Molluscum contagiosum virus: genome structure and mechanism of tumourigenesis

by

Colin David Porter

A thesis submitted for the Degree of Doctor of Philosophy (1989)

Charing Cross and Westminster Medical School,
University of London

For

Mum and Dad,

and for my wife,

Lorraine

ABSTRACT

Molluscum contagiosum virus (MCV) is an unclassified poxvirus which induces benign skin tumours in man. MCV can not be propagated *in vitro*. Whilst virus replication occurs only in the differentiating squamous cells, the proliferating cells are the underlying basal cells of the epidermis. MCV has been purified from skin lesions and the viral DNA has been subjected to molecular analysis. Three viral subtypes have been identified on the basis of different restriction endonuclease cleavage patterns and physical maps. One of these subtypes has been observed on only one occasion whilst the other two occur at a relative frequency of 3:1. There is no correlation of viral subtype with features of the clinical presentation.

MCV DNA is 185-195 kilobases long and possesses terminal cross-links and inverted repeats, typical of previously characterised poxviruses. Restriction maps have been obtained for all three subtypes: these differ extensively, in contrast to the situation for species of the orthopoxvirus genus, but are essentially colinear.

Given that vaccinia virus-infected cells secrete a viral polypeptide homologous to epidermal growth factor (EGF) and that EGF receptors within the epidermis occur predominantly on the basal cells, the MCV-induced proliferation of the latter may result from the secretion of a similar growth factor. MCV DNA does not hybridise to the human EGF

gene or to the vaccinia virus growth factor gene. (Neither do these genes cross-hybridise since the conserved amino-acid features are limited and involve different codon usage.) Hybridisation studies with oligonucleotide probes designed to select DNA sequences capable of encoding the conserved features of members of the EGF-like family of growth factors have led to the cloning of selected portions of the viral genome for sequence analysis, although this approach was ultimately unsuccessful in identifying an EGF-like sequence. In the course of these experiments the nucleotide sequence of the homologue of the vaccinia virus major envelope antigen was also obtained.

CONTENTS

Title page 1

Dedication 2

Abstract 3

Contents 5

List of figures 10

List of tables 12

Abbreviations used 13

Acknowledgements 16

1. Introduction 18

1.1 *Molluscum contagiosum* 18

1.1.1 Pathology of the *Molluscum contagiosum* lesion 19
1.1.2 Clinical features of *Molluscum contagiosum* 21
1.1.3 Epidemiology of *Molluscum contagiosum* 23

1.2 The poxviruses 25

1.2.1 Poxvirus taxonomy: poxviruses associated with human disease 25
1.2.2 Structure and replication of poxviruses 27
1.2.3 Poxvirus genome structure 30
1.2.4 Molecular biology of the poxviruses 34
1.2.5 Tumourigenic poxviruses 38
1.2.6 *Molluscum contagiosum* virus 42

1.3 Epidermal growth factor and its homologues 46

1.3.1 Epidermal growth factor and its receptor 46
1.3.2 Transforming growth factor-alpha 48
1.3.3 Poxviral growth factors 50
1.3.4 Sequence homologies of EGF-like polypeptides 53
1.3.5 Structure-function studies of the EGF-like polypeptides 57

1.4 Function of vaccinia virus growth factor: possible mechanisms of poxvirus-induced tumourigenesis 62

2. Aims of project 69

3. Materials and Methods 71

3.1 Materials and reagents 71
3.2 Water; sterilisation of liquids, plastics and glassware 71
3.3 Source of clinical material 72
3.4 Processing of Molluscum contagiosum lesions 73

3.4.1 Virus purification 73
3.4.2 Total DNA preparation 74
3.4.3 DNA purification 74

3.5 Restriction endonuclease digestion of DNA 75
3.6 Agarose gel electrophoresis of DNA 76
3.7 Field inversion gel electrophoresis of intact MCV DNA 77
3.8 Southern transfer of DNA 79
3.9 Preparation of oligonucleotides 80

3.9.1 Automated oligonucleotide synthesis 80
3.9.2 Oligonucleotide work-up 84
3.9.3 Purification of full-length oligonucleotides 86
3.9.4 Calculation of oligonucleotide concentration 89
3.9.5 Assessment of oligonucleotide purity 90
3.9.6 Theoretical assessment of oligonucleotide specificity 93

3.10 Radiolabelling of DNA 94

3.10.1 Random oligonucleotide primer extension 94
3.10.2 5' end-labelling of oligonucleotides 95

3.11 Filter hybridisation 96

3.11.1 Hybridisation with probes constructed by random oligonucleotide primer extension 96
3.11.2 Hybridisation with end-labelled oligonucleotide probes 97

3.12 Autoradiography 99
3.13 DNA clone construction 100

3.13.1 Fragment isolation from low melting-point agarose 100
3.13.2 Addition of synthetic EcoRI linkers to blunt-ended DNA 101
3.13.3 Preparation of vector DNA for cloning 103
3.13.4 Ligation of insert and vector DNAs 104

3.14 Growth, storage and transformation of E. coli 104

3.14.1 Host strains of E. coli 104
3.14.2 Growth media and antibiotics 105
3.14.3 Transformation with plasmid- and phage-based DNA clones 106
3.14.4 Storage of plasmid and phage clones 109

3.15 Preparation and restriction mapping of plasmid DNA 110

3.15.1 Small-scale isolation of plasmid DNA by alkaline lysis 110
3.15.2 Large-scale isolation of plasmid DNA by caesium chloride density centrifugation 111
3.15.3 Restriction mapping of cloned DNA 113

3.16 Preparation of M13 DNA 114

3.16.1 Isolation of M13 phage particles and single-stranded DNA for use as sequencing templates 114
3.16.2 Isolation of double-stranded M13 RF DNA 115

3.17 Unidirectional digestion of M13 clones with exonuclease III for the production of a nested set of deletion clones for sequencing 116

3.17.1 Establishing the conditions for exonuclease III digestion 117
3.17.2 Construction of deletion clones 118

3.18 DNA sequencing 121

3.18.1 Sequencing with DNA polymerase I "Klenow" fragment 121
3.18.2 Sequencing with T7 DNA polymerase (Sequenase™) 122
3.18.3 Polyacrylamide gel electrophoresis of sequencing reactions 124
3.18.4 Sequencing through "compressions" 125

3.19 Computer analysis of DNA sequence data 126

3.19.1 Reading of sequencing autoradiographs 126
3.19.2 Sequence data compilation 127
3.19.3 Sequence data analysis 127

4. Molluscum contagiosum virus: an epidemiological survey 129

4.1 There are three subtypes of Molluscum contagiosum virus 129
4.2 The study group of patients; occurrence of MCV subtypes 131
4.3 Statistical analysis of the occurrence of MCV subtypes 134
4.4 Summary of chapter 138

