evolutionary relationships between the enteroviruses and other groups of picornaviruses.

A total of ten nucleotide sequence changes, resulting in three predicted amino acid substitutions, were observed between the neurovirulent type 3 strain P3/Leon/37 and the attenuated vaccine strain P3/Leon 12 a₁b. Between the vaccine strain and the neurovirulent revertant P3/119, seven nucleotide sequence changes were seen, which resulted in three predicted amino acid substitutions. Since, with the exception of one of the non-coding changes, the mutations observed in P3/119 were not direct reversions of the mutations observed between P3/Leon/37 and P3/Leon 12 a₁b, it has not yet been possible to definitively identify the changes responsible for the attenuation of and reversion to neurovirulence in poliovirus type 3.

In view of the number of replicative cycles since the divergence of these three strains and of current ideas on the relative genetic instability of RNA virus genomes (Holland et al., 1982), the number of mutations observed is remarkably small. This small number of changes is at first perhaps surprising when the strikingly different pathogenic
capacities of the three strains are considered. However, the attenuation of virus neurovirulence by very limited nucleotide sequence changes is not without precedence, as a single amino acid substitution in the surface glycoprotein has been shown to be responsible for the attenuation of neurovirulence in rabies virus (Dietzschold et al., 1983). The mutations seen between the neurovirulent and attenuated strains in this study are considerably fewer in number than the 57 changes observed between the neurovirulent and attenuated type 1 strains P1/Mahoney and P1/Sabin by Nomoto et al. (1982b). Possible reasons for this are, firstly, slight differences in the derivation of the type 1 and type 3 vaccine strains (see, Sabin and Boulger, 1973; chapter 1) and secondly, the careful selection of three closely related strains with known passage histories for this study (chapter 2).

Some tentative conclusions may be drawn from the results presented in this chapter. The silent nucleotide substitution at position 472 may be responsible for both the attenuation of neurovirulence in P3/Leon 12 a,b and reversion to a neurovirulent phenotype in P3/119. However, it is unlikely that such major biological changes result from such a minor biochemical transition unless this region were playing a critical role in some function of, for example, virus replication. A more likely and less controversial hypothesis is that the serine-phenylalanine mutation in the VP3 protein of strain P3/Leon 12 a,b is, if not solely responsible for, at least heavily implicated in attenuation. Similarly, the valine-alanine change in the VP1 protein of P3/119 is implicated in the reversion of this strain to a neurovirulent phenotype. Definitive proof of these hypotheses requires the construction of in vitro recombinants between neurovirulent and attenuated polioviruses whose biological properties can be examined in vitro as well as in vivo. This is discussed in the next chapter.
CHAPTER 5

GENERAL SUMMARY
The major objective of this study was to identify the nucleotide sequence differences which account for the neurovirulent or attenuated phenotypes of three closely related strains of poliovirus type 3. To achieve this, an efficient RNA-cDNA hybrid cloning method was devised. Although this was not the first report of hybrid cloning, previous unfavourable comments on its efficiency (Wood and Lee, 1976; Zain et al., 1979) and suggestions that the method gives rise to cloning artefacts (Okayama and Berg, 1982) have almost certainly discouraged its widespread use. Thorough investigation of each of the manipulative steps involved has shown that these problems can be overcome (Cann et al., 1983). Comparable in efficiency to the more commonly used double-stranded cDNA cloning technique, the hybrid method has the advantages of experimental simplicity and that cDNA clones corresponding to the entire virus genome can be obtained from a single experiment. The method has proved to be ideally suited to the molecular cloning of picornavirus genomes. It is possible that the transformation efficiency of RNA-cDNA hybrids could be further increased by treatment with E. coli DNA ligase and DNA polymerase I before ligation, thus carrying out repair of the hybrid molecule in vitro, as in the method of Okayama and Berg (1982). This modification has not yet been tested.

Together with the work of Dr. G. Stanway on the neurovirulent strains P3/Leon/37 and P3/119, many hundreds of cDNA clones were examined and more than 22 kbp of nucleotide sequence determined. These experiments resulted in the identification of a small number of mutations in the genomes of the strains studied which must be responsible for their differences in neurovirulence. However, it has not yet been possible to identify the individual mutations involved in attenuation and reversion and further experiments are currently in progress. These experiments represent a number of different approaches. Firstly, the sequences of other
neurovirulent vaccine revertants are being determined, to ascertain whether the mutations observed in P3/119 are shared by other strains. Secondly, the work of Racaniello and Baltimore (1981a) has demonstrated that the construction of recombinant virus genomes \textit{in vitro}, at the level of cloned cDNA, is possible. Transfection of susceptible cells with these recombinant genomes gives rise to new, viable viruses with a defined set of mutations which can then be examined phenotypically. Initial experiments with recombinants between the neurovirulent strain P3/Leon/37 and the vaccine strain P3/Leon 12 a₁,b and also between the vaccine strain and the neurovirulent revertant P3/119 are in progress. Final proof that the mutations involved in attenuation and reversion have been accurately identified could be provided by the construction of an attenuated strain by recombination between P3/Leon/37 and P3/119. Parallel studies on neurovirulent and attenuated type 1 strains should help to explain the different stabilities of the type 1 and type 3 vaccines (see chapter 1) and may suggest how the type 3 vaccine can be modified to improve its stability. Finally, it is hoped to use site-directed mutagenesis of cloned virus genomes \textit{in vitro} to produce strains with specific biological properties.

Although the main aim of the work presented here has been the investigation of the molecular basis of attenuation in poliovirus, the information obtained has wider significance. The complete nucleotide sequence of P3/Leon 12 a₁,b was the first to be determined from a type 3 poliovirus. This has been compared with that of type 1 and the extent of the molecular homology between the strains demonstrated (Stanway \textit{et al}, 1983a). As part of a larger study based on the analysis of monoclonal antibody resistant mutants, the major neutralising antibody binding site of poliovirus type 3 has been identified (Minor \textit{et al}, 1983). The nucleotide sequence information obtained is also being used currently in the design of synthetic antigenic peptides, a development which may hold many advantages for the prevention and perhaps treatment of poliomyelitis.
and other related picornavirus infections.

At the outset of the work described in this dissertation, it was difficult to envisage that improved alternatives to the Sabin vaccines would ever be a realistic proposition. It now seems that the immediate future holds just such a prospect.
CHAPTER 6

MATERIALS AND METHODS
6. MATERIALS

6.A.1 - Reagents

Specialised reagents were obtained from the following sources:

Amersham International, Amersham, UK.
BDH - British Drug Houses Ltd., Poole, UK.
Boehringer-Mannheim, Lewis, UK.
BRL - Bethesda Research Laboratories, Cambridge, UK.
CBL - Cambridge Biotechnology Laboratories, Cambridge, UK.
Difco Laboratories, Detroit, Michigan, USA.
Fisons, Loughborough, UK.
NEN - New England Nuclear, Boston, Massachusetts, USA.
Pharmacia Fine Chemicals, Uppsala, Sweden.
PL Biochemicals, Milwaukee, Wisconsin, USA.
S & S - Schleicher & Schull Ltd., Dassel, W. Germany.
Serva Feinbiochemica, Heidelberg, GDR.
Sigma Chemical Company, Poole, UK.
Uniscience, Cambridge, UK.