5. Molluscum contagiosum virus genome structure and restriction maps; the relationship of viral subtypes I, II and III; subtype genetic variation 139

5.1 Preparation and preliminary characterisation of MCV I DNA 139
5.2 Hybridisation studies for restriction map determination 147
5.3 Construction of MCV I restriction maps: interpretation of hybridisation data 151
5.4 Other data used in MCV I restriction map construction 159
5.5 Characterisation of MCV II DNA and construction of restriction maps 167
5.6 Characterisation of MCV III DNA and construction of restriction maps 178
5.7 Confirmation of MCV genome size by field inversion gel electrophoresis 185
5.8 Comparison of the genome organisation of MCV I, II and III by cross-hybridisation 187
5.9 Variations in the restriction profiles of different isolates of MCV DNA 195
5.10 Summary of chapter 199

6. Use of synthetic oligonucleotide probes to detect MCV DNA sequences capable of encoding a conserved domain of EGF-like growth factors 201

6.1 Screening of MCV DNA with cloned human EGF and VGF probes 201
6.2 Conserved features of the EGF-family of growth factors: design of oligonucleotide probes 1-3 202
6.3 Synthesis and use of oligonucleotides 1-3 208
6.4 Cloning the oligonucleotide-positive locus 214
6.5 Subcloning from pMBH-I and pMBH-II into M13mp8 220
6.6 Generation of nested deletion clones; DNA sequencing 223
6.7 Identifying the oligonucleotide target sequence 225
6.8 Analysis of sequence data 227
6.9 Sequence comparison with MCV I 232
6.10 Summary of chapter 236

7. Use of synthetic oligonucleotide probes, part 2: the MCV homologue of the vaccinia virus major envelope antigen 227

7.1 Design of oligonucleotides 5 and 6 237
7.2 Synthesis and use of oligonucleotides 5 and 6 239
7.3 Cloning the oligonucleotide-positive locus 242
7.4 Subcloning from pH/M-IIB into M13mp18 245
7.5 DNA sequence analysis and the oligonucleotide target sequence 246
7.6 MCV homologue of the vaccinia virus p37 gene 247
7.7 Summary of chapter 251

8. Use of synthetic oligonucleotide probes, part 3 252

8.1 Design of oligonucleotides 8-13 252
8.2 Synthesis and use of oligonucleotides 8-13 256
8.3 Cloning the oligonucleotide-positive loci 259
8.4 Subcloning from pH/Q-I into M13mp18 and 19 263
8.5 DNA sequence analysis and the oligonucleotide target sequence 263
8.6 Summary of chapter 267

9. Summary and discussion of results 268

9.1 The molecular epidemiology of subtypes of the Molluscum contagiosum virus 268
9.2 Structure and organisation of the Molluscum contagiosum virus genome 273
9.3 Molluscum contagiosum virus-induced tumourigenesis: use of synthetic oligonucleotide probes to identify members of the epidermal growth factor gene family 280
9.4 Use of oligonucleotides to identify new members of gene families 288
9.5 Scope for further work 291

References 294

Publications arising from this project 304

Appendices 305

1 Material and reagents 305
2 Patient details of MCV isolates, grouped according to subtype 310
3 Hybridisation and cross-hybridisation data for restriction mapping the genomes of MCV subtypes I, II and III 314
4 Nucleotide sequence data 320

LIST OF FIGURES

Chapter 1

Figure 1. Genome structure of vaccinia virus strain WR.

Figure 2. Schematic structure of human EGF as determined by solution nuclear magnetic resonance.

Figure 3. Proposed mechanism of *Molluscum contagiosum* virus-induced tumourigenesis.

Chapter 4

Figure 4. Comparative restriction digests of MCV I, MCV II and MCV III DNAs for the enzymes BamHI and ClaI.

Chapter 5

Figure 5. Restriction digests of MCV I DNA and identification of rapidly renaturing terminal fragments.

Figure 6. An autoradiograph illustrative of those obtained during restriction mapping analysis.

Figure 7. Restriction maps of MCV I DNA for the enzymes BamHI, ClaI, HindIII, PstI, SmaI and XhoI.

Figure 8. Restriction digests of MCV II DNA with BamHI, ClaI and HindIII.

Figure 9. Restriction maps of MCV II DNA for the enzymes BamHI, ClaI and HindIII.

Figure 10. Restriction digests of MCV III DNA with BamHI, ClaI and HindIII.

Figure 11. Restriction maps of MCV III DNA for the enzymes BamHI, ClaI and HindIII.

Figure 12. Field inversion gel electrophoresis of intact MCV DNA.

Figure 13. Comparison of the restriction maps of the genomes of MCV I, MCV II and MCV III for the enzymes BamHI, ClaI and HindIII.

Figure 14. Variations in the restriction profiles of six isolates of MCV II DNA.

Chapter 6

Figure 15. Hybridisation of oligonucleotide 2 to BamHI, ClaI and HindIII digests of MCV I and MCV III DNA.

Figure 16. The locus of oligonucleotide 2 hybridisation to MCV subtypes I, II and III.

Figure 17. The scheme for cloning the overlap of MCV I BamHI fragment B and HindIII fragment C containing the locus of oligonucleotide hybridisation.

Figure 18. Restriction maps of clones pMBH-I and pMBH-II.

Figure 19. Open reading frames within the 2,571 bases of nucleotide sequence obtained from clone mII.

Figure 20. Amino-acid sequences of the putative translates of orfs A, B and C from the nucleotide sequence of clone mII.

Figure 21. Comparison of the 364 bases of sequence data obtained from clone mIa (from MCV I) with that of mII (MCV II).

Figure 22. Comparison of the C-terminal amino-acid sequences of the translates of MCV II orf A and orf B with their MCV I equivalents.

Chapter 7

Figure 23. Hybridisation of oligonucleotide 6 to BamHI, ClaI and HindIII digests of MCV II DNA.

Figure 24. Restriction map of clone pH/M-IIB.

Figure 25. Comparison of the amino-acid sequences of the C-termini of the vaccinia virus 37 kD major envelope antigen (p37) and its MCV II homologue.

Chapter 8

Figure 26. Hybridisation of oligonucleotides 11 and 12 to BamHI and HindIII digests of MCV I DNA.

Figure 27. Restriction maps of clones pH/Q-I and pH/K_L-I.

LIST OF TABLES

Chapter 1

Table 1. Poxviruses associated with proliferative diseases in their natural hosts.

Table 2. The EGF-like growth factor family.

Table 3. Other EGF-like polypeptide sequences.

Chapter 3

Table 4. Oligonucleotide design for the conserved domains A and B of the EGF-family of growth factors.

Table 5. Parameters for the series of oligonucleotides synthesised in this project.

Chapter 4

Table 6. Incidence of MCV I and MCV II in various patient categories.

Chapter 5

Table 7. Sizes of restriction fragments of MCV I DNA generated by BamHI, ClaI, HindIII, PstI, SmaI and XhoI.

Table 8. Sizes of restriction fragments of MCV II DNA generated by BamHI, ClaI and HindIII.

Table 9. Sizes of restriction fragments of MCV III DNA generated by BamHI, ClaI and HindIII.

Chapter 9

Table 10. Cumulative data relating to the occurrence of MCV subtypes

ABBREVIATIONS USED

Abbreviation	Meaning
2', 3', 5'	Refers to positions on the (deoxy)ribose moiety of (deoxy)ribonucleotides
A	Adenine
A_{260}	Absorbance (optical density) at 260 nm
ATP	Adenosine-5'-triphosphate
Bam/, etc.	Refers to a particular restriction endonuclease (eg. BamHI) fragment, etc.
bp	Base pairs
BSA	Bovine serum albumin
C	Carboxy (eg. C-terminal)
C	Cytosine
cDNA	Complementary DNA (q.v.)
Ci	Curie (3.7×10^{10} disintegrations per sec)
CPE	Cytopathic effect
dAMP, etc.	2'-deoxyadenosine-5'-monophosphate, etc.
dATP, etc.	2'-deoxyadenosine-5'-triphosphate, etc.
dITP	2'-deoxyinosine-5'-triphosphate
ddATP, etc.	2',3'-dideoxyadenosine-5'-triphosphate, etc.
DMSO	Dimethylsulphoxide
DMTr	The dimethoxytrityl 5'-protecting group
DNA	Deoxyribonucleic acid
dpm	Disintegrations per minute
DTT	Dithiothreitol
E_{260}, E_{280}	Absorption coefficient at 260 or 280 nm
E. coli	*Escherichia coli*

EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
G	Guanine
g, g_{av}	(Average) radial centifugal force
kb	Kilobase pairs
kD	Kilo Daltons
M	Molar
MCV (I/II/III)	Molluscum contagiosum virus (subtype I/II/III)
MGF	Myxoma virus growth factor
mol	Mole (of substance)
M_r	Relative molecular weight
mRNA	Messenger RNA (q.v.)
N	Amino (eg. N-terminal, N-glycosylated)
n-mer	An oligodeoxyribonucleotide of length n bases
orf	Open reading frame
P	Probability, X^2_y (q.v.) significance value
$p(dN)_6$	Random hexadeoxynucleotides
RF	Replicative form of bacteriophage M13 DNA (q.v.)
RNA	Ribonucleic acid
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SFGF	Shope rabbit fibroma virus growth factor
SFV	Shope rabbit fibroma virus
SSC	150 mM NaCl, 15 mM sodium citrate, pH 7.0
T	Thymine
T_m	The melting temperature (of an oligonucleotide)
TBE	89 mM Tris base, 89 mM boric acid, 2.8 mM EDTA, pH 8.3 (q.q.v.)
TE	10 mM Tris.Cl pH 8.0, 1 mM EDTA (q.q.v.)
TGF	Transforming growth factor-alpha

Tris	Tris(hydroxymethyl)aminomethane
Tris.Cl	A buffer composed of tris base (q.v.) and tris hydrochloride
tRNA	Transfer RNA (q.v.)
VGF	Vaccinia virus growth factor
X	Any of the nucleotide bases A, C, G or T (q.q.v.)
X^2_y	Chi-squared significance test with Yates' correction
YMTV	Yaba monkey tumour virus

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to Dr. Len Archard for his continued support, supervision and enthusiasm throughout this project, and for providing me with the opportunity to learn something of the fields of molecular biology and virology in his laboratory. My thanks go also to Professor Alan Malcolm for his interest and helpful advice, and, together with Dr. Archard, for providing an encouraging and stimulating research environment in the Department of Biochemistry, Charing Cross and Westminster Medical School.

I am indebted to Dr. Jeoff Cream, Consultant Dermatologist, Charing Cross Hospital for co-ordinating sample collection and for his collaboration in supplying clinical details. I also thank Drs. S. Breathnach, M. Muhlemann and N. Rosedale for their co-operation in providing the clinical material.

Neil Blake contributed greatly to generating nucleotide sequence data, for which I am very grateful: I thank him also for making available to me the complete sequence of the MCV homologue of the vaccinia virus p37 gene which he has since determined. I also thank all the members of the laboratory with whom I have worked, for their friendship, help and suggestions. In particular, I would like to thank Dr. Neil Bowles and Dr. Colin Wickenden for their invaluable tuition during the initial period of my stay in the laboratory and for

their continued assistance and friendship.

I thank the Cancer Research Campaign for their financial support during my studentship, and, together with the Biochemical Society of Great Britain and the Society for General Microbiology, for financial assistance for attending scientific meetings.

Lastly, but by no means least, I thank my mother and father, and my wife, Lorraine, for their love, interest and encouragement, without which this thesis would not have been possible.

CHAPTER 1

Introduction

This chapter begins with a description of the skin disease *Molluscum contagiosum*. Next, the poxviruses and preliminary studies of the aetiological agent of *Molluscum contagiosum*, a member of this virus group, are discussed. The family of growth factors related to epidermal growth factor and including several poxviral gene products is reviewed and, finally, possible mechanisms of poxvirus-induced tumourigenesis are discussed, including a proposed scheme for *Molluscum contagiosum* virus.

1.1 *Molluscum contagiosum*

Molluscum contagiosum is a benign skin tumour of man resulting from enhanced basal cell proliferation. The tumours are generally 2-5 mm in diameter and are elevated above the surface of the skin as firm, flesh-coloured papules with an erythematous base and a central depression, from which a white, curdy core can be expressed. The disease was first described by Bateman (1814) and has been reviewed by Postlethwaite (1970) and Brown *et al*. (1981).

1.1.1 Pathology of the Molluscum contagiosum lesion

The epidermis of skin consists of several layers of cells at different stages of differentiation (Montagna and Parakkal, 1974). The lowest, above the basement membrane that separates the epidermis from the dermis, is the stratum basale, the single cell layer of cuboidal basal cells which are the proliferating stem cells of the epidermis. Displaced daughter cells move into the higher levels, adopting a more planar, squamous shape and becoming differentiated keratinocytes. The stratum spinosum and stratum granulosum in turn overlie the stratum basale, and the stratum corneum is the outer surface of the skin. Homeostasis requires a balance between desquamation and self-renewal by basal cell turnover.