Acrylamide (2x crystallised, Analytical Grade) Serva, Uniscience.
Agar (Difco Bacto Agar) Difco.
Ammonium persulphate (AnalR) BDH.
Antibiotics Sigma.
(ampicillin, chloramphenicol, tetracycline)
DL-dithiothreitol (DTT) Sigma.
Ficoll (Type 400 - Mr 400,000) Sigma.
Formamide (AnalR) BDH.
8-hydroxyquinoline Sigma.
N,N'-methylene bis acrylamide  Uniscience.
Nitrocellulose membrane filters  S&S.
Phenol  Fisons.
Polyethylene glycol (PEG) (Mr 8000)  Sigma.
Polyvinyl pyrrolidone (technical grade)  Sigma.
Sodium dodecyl sulphate (SDS) (Specially pure)  BDH.
Tris-(hydroxymethyl) aminomethane acetate  Sigma.
(Tris acetate)
Tris-(hydroxymethyl) methylamine (Tris)  Fisons.
Urea  Fisons.

Radioisotopes
All radioisotopes used were obtained from Amersham International:–

[$\alpha-^{32}\text{P}]d\text{ATP}$ and [$\alpha-^{32}\text{P}]d\text{CTP}$ (2000-3000 Ci/mmol)
[$\gamma-^{32}\text{P}]\text{ATP}$ (3000 Ci/mmol)
[$\alpha-^{35}\text{S}]d\text{ATP}$ (at least 400 Ci/mmol)
[5-$^{3}\text{H}]d\text{CTP}$ (18-30 Ci/mmol)
[8-$^{3}\text{H}]d\text{GTP}$ (5-20 Ci/mmol)

Enzymes
T4 DNA ligase  NEN.
E. coli DNA polymerase I  Boehringer, BRL.
Large fragment E. coli DNA polymerase I  Boehringer, BRL, CBL.
(Klenow fragment)
Nuclease S1  Sigma.
T4 polynucleotide kinase  PL.
Proteinase K  Sigma.
Terminal deoxynucleotidyl transferase  PL.
Restriction endonucleases (various)  BRL.
Reverse transcriptase was a gift from Dr J.W. Beard, NIH, Bethesda, USA.
6.A.2 - Solutions, Buffers and Growth Media

5 x Denhardt's Solution

0.1% (w/v) BSA, 0.1% (w/v) ficoll, 0.1% (w/v) poly-vinylpyrrolidone in 6 x SSC. Filter sterilised and kept at 4°C.

EDTA

0.5M EDTA in distilled water, adjusted to pH 8.0 with NaOH.

Ethidium bromide (EtBr)

5 mg/ml in TE.

Saline citrate (SSC)

0.15M NaCl, 0.15M sodium citrate in distilled water.

Tris-EDTA (TE)

10mM Tris-HCl pH 7.4, 1mM EDTA in distilled deionised water.

Alkaline agarose buffer

30mM NaOH, 2mM EDTA in distilled water.

Alkaline agarose sample buffer

50% (v/v) glycerol, 60mM NaOH, 4mM EDTA, 0.01% (w/v) bromophenol blue in distilled water.

Elfo buffer

40mM Tris, 1mM EDTA in distilled water, adjusted to pH 7.7 with glacial acetic acid.

Elfo sample buffer

100mg agarose, 5g glycerol, 50ml 10mM Tris-acetate, 2ml 0.5M EDTA, 0.01% (w/v) bromophenol blue. Autoclaved 15min at 15lb/in², cooled and expelled from syringe to form agarose "beads".

Phage buffer

7g/l Na₂HPO₄, 3g/l KH₂PO₄, 5g/l NaCl, 1mM MgSO₄, 0.1mM CaCl₂ in distilled water. Autoclaved 15min at 15lb/in². 0.001% (v/v) gelatin added as a 1% (w/v) sterile solution.

TBE1 buffer

125mM Tris, 40mM boric acid, 2.5mM EDTA in distilled water.

TBE2 buffer

90mM Tris, 70mM boric acid, 2.5mM EDTA in distilled water.
Luria broth

10g/l Difco Bacto peptone, 5g/l Difco Bacto yeast extract, 5g/l NaCl in distilled water.

L-tet

Luria broth containing 10µg/ml tetracycline.

L-amp

Luria broth containing 100µg/ml ampicillin.

All agar plates contained 1.5% (w/v) Difco Bacto agar

Soft top agars contained 0.7% (w/v) Difco Bacto agar

Minimal agar

Agar - 3% (w/v) Difco Bacto agar in distilled water.

Salt solution - 10.5g/l K₂HPO₄, 4.5g/l KH₂PO₄, 1g/l (NH₄)₂SO₄, 0.5g/l sodium citrate, 0.2g/l MgSO₄ in distilled water. Solutions autoclaved separately, then 100ml of agar mixed with 100ml of salt solution. When cool, 1µl thiamine-HCl (1mg/ml) and 2ml of glucose solution, (200g/l) were added before pouring.

2 x TY broth

16g/l Difco Bacto peptone, 10g/l Difco Bacto yeast extract, 5g/l NaCl in distilled water.
6. B METHODS

Preparation of Phage $\lambda$ DNA

DNA size markers were prepared by digesting phage DNA with the restriction endonucleases EcoRI and HindIII. An inoculum of $\lambda^+ c$ was obtained from Professor W. J. Brammar. This phage is a clear plaque mutant of wild-type $\lambda$ and has the same restriction endonuclease map as the wild-type phage (Meselson, 1964). 1 l of Luria broth containing 10 mM MgCl$_2$ was inoculated with a 1:20 dilution of an overnight culture of E. coli C600 and grown with vigorous aeration at 37°C. When the culture reached a density of approximately $2 \times 10^8$ cells/ml (OD$_{650}$ 0.5), $\lambda^+ c$ was added to a multiplicity of infection of 0.5 and incubation continued until complete lysis of the culture occurred (2–3 h after infection). 1 ml of chloroform was added and incubation continued for 10 min to complete lysis. The medium was then clarified by centrifugation for 10 min at 16,000 x g. Phage were precipitated by the addition of NaCl to 40 µg/ml and PEG to 100 µg/ml, followed by incubation at 4°C for 18–24 h and were harvested by centrifugation for 10 min at 16,000 x g. The phage pellets were resuspended gently in 40 ml of cold phage buffer and incubated with 10 µg/ml DNase I and 10 µg/ml RNase A/T1 at 37°C for 30 min. The phage were purified on CsCl step gradients as follows. Step gradients were prepared in 14 ml nitrocellulose tubes and consisted of 1 ml 1.7 g/ml CsCl, 2 ml 1.5 g/ml CsCl and 3 ml 1.3 g/ml CsCl in phage buffer, onto which the phage suspension was carefully layered. The gradients were centrifuged for 2 h at 75,000 x g in a Sorvall AH627 rotor and the phage bands collected by side puncture of the tubes. DNA was obtained by phenol extraction of the purified phage and stored at 4°C in TE.