In the Molluscum contagiosum lesion, the above picture is disturbed: the tumour consists of localised hypertrophied and hyperplastic epidermis extending into the dermis beneath and projecting above the surface as a papule, but is benign in that it does not traverse the basement membrane. Intracytoplasmic inclusion bodies ("molluscum bodies") were described independently by Henderson (1841) and Paterson (1841). These Henderson-Paterson bodies first appear in cells of the stratum spinosum and enlarge as the affected cells develop and migrate through the stratum granulosum towards the surface: eventually they almost fill the cytoplasm. Unaffected cells within the lesion continue to develop normally and produce keratin. In the stratum corneum, the degenerating cells with their inclusion bodies become enmeshed in the

keratin network: this dissolves in the centre of the lesion to form the expressible lesion core which is rich in inclusion bodies (Sutton and Burnett, 1969).

Molluscum contagiosum lesions contain particles similar in appearance to those seen in poxvirus-infected cells. Their fine structure was examined by electron microscopy which, together with their intracytoplasmic location and the appearance of the inclusion bodies, confirmed that these particles are poxviruses (reviewed by Postlethwaite, 1970). No virus particles are present in the basal cells but in overlying cells a cytoplasmic "matrix" develops, in which virus particles are assembled, eventually becoming the inclusion body (Sutton and Burnett, 1969). Virus particles are present exclusively in the cytoplasm of infected cells. Immunofluorescence studies did not detect virus antigen in basal cells or in uninfected cells adjacent to the lesion (Shirodaria and Matthews, 1977). Sutton and Burnett (1969) also noted some hyperplasia of the subjacent dermal fibroblasts, although these cells did not harbour virus particles.

The Molluscum contagiosum tumour results from enhanced proliferation of basal cells (Epstein et al., 1966). This has been demonstrated by injection of tritiated thymidine into skin lesions in situ with subsequent lesion sectioning and autoradiography. Although there was no infection of basal cells, their division time was reduced from 6.1 to 3.4 days. The renewal time was also estimated by arresting cell

division with colcemid and counting metaphase figures, giving a value of 2.8 days. Cytoplasmic labelling with thymidine was present in the cell layers containing inclusion bodies, indicating viral DNA replication in these sites. Observation of sections from lesions at various times after labelling revealed that infected cells migrated at a uniform rate through the epidermis and were shed within 9-15 days, contrasting with the less regulated 9-30 days for uninfected cells to reach the stratum granulosum. Synthesis of viral DNA was confined to the first 4-5 days (Epstein and Fukayama, 1973). Infected cells showed no nuclear labelling with tritiated uridine, indicating a shut-off of host transcription, but showed some labelling over the cytoplasmic inclusion bodies due to transcription of viral DNA: this is a usual feature of poxvirus-infected cells (see section 1.2.2).

1.1.2 Clinical features of Molluscum contagiosum

The incubation period for infection with Molluscum contagiosum virus is one week to six months. Without treatment, individual lesions resolve spontaneously within a few months. However, autoinoculation with virus from lesion cores is common and so the disease may last much longer. Treatment involves lesion disruption by expression of the core, or removal by curettage or freezing with liquid nitrogen. Rare complications of the disease result from lesion rupture, in which core material is discharged into the dermis, resulting in an inflammatory response.

Molluscum contagiosum lesions are typically 2-5 mm in diameter. However, large, often solitary lesions can occur with diameters of 15 mm or more ("giant" Molluscum contagiosum; Brown et al., 1981). Occasionally lesions are intrafollicular - ie. they occur in the invaginated epidermis of a hair follicle - causing sycosis barbae (Krishnamurthy, 1974).

An immune response to Molluscum contagiosum was demonstrated in only 46 of 67 patients (69%), although virus antigen was detected in all lesions examined (Shirodaria and Matthews, 1977). The antibody response was predominantly of the IgG class, indicative of recurrent infection. The lack of a response in the remainder has been suggested to be due to the need for lesion disruption to present antigen to the immune system: the superficial site of viral replication in the epidermis and the outward migration of infected cells both minimise contact with the immune system. Support for this suggestion is the induction of a response following lesion trauma in four of nine patients previously negative for antibody (Shirodaria and Matthews, 1977). Involvement of the immune system in resolution of the disease is also implicated by the observation that lesions may be numerous, widespread and long-lasting in the immunosuppressed, in those with atopic dermatitis, and in those with acquired immunodeficiency syndrome (Brown et al., 1981; Cotton et al., 1987; Katzman et al., 1987).

1.1.3 Epidemiology of Molluscum contagiosum

The distribution of Molluscum contagiosum is worldwide. Its incidence in Papua New Guinea in 1971 was 10% of children, and in Fiji in 1966 it was 4.5% of an entire village of 672 people, with the peak incidence in children of 2-3 years of age (Stuart et al., 1971; Postlethwaite et al., 1967). Between 1956 to 1963 the incidence amongst 368 skin outpatients in Scotland was 1.2%, with a peak incidence at 10-12 years of age (Postlethwaite et al., 1967). In the latter study, Molluscum contagiosum was 3- to 4-fold more prevalent in boys than in girls for the age group 7-15 years, for which there were 298 patients: age groups above and below this group showed approximately equal sex distributions. A smaller study by Overfield and Brody (1966) recorded a similarly higher incidence amongst boys in Alaska. The peak incidence and the sex bias of the Scottish study contrast with those of the Fijian study.

Transmission of Molluscum contagiosum virus is mechanical, by direct or indirect skin-to-skin contact. Reported means include intrafamilial transmission between siblings, and transmission from a surgeon's hand to patients, between wrestlers, between children sharing bath towels and gymnasium equipment, and between sexual partners (reviewed by Brown et al., 1981). Patients usually have clustered lesions, which may occur at most sites of the body, although they are rare on the palms and soles. Lesions at genital sites are

common in adults: sexual promiscuity, the occurrence of other sexually transmitted diseases, and the presence of lesions in sexual partners suggest sexual transmission of Molluscum contagiosum in these cases (Cobbold and MacDonald, 1970). In common with the incidence of other sexually transmitted diseases, the genital incidence of Molluscum contagiosum has increased over recent years (Becker et al., 1986).

Although generally regarded as a disease unique to man, occasional reports have appeared describing Molluscum contagiosum in other species: Douglas et al. (1967) reported lesions in captive chimpanzees, Bagnall and Wilson (1974) reported lesions in a kangaroo, and Cooley et al. (1987) reported widespread lesions in a horse with immunosuppression due to granulomatous enteritis. However, these reports were based on cytochemistry and electron microscopy and so do not confirm that the virus was the same as that responsible for the human disease.

1.2 The poxviruses

Physically, and in terms of their genetic capability, the poxviruses are amongst the largest viruses known. They have a characteristic morphology and, with the exception of the iridoviruses, are unique in that their complete replicative cycle occurs in the cytoplasm. The poxviruses have been extensively reviewed by Andrews *et al.* (1978), Baxby (1984), Fenner (1985) and Moss (1985).