End Labelling of DNA

Restriction endonuclease fragments of DNA with protruding 5' termini were labelled using the "fill-in" reaction of the Klenow fragment of E. coli.
DNA polymerase I (Klenow et al., 1971). The restriction endonuclease digested DNA (1–5 pmoles of ends) was added to a mixture containing Hin buffer (20 mM Tris-HCl pH 7.4, 7 mM MgCl₂, 60 mM NaCl), 0.5 mM dGTP, dCTP and dTTP, 1–5 μCi of [α-³²P]dATP and 2 u of enzyme. After incubation for 30 min at room temperature, the reaction was stopped by heating to 65°C for 10 min.

Restriction endonuclease fragments with flush ends or protruding 5′ termini were labelled using T4 polynucleotide kinase, as follows. 5′ phosphate groups were removed with bacterial alkaline phosphatase or calf intestinal phosphatase. Restriction endonuclease digested DNA (about 20 pmoles ends) was added to a mixture containing 10 mM Tris-HCl pH 8.0, 10 mM MgCl₂ and 1 u of enzyme, incubated at 37°C for 30 min and phenol extracted to stop the reaction. The DNA was ethanol precipitated, washed in 70% ethanol and redissolved in a mixture containing 70 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 5 mM DTT, 1–5 μCi of [γ-³²P]ATP and 5 u of T4 polynucleotide kinase. After incubation for 30 min at 37°C, the reaction was stopped by heating to 65°C for 10 min.

**Extraction and Purification of Virus RNA**

Polioviruses, grown in Hep2c cells, were purified on 15–45% sucrose gradients (Minor, 1980) and were supplied by Dr. P. D. Minor as gradient fractions transported at -70°C. RNA was extracted directly from the sucrose fraction by the addition of SDS to 0.5% (w/v) and three extractions with an equal volume of phenol/chloroform/isoamylalcohol/8-hydroxyquinoline (100v/100v/4v/0.8w). Alternatively, RNA was extracted by the addition of SDS to 0.2% (w/v) and proteinase K to 20 μg/ml. After incubation at 37°C for 30 min, three phenol extractions were performed as above (Cann et al., 1983). This second method was employed in most later experiments. Following phenol extraction, the RNA was ethanol precipitated, collected by centrifugation for 10 min at 9,000 x g in sterile, siliconised "Corex" tubes (DuPont Ltd) and redissolved in sterile
distilled deionised water at a concentration of 1 mg/ml. 0.1-1 µg of RNA was analysed by electrophoresis on a 1% (w/v) agarose gel in Elfo buffer.

**Reverse Transcription**

Conditions used to synthesise cDNA were similar to those published by other workers (Efstratiadis et al., 1976; Kacian and Myers, 1976a; Buell et al., 1978; Retzel et al., 1980). Standard reverse transcription reactions contained 50 µg/ml RNA, 50-100 mM NaCl, 50 mM Tris-HCl pH 8.3, 8 mM MgCl₂, 20 mM 2-mercaptoethanol, 0.5 mM dCTP dGTP and dTTP, 0.1 mM dATP, 100 µCi/ml [α-32P]dATP, 5 µg/ml p(dT)₁₇ and 250-500 u/ml AMV reverse transcriptase (Houts et al., 1979). Reverse transcription reactions used to synthesise cDNA for cloning experiments contained all four deoxynucleotide triphosphates at 0.5 mM, 100 mM NaCl and 25 µCi/ml [α-32P]dATP. After incubation for 1 h at 37°C, the reaction was stopped by the addition of EDTA to 20 mM and heating to 65°C for 10 min. The yield of cDNA was estimated by measuring the incorporation of [α-32P]dATP into perchloric acid precipitable material by the following method. 2 µl of the reaction mixture were added to 0.25 ml of 0.2 mg/ml high molecular weight calf thymus DNA solution. 0.5 ml 1 M perchloric acid was added and the mixture left on ice for 10 min. The sample was then washed onto a 2.5 cm Whatman GFC filter with 10-20 ml of cold 5% trichloracetic acid and washed with 10 ml of industrial methylated spirits. Filters were then dried and placed in a scintillation vial with 3 ml scintillation fluid and counted in a Hewlett Packard "Prias" scintillation counter. The approximate yield of cDNA was calculated using the following formula:

\[
\text{Yield of cDNA (µg/ml)} = \frac{\text{Weight of dATP input} \times 4 \times \% \text{ total cpm incorporated}}{320/340 \times \text{weight of RNA input}}
\]

**Homopolymer Addition**

RNA-cDNA hybrids were cloned directly into the PstI site of a plasmid.
vector by a homopolymer annealing method (Otsuka, 1982). Conditions used for "tailing" RNA-cDNA hybrids were essentially the same as those used for the addition of homopolymer tails to double-stranded DNA (Roychoudhury et al., 1976; Deng and Wu, 1981). After incubation, reverse transcription mixtures were phenol extracted and passed through a small Sephadex G100 column formed in a pasteur pipette. cDNA eluted in the void volume of the column and was recovered from the appropriate fractions. RNA-cDNA hybrids at a concentration of 10-20 μg/ml were tailed in a reaction mixture containing 140 mM potassium cacodylate pH 6.9, 1 mM CoCl₂, 0.1 mM DTT, 20 nmoles/ml dCTP, 20-40 pCi of [α-32P]dCTP or [5-3H]dCTP and 200-500 u/ml calf-thymus terminal deoxynucleotidyl transferase. After 15 min at 37°C, the reaction was placed on ice and the incorporation of radioactivity determined. The approximate number of residues added per end was calculated using the following formula:

\[ \text{pmol dNTP in sample} \times \% \text{total cpm incorporated} \]

\[ \text{pmol ends in sample} \]

If less than 15-20 dC residues/3' end had been added, incubation was continued.

PstI cut pAT 153 (Twigg and Sherratt, 1980) was tailed with dGTP under similar conditions. Linear plasmid DNA (80 pmole ends/ml) was tailed in a reaction mixture containing 140 mM potassium cacodylate pH 6.9, 4 mM MgCl₂, 0.1 mM DTT, 32 nM [8-3H]dGTP (specific activity 0.37 Ci/mmol) and 250 u/ml terminal transferase. The reaction mixture was incubated at 37°C for 30 min when incorporation of radioactivity had normally stopped and approximately 20 dG residues/3' end had been added.