1.2.1 Poxvirus taxonomy: poxviruses associated with human disease

The single taxonomic family *poxviridae* comprises two sub-families: the *Chordopoxvirinae* and *Entomopoxvirinae*. The latter consists of poxviruses of insects and comprises three genera. The former consists of poxviruses of vertebrates and comprises six genera: the orthopoxviruses, parapoxviruses, capripoxviruses, leporipoxviruses, avipoxviruses and suipoxviruses. Genera are distinguished serologically and by host species. There are several other members of the sub-family *Chordopoxvirinae* not yet officially classified, of which *Molluscum contagiosum* and Yaba monkey tumour virus (and the serologically-related Tanapox virus) are examples. The orthopoxviruses include variola, the agent of smallpox, and vaccinia, the virus used in the immunisation campaign to eradicate smallpox. Because of this, the orthopoxviruses, and vaccinia virus in particular, are the best studied of the

poxviruses.

A number of poxviruses can cause disease in humans (Fenner, 1985). Variola virus and Molluscum contagiosum virus are the only poxviruses responsible for specifically human disease. The latter is responsible for benign skin tumours and is discussed more fully below (section 1.2.6). Several other poxviruses are responsible for zoonotic disease in humans.

Variola virus was responsible for a systemic and often fatal disease. An intense viraemia led to infection of the epidermis and the appearance of multiple skin lesions which resolved several weeks later. Vaccinia virus is a related orthopoxvirus that has been used to provide protective immunity to variola. Skin lesions caused by vaccinia virus have a similar pathology to those of variola virus but are localised. There are rare complications of vaccination associated with immunosuppression. Other orthopoxviruses that can cause disease in humans are cowpoxviruses, with a similar pathology to that of vaccinia virus, and monkeypox virus, which causes a smallpox-like disease in monkeys and may rarely infect humans but is poorly transmissable.

Orf virus and milker's nodules (or paravaccinia) virus are parapoxviruses that can be transmitted from sheep and cows respectively and cause large skin lesions lasting for several weeks. The unclassified but related Tanapox and Yaba monkey tumour viruses of monkeys may also rarely infect humans. The

former causes single large skin lesions which regress after a few weeks, whilst the latter causes benign, subcutaneous tumours in monkeys and humans (see section 1.2.5).

1.2.2 Structure and replication of poxviruses

The following account of poxvirus replication relates to vaccinia virus, the most widely studied of the poxviruses: the most commonly used strain of this virus is "WR", the Western Reserve vaccine strain. Although poxviruses show variations in size and morphology, growth characteristics and host range, it is thought to be largely valid for most (Andrews *et al.*, 1978; Baxby, 1984; Moss, 1985). The structural description relates to virions released from cultured cells after experimental lysis: those released in the absence of lysis possess a further lipid envelope (see below).

The virus particle is a brick shaped body 218 x 270 nm in size. It consists of a membrane-delimited central core with a biconcave disc morphology: two protein lateral bodies rest in the concavities and the whole is enclosed in an outer phospholipid membrane which differs from the host cellular membranes. Virus particles consist of 90% protein, 6% lipid, 3% DNA and trace amounts of carbohydrate and RNA. The virion core contains the virus genome, a single molecule of double-stranded DNA. The core also contains enzymes necessary to establish replication at the cytoplasmic location, including those for viral gene transcription and mRNA modification.

The presence of these enzymes accounts for the phenomenon of non-genetic reactivation whereby a heat-inactivated poxvirus can be "rescued" by the enzyme activities of a second ultraviolet-irradiated poxvirus. Similarly, transfected purified poxviral DNA can be rescued by co-infection with another poxvirus (Sam and Dumbell, 1981).

Vaccinia virus grows in a wide variety of cells in culture and causes plaques due to cell lysis. Virions may enter the cell by endocytosis or by fusion of the outer lipid coat with the cell membrane and entry of partially uncoated virions (Chang and Metz, 1976): the latter is thought to be the infective route. "Second-stage uncoating" releases the DNA from the virion core and requires the combined action of host enzymes and newly synthesised virus proteins. Synthesis of viral proteins is temporally regulated. The two main classes are the "early" proteins, expressed before the onset of DNA replication, and the "late" proteins (see section 1.2.4; Pennington, 1974).

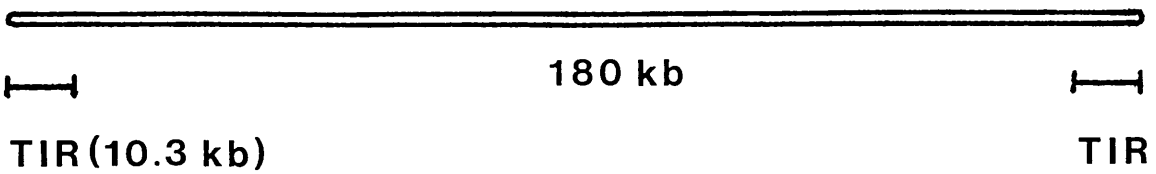
Viral DNA replication and virion assembly occur within intracytoplasmic inclusions ("factory sites") and does not require host nuclear functions. During the replicative cycle, host DNA replication and transcription are halted, host DNA is degraded and eventually cell lysis occurs. Mature virions exit through the plasma membrane via the Golgi system or are released by cytolysis. Those released prior to cytolysis ("extracellular virus") differ antigenically from those released later. This is due to their acquisition of a

further membrane of Golgi origin containing a virus late gene product of 37 kD (p37, the major envelope antigen; Hiller and Weber, 1985): the p37 antigen comprises 5-7% of the protein mass of the mature virion. Initially, a double-membrane forms and virion release occurs by fusion of the outer layer with the plasma membrane. The p37-containing membrane is not required for infectivity although it mediates more efficient cell adsorption.

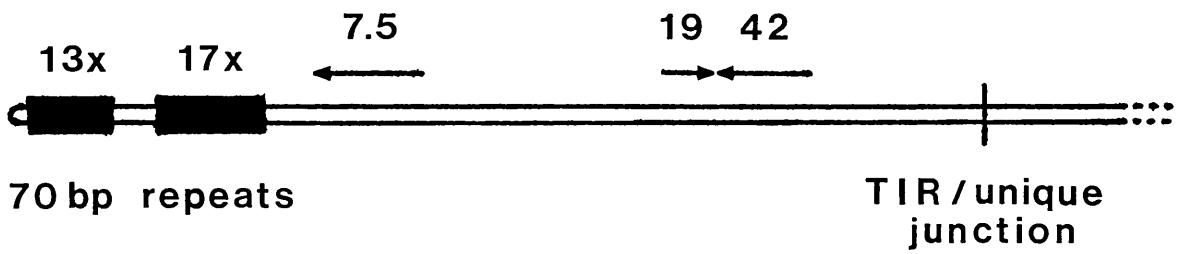
1.2.3 Poxvirus genome structure

Poxvirus genomes are single, linear molecules of double-stranded DNA, ranging in size from approximately 130 kb for parapoxviruses to 280 kb for avipoxviruses. Base composition ranges from 36% G+C for orthopoxviruses to 63% for parapoxviruses (Baxby, 1984). The genomic termini are inverted repeats and these in turn contain multiple tandem repetitions. There are two blocks of 13 and 17 repetitions of a 70 bp unit within a cloned terminal repeat of 10.3 kb for vaccinia virus strain WR, with an interruption of 325 bp between them (Wittek and Moss, 1980). The 3.5 kb terminal repeat of orf virus contains repetitions of a 17 bp unit (Robinson *et al.*, 1987). Poxviruses grown in culture show a heterogeneous repeat size (Wittek *et al.*, 1978), indicative of variation in the number of these repetitions. The genome ends are covalently closed, akin to the structure of chromosome telomeres, and so the genome is formed from a continuous polynucleotide strand (Geshelin and Berns, 1974). The terminal hairpin of vaccinia virus DNA has been sequenced and shown to possess two forms, which can be differentiated electrophoretically: their sequences around the hairpin are inverted and complementary to each other (Baroudy *et al.*, 1982). They are 92% AT-rich and incompletely base-paired, the latter accounting for the two forms with different mobilities since the nature of the unmatched bases is complementary. Similar "flip-flop" structures have been identified for Shope rabbit fibroma virus DNA (Delange *et al.*, 1986). The structure of the vaccinia virus WR genome is illustrated in figure 1.

(a)



(b)



(c)

```
 AAGTTAG TAAATT AT ATATATAA TTTTA TA ATTAAT TTA ATTTTAC TAATTT-3'
(  ::::::: :::::: :: :::::::: ::::: :: :::::: ::: ::::::  : ::::
 ATCAATCTATTTAATTATTATATATTCAAAATCATGTAATTATAATATAAAATTTA TAAA-5'
```

Figure 1. Genome structure of vaccinia virus strain WR. (a) The total genome length is 180 kb and the termini are covalently closed (see text). There is a terminal inverted repeat (TIR) of 10.3 kb, the structure of which is expanded in (b). Within the TIR are two blocks of a 70 bp unit (boxes) as well as three immediate early genes encoding polypeptides of 7.5, 19 and 42 kD: the positions and orientations of these genes are indicated (Wittek et al., 1980b). (The product of the second of these is the precursor (p19) of the vaccinia virus growth factor; see section 1.3.3.) In (c), the terminus is expanded to show the hairpin loop, which is incompletely base-paired. There are two forms of the hairpin: that drawn here and that formed by the complementary sequence (for which the unpaired bases differ and impart a different electrophoretic mobility). The data are taken from Baroudy et al., 1982.

The restriction endonuclease "maps" of many poxviral genomes are available. The best studied genus is the orthopoxviruses, for which the maps of HindIII, SmaI and XhoI sites for 16 strains of rabbitpox, vaccinia, monkeypox, variola, cowpox and ectromelia viruses were compared (Mackett and Archard, 1979). Terminal fragments were identified by their rapid renaturation and cross-hybridisation, and the maps were determined by hybridisation of individual DNA restriction fragments to Southern blots of genomic digests. Despite differences in the genome sizes of these viruses (from 182 kb for rabbitpox to 220 kb for cowpox), these studies indicated a high level of conservation of restriction sites throughout a central portion of 45 kb of their genomes. Restriction site conservation is a marker of sequence conservation. Species-specific differences of size and sequence map to the terminal regions which presumably encode the distinguishable phenotypes (including host-range) of these species. In this respect, the altered phenotype of virus "white pock mutants" (see below) is due to the deletion of near-terminal sequences (Archard and Mackett, 1979; Dumbell and Archard, 1980). With regard to host-range, a gene necessary for replication in human cells has been mapped to approximately 25 kb from the left end of the vaccinia virus genome (Gillard et al., 1986). The comparison of orthopoxviral restriction maps has been extended to 38 strains, including gerbilpox, racoonpox and camelpox viruses (Esposito and Knight, 1985).

The restriction maps for 12 strains of papular stomatitis, orf and milker's node viruses (parapoxviruses)

have been similarly compared and reveal conservation of a central region and species-specific terminal differences (Gassmann et al., 1985). Maps of 17 strains of orf virus were highly conserved throughout (Robinson et al., 1987). A similar study of six strains of capripoxviruses from sheep, goats and cattle revealed a high level of restriction site conservation throughout the genomes and an estimated 96-97% sequence homology (Gershon and Black, 1988). This study also described one strain thought to have arisen by natural recombination between two others.

Restriction maps for the tumourigenic Shope rabbit fibroma virus and Yaba monkey tumour virus are available (see section 1.2.5; Cabirac et al., 1985; Kilpatrick and Rouhandeh, 1985): they bear no resemblance to those of other previously characterised viruses.

Plaques due to wild-type cowpox virus growth on chorioallantoic membranes are haemorrhagic. "White pock mutants" give white plaques due to their failure to prevent leukocyte infiltration, and have right-terminal deletions of 32-41 kb of the wild-type genome (Archard and Mackett, 1979). This deletion is compensated for by copying a variable amount (4.5-41 kb) of the left-terminal DNA to the right-terminus, so regenerating a terminal repeat and cross-link (Archard et al., 1984). This observation provides evidence for transient circularisation with conjunction of the termini during DNA replication, as discussed below. White pock mutants of other orthopoxviruses have also been described (eg. monkeypoxvirus;

Dumbell and Archard, 1980).

1.2.4 Molecular biology of the poxviruses

The proposed mechanism of poxviral DNA replication is asymmetric and involves initiation within the terminal inverted repeat and continuous replication of one strand with displacement of the other. Such an initiation event could occur at both ends of the genome: however, should it occur at one end only, the termini of the displaced strand are proposed to self-anneal via the inverted repeats, forming a "pan-handle" structure from which replication can initiate as before (Wittek and Moss, 1980). This mechanism is analagous to that of adenovirus DNA replication. The presence of tandem repetitions within the terminal repeat is thought to enhance the kinetics of the self-annealing step, which would otherwise be unfavourable for such a large DNA molecule. The terminal cross-links must be cleaved by a single-stranded endonuclease to allow replication. It has been proposed that this enzyme is strand- and site-specific and nicks proximal to the last region of single-strandedness within the incompletely base-paired hairpin. DNA synthesis primed from this nick could displace the other strand, unfold and replicate round the hairpin, and so generate a palindromic repeat at the terminus of each strand. Thus, each strand could displace the other at the terminus by forming two new hairpins and self-primed replication of the strand into which the nick was first introduced could follow, as discussed above (Baroudy *et al.*,

1982). Such a mechanism is also able to account for interconversion of the two hairpin forms ("flip-flop"; see above) and for virus mutants in which deletions at one end of the genome are replaced by copies of DNA from the other end (Archard et al., 1984).