Tailed plasmid and RNA-cDNA hybrid were passed through Sephadex G100 columns, ethanol precipitated together at a ratio of 1 plasmid molecule to 1 RNA-cDNA hybrid molecule (based on an estimate of the average size of the cDNA produced in the reverse transcription reaction) and redissolved.
in sterile distilled deionised water at a concentration of 1 μg/ml plasmid DNA. The mixture was then adjusted to contain 100 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.2 mM EDTA and sealed in a 1.5 ml Eppendorf tube. After heating to 65°C for 10 min and 45°C for 2 h the mixture was allowed to cool slowly to room temperature in a covered waterbath (over 24 h) and subsequently kept on ice.

Preparation of Competent Cells

*E. coli* strains JA221 (hsdM⁺, hsdR⁻, lacY⁻, leuB6⁻, trpES, recA1/F) (Clarke and Carbon, 1978) and JA226 (rk⁻, mk⁺, lop11, thi⁻, leu⁻, strB, RecBC) (Dr. J. Windass, personal communication) were made competent for transfection with DNA by the following procedure. Luria broth (30-50 ml) was inoculated with a 1:20 dilution of an overnight culture, or with a single colony directly from an agar plate and grown with vigorous aeration at 37°C to a density of approximately 2 x 10⁸ cells/ml (OD₆₅₀ 0.5). The culture was then chilled on ice for 10 min and the cells pelleted by centrifugation in disposable plastic universal tubes, then resuspended in half the original volume of cold 100 mM CaCl₂ (a sterile solution in distilled, deionised water). After 20 min on ice, the cells were pelleted again and gently resuspended in 1/20th the original volume of cold CaCl₂. Cells were left on ice for at least 1 h before use.

Transformation of *E. coli*

Transformations of all strains of *E. coli* with plasmid DNA were performed as follows. Up to 10 μl of the annealed mixture was added to 50 μl of 1 x SSC and 0.1 ml of competent cells in 1.5 ml Eppendorf tubes or in disposable plastic universal bottles. After 30 min on ice, the cells were heated to 42°C for 2 min, then returned to the ice for 20 min. 1 ml of Luria broth was added and the tubes incubated for 1 h at 37°C to allow phenotypic expression to occur, after which time the cells were pelleted by centrifugation for 20 sec at 15,000 x g in an Eppendorf microfuge,
resuspended in 0.1 ml of Luria broth and spread onto the surface of an L-tet plate. Plates were incubated in an inverted position for 18-24 h.

**Colony Screening**

Initial screening was performed by testing for insertional inactivation of the ampicillin resistance gene of pAT 153. Tetracycline resistant colonies were streaked in grid formation onto two agar plates, L-amp and L-tet, using sterile wooden toothpicks. Colonies with a \( \text{tet}^r \text{ amp}^s \) phenotype were grown on nitrocellulose filters to be screened by hybridisation, using a modification of the method of Grunstein and Hogness (1975). The colonies were prepared for hybridisation by lysis and denaturation in 0.5 M NaOH (5 min), neutralisation in 0.5 M Tris-HCl pH 7.4 (2 x 2 min) and in 1.5 M NaCl, 0.5 M Tris-HCl pH 7.4 (4 min). After drying, the filters were baked in vacuo at 80°C for 2 h, then washed at 65°C for 2 h in 5 x Denhardt's solution. Hybridisations were performed in 9 cm petri dishes in sealed plastic boxes. Up to 4 filters were placed in a single dish containing 10-20 ml 5 x Denhardt's solution with a radiolabelled hybridisation probe. Approximately 1 x 10^6 cpm were used per filter and the specific activity of the radiolabelled probes varied between 0.1-5 pCi/\( \mu \)gDNA (see below). Slight variations in the specific activity of the hybridisation probe did not significantly affect the method (Grunstein and Wallis, 1979). Hybridisation was allowed to proceed at 65°C for 18-24 h. Filters were then washed at 65°C for a period of about 3 h in three 10-15 ml changes of 3 x SSC, 0.1% SDS (w/v). After drying, filters were covered with Saran-wrap and subjected to autoradiography.

Several different radiolabelled hybridisation probes were used and these were prepared as follows:

1) **High molecular weight cDNA probe**

Initial experiments on relatively small numbers of \( \text{tet}^r \text{ amp}^s \) colonies were performed using high molecular weight cDNA synthesised in a standard reverse transcription reaction mixture (see "Reverse Transcription").
These experiments were performed in order to obtain some initial information on the transformed colonies and to perfect the hybridisation procedure. More detailed information about the transformed colonies was obtained using the hybridisation probes described below.

ii) 3' enriched probe

In order to identify cDNA clones from the 3' terminus of the genome, a 3' enriched probe was synthesised from a reverse transcription reaction as described earlier, except that [α-32P]dATP at a specific activity of 2000-3000 Ci/mmol was used.

iii) Randomly primed probe

To isolate clones from other parts of the genome, hybridisation probes were prepared as above, but using low molecular weight salmon sperm DNA as a primer. This was prepared by the method of Taylor et al (1976). High molecular weight salmon sperm DNA at a concentration of 0.5 mg/ml was degraded to oligonucleotides by incubation in a reaction mixture containing 10 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 70 μg/ml DNase I at 37°C for 2 h. The Eppendorf tube was autoclaved briefly in a universal bottle to terminate the reaction and ensure the destruction of all the nuclease. The DNA was then ethanol precipitated, redissolved at a concentration of 10 mg/ml in sterile distilled deionised water and used at a concentration of 0.5 mg/ml, in place of p(dT)₁₇, to prime the reverse transcription reaction.

iv) Nick translated restriction endonuclease fragments

A restriction endonuclease fragment from a previously characterised clone was isolated by electrophoresis in a 1% (w/v) agarose gel (see "Isolation of DNA From Agarose Gels") and redissolved in sterile distilled deionised water at a concentration of 20-100 μg/ml. The DNA was boiled for 1 min to denature it and quenched on ice prior to the addition of the other reaction components. This step was found to increase the total incorporation of radiolabel in the reaction by up to 50%. Nick translation was carried out by a modification of the procedure of Rigby et al (1977)
in a reaction mixture containing 20-100 μg/ml DNA, HiPi buffer (20 mM Tris-HCl pH 7.4, 7 mM MgCl₂, 60 mM NaCl), 0.5 mM dCTP, dGTP and dTTP, 0.1 mM dATP, 200 μCi/ml [α-³²P]dATP, 10 ng/ml DNase I and 20-100 u/ml E. coli DNA polymerase I. After incubation for 30 min at room temperature, the reaction was stopped by phenol extraction. [α-³⁵S]dATP was later successfully used, under identical conditions, in place of [α-³²P]dATP.