As mentioned above, poxvirus genes are expressed in two temporal classes: "early" genes, expressed prior to viral DNA replication (including the "immediate early" genes, expressed prior to viral protein synthesis); and "late" genes, expressed after DNA replication. Immediate early genes are transcribed within the virion core by the core DNA-dependent RNA polymerase and their products are thought to include proteins required to complete uncoating and potentiate late gene transcription. Examples of early genes are the DNA-dependent DNA polymerase, a deoxyribonuclease, a ligase, poly-A polymerase, thymidine kinase and at least one structural protein. Late genes encode most of the structural proteins and the virion enzymes. Early genes continue to be transcribed at late stages, although translational control prevents the continued production of some early proteins (eg. thymidine kinase; McAuslan, 1963).

All transcripts are polyadenylated. The viral mRNA capping enzyme is responsible both for modification of their 5'-ends (by creation of a 7-methyl-guanosine cap) and for termination of early gene transcription downstream of specific TTTTTXT signal sequences (Rohrmann et al., 1986; Yuen and Moss, 1987; Shuman et al., 1987): the cognate sequence UUUUUXU

in the nascent mRNA is responsible for termination (Shuman and Moss, 1988). Termination of late gene transcription is not specific and so late transcripts have heterogeneous 3'-ends. Late gene transcripts are apparently synthesised discontinuously, their coding sequences being tagged on the 3'-ends of other leader (early or late) viral transcripts (Bertholet et al., 1987), poly(A) tracts (Schwer et al., 1987; Schwer and Stunnenberg, 1988) or cellular polyadenylated RNA (Van Meir and Wittek, 1988). It is not known whether the poly(A) sequence is necessary for transcription priming or is ligated to late mRNA. A means of preferentially translating the late coding region of chimaeric mRNAs is required. These structures may explain the continued presence of early transcripts late in infection (see above).

Sequence analysis of regulatory sequences involved in the temporal pattern of virus gene expression indicates that the regions upstream of early and late gene transcription initiation sites are different and also differ from those of host cell genes. Vaccinia virus early gene promoters are AT-rich (>80%) within the 45 bp preceding the transcription initiation site and a consensus sequence has been described (Mars and Beaud, 1987). In contrast, there is no late promoter consensus, possibly reflecting the tagging of late transcripts onto other leader mRNAs (see above). Viral transcripts do not contain introns and their non-translated 5'-leader and 3'-tail regions are short compared to host gene transcripts. The sequence TAAATG occurs upstream of most late genes: it commonly contains both transcription and translation

initiation sites and may mediate attachment to leader mRNA (Bertholet et al., 1987). Mutation of this latter sequence motif abolishes late gene transcription in vitro (Wright and Moss, 1987).

Transcriptional maps of the vaccinia virus genome have been obtained by in vitro translation of hybrid-selected mRNA and by nucleotide sequencing. There is no global segregation of early and late genes to different genomic regions, although a cluster of late genes has been described (Weinrich and Hruby, 1986). Poxviral genes are closely spaced, often partially overlapping (Pluciennicczak et al., 1985; Niles et al., 1986). The genes for three prominent immediate early proteins have been mapped to the vaccinia virus terminal repeat, proximal to the tandem repetitions (Wittek et al., 1980a). A 2,236 bp region containing two of these (the p19 and p42 genes, encoding 19 kD and 42 kD proteins) was sequenced in order to analyse regulatory elements (Venkatesan et al., 1982): the p19 gene product has since been shown to be the precursor of vaccinia virus growth factor, a member of the family of growth factors related to epidermal growth factor (see section 1.3.3).

Sequencing studies have provided information on the comparative genomic organisation of different poxvirus genera. A central region of the fowlpox virus (FPV) genome shows an arrangement of closely packed and partially overlapping open reading frames that is the same as that for vaccinia virus, with the exception of the location of the thymidine kinase

gene (Binns et al., 1988). There was little homology at the nucleotide level but significant homology (33-35% identity) at the amino-acid level. However, the organisation of genes within the terminal region of the FPV genome differs from that of the orthopoxviruses and sequence homology is lower (Tomley et al., 1988). Poxviral genes whose products show sequence as well as functional homology to prokaryotic and eukaryotic RNA polymerases (Broyles and Moss, 1986), mammalian and avian thymidine kinases (Boyle et al., 1987), herpes and adenovirus DNA polymerases (Earl et al., 1986) and a mammalian polypeptide growth factor (see section 1.3.3) have been described.

1.2.5 Tumourigenic poxviruses

Unlike orthopoxviruses, a number of poxviruses induce tumours in their natural hosts: see table 1. The lesions of orf and paravaccinia viruses are also proliferative. Transmission is by contact, or by arthropod vectors (demonstrated for Shope rabbit fibroma and fowlpox viruses). Molluscum contagiosum virus is tumourigenic in man and is discussed further below (section 1.2.6).

Fowlpox virus is an avipoxvirus that induces warty growths on the unfeathered skin of chickens (Woodruff and Goodpasture, 1931). The lesions consist of hyperplastic epithelial cells which possess cytoplasmic inclusion bodies. Lesions persist for several weeks before regression.

Yaba monkey tumour virus (YMTV) is an unclassified poxvirus causing benign skin tumours in primates. Tumour progression following subcutaneous injection of rhesus monkeys has been described by Sproul et al. (1963). The tumours are dermal or subcutaneous histiocytomas which proliferate for 2 weeks, reaching 2-5 cm diameter, and regress within 2-3 months. The proliferating cell type is the histiocyte, which migrates into the infected area and assumes a fibroblastic morphology. Proliferating cells may also possess cytoplasmic inclusions indicative of infection. The tumours are not fibrotic, in contrast to those of the fibroma viruses.

Shope rabbit fibroma virus (SFV) and the related fibroma viruses of hares and squirrels are leporipoxviruses inducing benign subcutaneous fibromas in cottontails and rabbits due to proliferation of fibroblasts (Shope, 1932). The tumours persist for 2 weeks and then regress. A recombinant between SFV and the related rabbit myxoma virus, in which the tumourigenic properties of SFV are combined with the host-range and disseminative properties of myxoma virus, has been described: malignant rabbit virus (MRV; Cabirac et al., 1985). The MRV genome is basically that of myxoma virus with the replacement of 5 kb of the terminal repeat and 3 kb of right unique sequence with 8 kb from the right terminus of SFV (Upton et al., 1988). This region of SFV contains several early genes, including the gene for the Shope rabbit fibroma virus growth factor (SFGF), a homologue of epidermal growth factor (see section 1.3.3). The myxoma virus homologue of

this gene is replaced with that for SFGF: it is possible that this replacement is responsible for the altered tumourigenic phenotype of MRV.