All hybridisation probes were passed through Sephadex G100 columns before use to remove unincorporated label and very low molecular weight labelled material, which otherwise resulted in non-specific hybridisation. The appropriate column fractions were boiled for 10 min to denature the probe before being added directly to the hybridisation dishes (pre-warmed to 65°C).

Isolation of DNA From Agarose Gels

DNA fragments were excised from agarose gels containing Elfo buffer and 0.5 μg/ml EtBr, using a sterile scalpel blade. The gel slice was placed in a sealed length of dialysis tubing containing 0.1 x Elfo buffer and electrophoresis carried out in 0.1 x Elfo buffer until the DNA migrated from the agarose onto the wall of the dialysis membrane. This process was monitored by brief periods of UV illumination, so as to minimise the risk of UV induced damage to the DNA. The electrical potential across the gel tank was then reversed for 30 sec to free the DNA from the dialysis membrane and the buffer containing the DNA removed. The dialysis tubing was washed out 2-3 times with sterile distilled deionised water and the pooled washings and buffer extracted 3-4 times with an equal volume of isobutanol in order to remove EtBr bound to the DNA and to concentrate the DNA by volume reduction. The DNA solution was then extracted 3-4 times with an equal volume of phenol/ chloroform/ isoamylalcohol/ 8-hydroxy-quinoline and twice with an equal volume of chloroform. Following ethanol precipitation, the DNA was redissolved in sterile distilled deionised water at an appropriate concentration for whatever purpose it was
required.

**Plasmid Isolation Methods**

Plasmids from colonies which showed a positive hybridisation response to poliovirus specific probes were isolated using two alternative rapid methods.

**Method 1** - This method was based on that of Holmes and Quigley (1981). 1 ml cultures of positively hybridising clones were grown in 1.5 ml Eppendorf tubes at 37°C for 12-18 h, in L-tet broth. The cells were pelleted by centrifugation for 20 sec at 15,000 x g in an Eppendorf microfuge and resuspended in 25 µl of STET buffer (8% (w/v) sucrose, 5% (w/v) triton X100, 50 mM EDTA, 50 mM Tris-HCl pH 8.0). 2 µl of a 10 mg/ml solution of lysozyme was added and after boiling for 40 sec, the tubes were centrifuged for 10 min in an Eppendorf microfuge. The pellets were discarded and 25 µl of isopropanol was added to the supernatants. The nucleic acids were then precipitated by placing the tubes in an ethanol-dry ice bath at -70°C for 10 min and centrifugation for a further 10 min. The nucleic acid pellet was redissolved in 25 µl of 0.3 M sodium acetate pH 6.5 and reprecipitated with 75 µl of ethanol, washed with 70% ethanol and redissolved in 20 µl of sterile distilled deionised water. Each preparation contained enough plasmid DNA for at least 2 restriction endonuclease digestions, the products of which were analysed on agarose gels containing 0.5 µg/ml EtBr.

**Method 2** - Plasmids with larger inserts were then examined using the following isolation method. The cell pellet from a 10 ml culture, grown overnight at 37°C in L-tet broth, was resuspended in 0.7 ml of 50 mM Tris-HCl pH 8.0, 50 mM EDTA, 15% (w/v) sucrose. Lysozyme was added to 1 mg/ml and the cell suspension incubated for 10 min at room temperature. 30 µl of 10% (w/v) SDS and 60 µl of 5 M potassium acetate pH 6.0 were then added and after incubation on ice for 30 min, the tubes were centrifuged for 15 min in an Eppendorf microfuge. The pellets were discarded and 1 µl
of a 2 mg/ml solution of RNase A/T1 was added to each of the supernatants. After incubation for 30 min at 37°C, the mixture was extracted once with an equal volume of phenol/chloroform (1v/1v), the DNA ethanol precipitated from the aqueous phase and redissolved in 0.1 ml of 0.3 M sodium acetate. Following two more ethanol precipitations, the DNA was washed with 70% ethanol and redissolved in 50 μl of sterile distilled deionised water.

**Large scale method** - Plasmids with large cDNA inserts (e.g. 2 kbp) were isolated in pure form by large scale CsCl-EtBr preparations using the following Triton lysis method. A 500 ml culture in L-tet broth was grown at 37°C to late exponential phase (OD$_{650}$ 0.75) and 150 μg/ml chloramphenicol added. Incubation was continued for 18-24 h to allow plasmid amplification to occur, then the cells were harvested by centrifugation, washed once in TE buffer and resuspended in 24 ml 25% (w/v) sucrose, 50 mM Tris-HCl pH 8.0. The cells were held on ice and 6 ml of 10 mg/ml lysozyme solution and 6 ml of 0.5M EDTA was added whilst stirring, followed by 40 ml of Triton mixture (50 mM Tris-HCl pH 8.0, 50 mM EDTA, 1% (w/v) Triton X100). When lysis had occurred (indicated by the increased viscosity of the suspension), the cell debris was pelleted by centrifugation for 45 min at 48,000 x g. 8 ml of 10% (w/v) SDS was added to the clear supernatant which was then extracted with 20 ml of phenol/chloroform (1v/1v). The nucleic acids were precipitated from the aqueous phase with ethanol and redissolved in 10 ml of 10 mM Tris-HCl pH 8.0. 50 μl of a 20 mg/ml solution of RNase A/T1 was added, the mixture incubated for 30 min at 37°C and then extracted with 10 ml of phenol/chloroform. The DNA was ethanol precipitated from the aqueous phase and redissolved in a total weight of 20g of TE buffer. 4 ml of a 5 mg/ml solution of EtBr and 23.76g of CsCl were added to bring the density of the solution to 1.393-1.394 g/cm$^3$. Gradients were formed by centrifugation in 36 ml polyallomer tubes in a Sorvall TV850 titanium vertical-bucket rotor at 128,000 x g for 18 h at a temperature of 15°C. The lower plasmid DNA band was collected with a 10 ml syringe by side-puncture of the tubes.
EtdBr was removed by repeated extractions of the DNA solution with equal volumes of CsCl/water saturated isopropanol, followed by two extractions with an equal volume of diethyl ether. The DNA was then dialysed against 500 ml of sterile distilled water with several changes over a period of at least 3-4 h, ethanol precipitated and redissolved in sterile distilled deionised water at a concentration of 1 mg/ml.

Restriction Endonuclease Digestions

All restriction endonucleases were used initially under the conditions specified by the suppliers, as follows:

- AluI - 6mM Tris-HCl pH 7.6, 6mM MgCl₂, 6mM 2-mercaptoethanol, 50mM NaCl.
- BamHI - 20mM Tris-HCl pH 7.0, 7mM MgCl₂, 2mM 2-mercaptoethanol, 100mM NaCl.
- ClaI - 10mM Tris HCl pH 8.0, 10mM MgCl₂, 100 μg/ml BSA.
- EcoRI - 100mM Tris-HCl pH 7.2, 5mM MgCl₂, 2mM 2-mercaptoethanol, 50mM NaCl.
- HaeIII - 50mM Tris-HCl pH 7.5, 5mM MgCl₂, 0.5mM DTT.
- HindIII - 20mM Tris-HCl pH 7.4, 7mM MgCl₂, 60mM NaCl.
- PstI - 20mM Tris-HCl pH 7.5, 10mM MgCl₂, 50mM (NH₄)₂SO₄, 100 μg/ml BSA.
- Sau3A - 6mM Tris-HCl pH 7.5, 6mM MgCl₂, 50mM NaCl, 50 μg/ml BSA.
- SnaI - 15mM Tris-HCl pH 8.0, 6mM MgCl₂, 15mM KCl.
- Sphi - 6mM Tris-HCl pH 7.5, 6mM MgCl₂, 6mM 2-mercaptoethanol, 50mM NaCl.
- XbaI - 6mM Tris-HCl pH 7.4, 6mM MgCl₂, 100mM NaCl.
- XhoI - 8mM Tris-HCl pH 7.4, 6mM MgCl₂, 6mM 2-mercaptoethanol, 150mM NaCl.

(For further details of these enzymes, see Roberts, (1980)).

However, since the reaction conditions for most of these enzymes are similar, most were later used in either "666" buffer (6 mM Tris-HCl pH 7.4, 6 mM MgCl₂, 6 mM 2-mercaptoethanol) adjusted to the correct concentration of NaCl, or in "core" buffer (50 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 50 mM NaCl) (BRL Ltd). All restriction endonuclease digestions were performed at 37°C.
i) **Restriction endonucleases with hexameric sites** — Cleavage positions of restriction endonucleases with hexameric recognition sites were mapped by analysis of single and double digestion products on 1% (w/v) agarose gels in Elfo buffer and containing 0.5 µg/ml EtBr. 1-2 µg of DNA, depending on the number of fragments generated, was loaded into each slot and the bands visualised by UV illumination. A photograph of the EtBr stained bands was taken and used to plot a graph of the mobility of known size marker fragments (λ+c DNA doubly digested with EcoRI and HindIII and linearised pAT 153 DNA were used as size markers, resulting in a visible size range from 565 bp to 21.5 kbp) from which the size of the unknown restriction fragments of the plasmid being mapped were determined.

ii) **Restriction endonucleases with tetrameric sites** — Cleavage positions of restriction endonucleases with tetrameric recognition sites were mapped using the method of Smith and Birnstiel (Smith and Birnstiel, 1976). Recombinant plasmids (1-5 µg) were cut using a restriction endonuclease with a unique site in the plasmid to be mapped (not in the cDNA insert), to produce fragments with 5' protruding termini e.g. using EcoRI, BamHI, HindIII or ClaI. The ends of the molecule were radiolabelled by the "fill-in" reaction of the Klenow fragment of E. coli DNA polymerase I (see, "End Labelling of DNA") to a specific activity of approximately 0.1-0.5 µCi/µg DNA. Asymmetrically labelled fragments were produced by a second restriction endonuclease cleavage at another unique site in the vector, leaving the radiolabelled end of the larger fragment of the vector molecule nearest the cDNA insert. The large end-labelled fragment was then partially digested with the restriction endonucleases to be mapped, in a 20 µl reaction mixture containing 0.5 µg high molecular weight calf thymus DNA and 0.5-1 µg radiolabelled plasmid. A series of samples forming a time course of the digestion were taken, e.g. at 1, 5, 15, 30 min and the reaction stopped by placing on ice, followed by a single phenol extraction. These samples were subjected to electrophoresis on a 1% (w/v)
agarose gel. The small radiolabelled fragment was not usually separated from the large before digestion, since, being vector-derived, this was of known size and served as an additional size marker to the end-labelled λ<sup>c</sup> EcoRI - HindIII digest. After electrophoresis, the gels were dried and subjected to autoradiography. Since the restriction endonuclease map of the vector was known, the map of the cDNA insert could be determined from the relative mobilities of the radiolabelled fragments.

Preparation of Double-Stranded M13 Replicative Form DNA

M13 mp8 and mp9 (Messing and Vieira, 1982) double-stranded replicative form (RF) DNA was prepared as follows. 1 ml of 2 x TY broth, seeded with a 1:100 dilution of an overnight culture of E. coli JM103 (Alac pro, thi<sup>−</sup>, strA, supE, endA, sbcB15, hsdR4, F' traD36, proAB, lac<sup>q</sup>, Z M15) (Messing et al., 1981), was infected with a single phage plaque and incubated with vigorous aeration at 37°C until the optical density of the culture reached OD<sub>650</sub> 0.5. This was then used to inoculate a JM103 seeded 5 ml 2 x TY broth culture, which when grown to the same optical density was used to inoculate a seeded 50 ml culture. This last "starter" culture was used to inoculate 500 ml of unseeded 2 x TY broth which was grown to stationary phase (18 h at 37°C). The cells were harvested from this final culture and double-stranded RF DNA was extracted and purified by the triton-lysis extraction method (see, "Plasmid Isolation Methods"), except that the RF DNA was subjected to a second CsCl-EtdBr centrifugation to ensure the complete removal of contaminating bacterial nucleic acids.

M13 Sub-cloning of DNA Fragments

cDNA inserts from recombinant plasmids were isolated from PstI digestions of the plasmids, subjected to electrophoresis on 1% (w/v) agarose gels (see, "Isolation of DNA From Agarose Gels") and redissolved at a concentration of 0.1-1 μg/ml. The DNA was redigested with a second restriction endonuclease, chosen with reference to the previously
determined restriction map, but most commonly AluI, HaeIII or Sau3A, although other enzymes were used when appropriate. The reactions were terminated by extraction with an equal volume of phenol/ chloroform/ isoamylalcohol/ 8-hydroxyquinoline to ensure the complete removal of the enzyme, followed by ethanol precipitation. The DNA was then redissolved at approximately 1 μg/ml in sterile distilled deionised water. cDNA restriction endonuclease fragments were ligated into appropriately digested M13 vectors, possessing complementary ends to the inserts, by the following procedure. Digested M13 RF DNA (1 μg/ml) and digested cDNA (0.1-1 μg/ml depending on the relative size of the restriction endonuclease fragment(s) concerned, a 1:2-1:10 fold molar ratio of cDNA:vector), were incubated at 14°C for 18-24 h in the presence of 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP.

Alternatively, random fragments generated by ultra-sonic treatment of the DNA, rather than restriction endonuclease digestion, were cloned by the following method. 50 μg of plasmid DNA was digested with PstI and subjected to electrophoresis on a 1% (w/v) agarose gel. The fragment corresponding to the cDNA insert was cut out, electro-eluted from the gel and redissolved in 50 μl of sterile distilled deionised water. The cohesive PstI termini were self-ligated in a 50 μl reaction mixture containing 35 μl of DNA, 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 1 mM ATP and 5 u T4 DNA ligase by incubation for 18 h at 14°C. Sonication was performed in 1.5 ml Eppendorf tubes using an ultrasonic waterbath (Kerry Ultrasonics Ltd, Hitchin, UK) for 2-3 min and the extent of ultrasonic shearing monitored by electrophoresis of small aliquots on an agarose gel. Sheared DNA ends were repaired using the "fill-in" reaction of the Klenow fragment of E. coli DNA polymerase I (see, "End Labeling of DNA"), without radiolabel and with each of the four deoxynucleotidetriphosphates at 50 μM. Repaired DNA was subjected to electrophoresis on a 1% (w/v) agarose gel and a size fraction of approximately 0.5-1.5 kbp cut out and electro-eluted as before. DNA recovered from the gel was
redissolved in 20 μl of sterile distilled deionised water. M13 mp8 digested with SmaI and treated with calf-intestinal phosphatase to prevent religation of vector molecules (see, "End Labelling of DNA"), was provided by Dr. G. Stanway. 0.5-2 μl of sonicated DNA were ligated with 1 ng of SmaI digested vector as described above.

**Transfection of E. coli**

M13 ligation mixtures (see, "M13 Sub-cloning of DNA Fragments") were used to transform competent JM103 cells (see, "Preparation of Competent Cells") using the following method. 0.1 ml of competent cells were added to each 10 μl ligation mixture and incubated on ice for 40 min, then heat shocked at 42°C for 2 min. The cells were plated out on well dried minimal agar plates in 3 ml of 2 x TY soft top agar containing 6 μg/ml BCIG, 5 μg/ml IPTG and 0.15 ml of a stationary phase culture of JM103. After allowing the top agar to solidify, the plates were incubated in an inverted position at 37°C for 18 h. Recombinant phage containing cDNA inserts formed turbid plaques ("whites") on incubation, easily distinguishable from the blue plaques formed by phage without inserts (Messing et al., 1981).

**Preparation of Single-Stranded Recombinant M13 DNA**

Single-stranded M13 DNA to be used as templates for nucleotide sequencing was prepared as follows. Recombinant plaques were harvested using sterile wooden toothpicks and used to infect 1 ml cultures in 2 x TY broth, seeded with a 1:100 dilution of a stationary phase culture of JM103. These cultures were grown at 37°C with vigorous aeration for 6 h and then the cells pelleted by centrifugation in 1.5 ml Eppendorf tubes at 15,000 x g for 5 min in an Eppendorf microfuge. The cell pellets were discarded and the phage precipitated from the supernatants by the addition of 0.2 ml of a solution of 10% (w/v) PEG, 2.5 M NaCl. After incubation for 30-60 min at room temperature, the phage were harvested by centrifugation for 15 min in
an Eppendorf microfuge. All the supernatant liquid was removed with a
drawn out capillary to minimise contamination of the phage pellet with
PEG. Tubes without phage pellets were discarded at this stage. The
remaining pellets were resuspended in 0.1 ml of 1.1 M sodium acetate pH
7.0. The phage suspension was vortexed with 50 µl of aqueous phenol, 50 µl
of chloroform/isoamylalcohol (50v/1v) was added and the tubes vortexed
again. They were then centrifuged for 1 min in an Eppendorf microfuge.
The aqueous phase was re-extracted with 0.1 ml of chloroform/
isoamylalcohol and the DNA precipitated by the addition of 0.25 ml of
ethanol and chilling the tubes for 15 min in an alcohol - dry-ice bath.
Following centrifugation for 10 min in a microfuge, the supernatants were
removed with a drawn out capillary, the pellets dried under vacuum and
redissolved in 20 µl of 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA.

Single-Tracking of Recombinant M13 Phage

Template DNA from "white" plaques was screened by a single-tracking
sequencing procedure using ddTTP to distinguish between "white" plaques
arising from phage deletion mutants and recombinant phage and also to
provide some initial information on the size, genomic location and
orientation of the cDNA inserts. 1 µl of template DNA was placed in a
0.4 ml Eppendorf tube with 1 µl of priming mixture, containing 0.2 ng
primer (17 bp "LMB2" primer, supplied by Dr. P. Meacock, ICI Joint
Laboratory, University of Leicester), 10 mM Tris-HCl pH 7.4, 10 mM MgCl₂,
50 mM NaCl. The tubes were sealed, placed in a boiling water bath for
5 min, then allowed to cool to room temperature for 15 min. 1 µl of the
primed template was placed in an open 1.5 ml Eppendorf tube with 1.5 µl of
the following mixture, containing 0.5 µl T₄ mixture (see table 6.1),
0.5 µl 0.8 mM ddTTP, 0.5 µl 1 mM DTT, 1 µCi [α-³²P]dATP, 0.2 u E. coli
dNA polymerase I Klenow fragment. After 15 min at room temperature, 0.5 µl
of 0.5 mM dATP was added and the reaction allowed to proceed for a further
15 min, when the tubes were either frozen at -20°C, or, if to be subjected
to electrophoresis immediately, 2 μl of formamide-dye mixture (containing 95% (v/v) formamide, 20 mM EDTA, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol) was added and the tubes placed in a boiling water bath for 5 min. After this, the samples were loaded onto 6% (w/v) polyacrylamide gels (40 cm x 20 cm x 0.37 mm) using a drawn-out capillary and subjected to electrophoresis at a constant 40 W (approximately 25 mA, 1600 V) for 2 h. These gels contained 8 M urea and TBE1 buffer. The gels were fixed by immersion in 10% (v/v) acetic acid, excess liquid removed, then the gel covered with Saran Wrap (Dow Chemicals, USA) and subjected to autoradiography for between 12 h and 3 days.

**TABLE 6.1**

<table>
<thead>
<tr>
<th>A&lt;sup&gt;*&lt;/sup&gt;</th>
<th>G&lt;sup&gt;*&lt;/sup&gt;</th>
<th>C&lt;sup&gt;*&lt;/sup&gt;</th>
<th>T&lt;sup&gt;*&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>μl</td>
<td>μl</td>
<td>μl</td>
<td>μl</td>
</tr>
<tr>
<td>0.5 mM dATP</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.5 mM dGTP</td>
<td>20</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>0.5 mM dCTP</td>
<td>20</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>0.5 mM dTTP</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>10 x TE&lt;sup&gt;1&lt;/sup&gt;</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>ddNTP</td>
<td>5 nM</td>
<td>0.2 mM</td>
<td>0.2 mM</td>
</tr>
</tbody>
</table>

<sup>1</sup>100 mM Tris-HCl pH 8.0, 1 mM EDTA

**DNA Sequencing**

Two methods were used to determine the nucleotide sequence of poliovirus specific cDNA's. These were similar in that they both employ sub-cloning of cDNA into phage M13 vectors and use the dideoxynucleotide sequencing method (Sanger et al, 1977), but differ in experimental detail and will be described separately.

**Method 1** - Nucleotide sequencing of recombinant M13 phage bearing restriction endonuclease fragments of poliovirus cDNA was carried out by a
similar procedure to the single tracking method, essentially using the method of Sangar et al. (1977). Priming was carried out as above in sealed 0.4 ml Eppendorf tubes containing 2 µl template DNA, 1 ng LMR2 primer, 10 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 50 mM NaCl. After priming, 5 µCi of [α-³²P]dATP was added to the primed template and four 1 µl aliquots placed in open 1.5 ml Eppendorf tubes. 1.5 µl of a mixture containing 0.5 µl A*/G*/C*/T* mixture (see table 6.1), 0.5 µl ddATP/ddGTP/ddCTP/ddTTP (table 6.1), 0.5 µl 1mM DTT, 0.2 u E. coli DNA polymerase I Klenow fragment was added to each tube. After 15 min at room temperature, 0.5 µl 0.5mM dATP was added and after a further 15 min, the tubes were frozen or formamide-dye mixture added as above. Electrophoresis and autoradiography were carried out as in the single-tracking procedure, except that gels were run for 2, 4 or 6 h with a change of running buffer each 2 h, depending on the length of sequence to be determined from each template.

**Method 2** - M13 subclones were made by the sonication of DNA to give randomly sheared fragments (Deininger, 1983). No single-track screening was performed since the phosphatase treatment of the vector prevented a high background of M13 deletion mutants (see "M13 Sub-cloning of DNA Fragments"). Nucleotide sequence of the recombinant phage was obtained by an updated dideoxynucleotide method (Biggin et al, 1983). Priming of template DNA was carried out as before in sealed 0.4 ml Eppendorf tubes, in a 5 µl mixture containing 2.5 µl of template DNA, 1 ng primer (17 bp "LMR2" primer, supplied by Dr. P. Meacock or 15 bp primer, BRL Ltd.), 10 mM Tris-HCl pH 8.5, 5 mM MgCl₂. After priming, 4 µl 3 mM DTT, 2 µCi [α-³⁵S]dATP and 0.4 u E. coli DNA polymerase I Klenow fragment were added. To start the reaction, 2 µl aliquots of the primed mixture were added to open 1.5 ml Eppendorf tubes containing 1 µl of A/G/C/T nucleotide mixture (see table 6.2). After 15 min at room temperature, 0.5 µl 0.5 mM dATP was added and after a further 15 min, the tubes were either frozen or prepared for electrophoresis as before. Electrophoresis was performed on 50 cm x 20 cm x 0.25 mm 6% acrylamide TBE2 gradient gels, containing 8 M
urea (Biggin et al., 1983), at a constant 40 W (approximately 20 mA, 2000 V) with 0.5 x TBE2 in the upper chamber and 1 x TBE2 in the lower. The gel was fixed in 10% (v/v) acetic acid when the bromophenol blue dye marker reached the bottom, transferred to Whatman 3MM paper and dried on a Bio Rad 1125B gel drier (Bio Rad Laboratories Ltd, Watford, UK), then subjected to autoradiography for 12 h - 7 days.

TABLE 6.2

COMPOSITION OF NUCLEOTIDE SEQUENCING REAGENTS (35S)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>µl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 mM dGTP</td>
<td>250</td>
<td>12.5</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>0.5 mM dCTP</td>
<td>250</td>
<td>250</td>
<td>12.5</td>
<td>250</td>
</tr>
<tr>
<td>0.5 mM dTTP</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>12.5</td>
</tr>
<tr>
<td>5 mM ddATP</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10 mM ddGTP</td>
<td>-</td>
<td>8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10 mM ddCTP</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>10 mM ddTTP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>25</td>
</tr>
<tr>
<td>TE 1</td>
<td>250</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

1 10 mM Tris-HCl pH 8.0, 1 mM EDTA

Data Handling

Nucleotide sequence data was entered and analysed on a PDP 11/44 computer using a package of programs obtained from R. Staden, MRC Laboratory of Molecular Biology, Cambridge, UK (Staden, 1980). A summary of the main programs from this package is given in chapter 4.
REFERENCES
REFERENCES


Assad, F. (1981) Re-assessment of the place of inactivated poliomyelitis vaccine in national immunization programs. In (Hennessen, W. & van


Black, D. N., Stephanson, P., Rowlands, D. J. & Brown, F. (1979) Sequence & location of the poly(C) tract in aphtho & cardiovirus RNA. Nucleic Acids Res. 6, 2381-90.

Detection by monoclonal antibodies of an antigenic determinant critical for poliovirus neutralization present on VP1 & on heat-inactivated virions. Virology 126, 707-10.


double-stranded RNA: Differential infectivity in enucleate cells. J. Virol. 27, 582-86.


Gauntt, C.J. (1980) Fragility of the rhinovirus type 14 genome to incubation at 60°C. Intervirology 13, 7-15.


Holand, J. J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S. & Vandepol, S.


J. Biochem. 22, 371-81.


classification of enteroviruses & ECHO virus type 34. Arch. Ges.
Virusforsch. 30, 89-92.

Rowlands, D.J., Harris, T.J.R. & Brown, F. (1978) More precise location of
the polycytidylic acid tract in foot-and-mouth disease virus RNA. J.
Virol. 26, 335-43.

Rowlands, D.J., Clarke, B.E., Carroll, A.R., Brown, F., Nicholson, B.H.,
antigenic variation in foot-and-mouth disease virus. Nature 306,
694-97.

Roychoudhury, R., Jay, E. & Wu, R. (1976) Terminal labelling & addition of
homopolymer tracts to duplex DNA fragments by terminal deoxy-
nucleotidyl transferase. Nucleic Acids Res. 3, 863-77.

Rueckert, R.R. (1965) Studies on the structure of viruses of the Columbia
SK group II. The protein subunits of ME virus, EMC virus & Mango

p. 131-213 In (Fraenkel-Conrat, H. & Wagner, R.R., eds.) Comprehensive

Rueckert, R.R., Palmemberg, A.C. & Pallansch, M. (1980) Evidence for a self-
cleaving precursor of virus coded protease, RNA replicase & VPg.
p. 263-75. In (Koch & Richter, eds.) Biosynthesis, Modification &
Processing of Cellular & Viral Polyproteins, Academic Press, New
York.


Sabin, A.B. (1961) Reproductive capacity of polioviruses of diverse origins
at various temperatures. Perspectives in Virology 2, 90-110.

872-76.

Rev. Inf. Dis. 3, 543-64.


Vogt, M., Dulbecco, R. & Wenner, H.A. (1957) Mutants of poliomyelitis virus with reduced efficiency of plating in acid medium & reduced...
neuropathogenicity. Virology 4, 141-55.