Virus	Genus	Natural host	Lesion/proliferating cell type
Fowlpox	Avipoxvirinae	Fowl	Hyperplastic epithelium
Molluscum contagiosum	Unclassified	Man	Epidermal basal cell tumour
Shope rabbit fibroma	Leporipoxvirinae	Cottontail and rabbit	Fibroma
Yaba monkey tumour	Unclassified	Monkey	Histiocytoma

Table 1. Poxviruses associated with proliferative diseases in their natural hosts.

1.2.6 Molluscum contagiosum virus

The lesion pathology, clinical features and epidemiology of Molluscum contagiosum were described in section 1.1. The Molluscum contagiosum virus (MCV) particle is approximately 100 x 200 x 300 nm in size. It has been purified from clinical lesions and shown to be similar to vaccinia virus in terms of its chemical composition and physical properties (Pirie et al., 1971). The structural proteins of the virus have been studied by polyacrylamide gel electrophoresis, revealing >40 polypeptides (Oda et al., 1982). Nine isolates of MCV were studied, showing variations in the sizes of two major structural proteins present in the surface or subsurface layers of virions but not in virion cores. There is no antigenic cross-reactivity of MCV with vaccinia, cowpox or fowlpox viruses in complement fixation tests (Mitchell, 1953) or with vaccinia, mousepox (ectromelia) or rabbitpox viruses by immunofluorescent staining of sectioned lesions (Shirodaria and Matthews, 1977). MCV is serologically distinct from previously classified genera and so is termed an "unclassified" poxvirus.

Molluscum contagiosum virus has been transmitted by inoculation of human subjects (Juliusberg, 1905) but similar experiments with monkeys, apes, sheep, fowl, guinea pigs and mice were unsuccessful. Attempts to grow the virus on the chick embryo chorioallantoic membrane and in human embryonic skin organ cultures were also unsuccessful. Primary and continuous cultures of cells have also been used: although

there was limited transmission of a cytopathic effect (CPE; cell rounding and clumping) and induction of interferon for a few passages, no culture system for the propagation of MCV has been established (reviewed by Postlethwaite, 1970). Attempts to "rescue" the virus by simultaneous infection of the chorioallantoic membrane with vaccinia, fowlpox or cowpox viruses were also unsuccessful (Mitchell, 1953). Francis and Bradford (1976) more recently reported propagation of MCV in human amnion cells but attempts to repeat this were unsuccessful.

The CPE induced by MCV was studied further by McFadden et al. (1979). A CPE occurred in several primate cell types (primary and continuous, simian and human fibroblasts or epithelial cells), but not in murine cells, and persisted for 10 days, after which the cells returned to normal and were able to respond to fresh challenge with MCV. Similar results were obtained when fresh human epidermis explants were used. Using immunofluorescence and electron microscopy, the authors demonstrated viral attachment and release of virion cores into the cytoplasm: however, cores were not uncoated and infection was thus aborted. Simultaneous infection with vaccinia virus, SFV or YMTV did not rescue MCV from the block to uncoating. The CPE was blocked by inhibitors of RNA and protein synthesis showing it to be due to virus early gene expression. RNA/DNA hybridisation was used to demonstrate transcription of MCV DNA from the intact cores.

The difficulties in culturing MCV, like those for the

human papillomaviruses, are thought to be due to the need for keratinocytes at a defined stage of differentiation which is difficult to obtain or sustain *in vitro*. In this respect, MCV behaves as a host-dependent conditional lethal mutant (Postlethwaite, 1970). The block in the infectious cycle *in vitro* is apparently completion of the uncoating process.

The MCV genome has been characterised by electron microscopy and shown to be a linear, duplex DNA molecule with a molecular weight of 118 x 10^6 Daltons (Parr *et al.*, 1977), equivalent to 179 kb. Denaturation studies revealed the presence of covalently closed termini, the genome being a single strand of self-complementary DNA. In both these respects MCV DNA is similar to that of the orthopoxvirus vaccinia (see section 1.2.3). However, there were differences in their denaturation profiles: the most easily denatured regions (ie. the most AT-rich) of MCV DNA were those near the ends, in contrast to vaccinia virus DNA. Overall, denaturation conditions equivalent to an additional 10 ^{o}C were required to denature MCV DNA to levels comparable to vaccinia virus DNA, indicating a greater GC-composition for the former.

Parr *et al.* (1977) studied DNA from ten isolates of MCV by agarose gel electrophoresis after digestion with the restriction endonuclease BamHI and obtained estimates of the genome size in agreement with that above. They described three different cleavage profiles, indicating three genetic subtypes of the virus, each distinct from that of orthopoxvirus DNA. There were six occurrences of subtype I (three

of these isolates were from the same patient), one of subtype II and three of subtype III. Subtypes II and III were similar to each other but distinct from the first and are probably variants of the same subtype. Shortly after the commencement of this project, Darai et al. (1986) reported the BamHI restriction endonuclease analysis of DNA from 14 MCV isolates from independent patients and described two viral subtypes on the basis of different cleavage profiles: I and II, respectively similar to the subtypes II/III and I described by Parr et al. They observed only one incidence of subtype II and established its close relationship to subtype I by DNA cross-hybridisation. There was an indication from these two reports that isolates from genital lesions were of subtype II (Darai's terminology) whilst those from lesions at other sites were predominantly subtype I. The latter study was subsequently expanded to 40 isolates, with a single incidence of MCV II (Reisner et al., 1986). No restriction mapping data were available for MCV DNA prior to the commencement of this project.

1.3 Epidermal growth factor and its homologues

Epidermal growth factor is one of several mammalian polypeptide growth hormones involved in the regulation of cell proliferation. There is a large family of homologous polypeptides, including the transforming growth factor-alpha and a poxviral gene product. The basic structural features of these molecules are also displayed by several blood clotting factors and non-mammalian developmental gene products.

1.3.1 Epidermal growth factor and its receptor

Epidermal growth factor (EGF) was initially described as a polypeptide hormone, isolated from the mouse submaxillary gland, that caused precocious eyelid opening and tooth eruption when injected into newborn mice (Cohen, 1962). Both effects are a consequence of enhanced epidermal growth and keratinisation. EGF also stimulates the proliferation of keratinocytes and fibroblasts *in vitro* (reviewed by Cohen, 1987). For his discovery and subsequent work Stanley Cohen was recently awarded the Nobel Prize. Mouse EGF is a 53 amino-acid polypeptide with three disulphide-bridged loops (Savage *et al.*, 1972; Savage *et al.*, 1973). The human homologue is the same size and is identical to the gastric anti-secretory hormone urogastrone (Gregory, 1975).

The major site of synthesis in the mouse is the submaxillary gland, where EGF is proteolytically cleaved by an

arginine esterase from a 128kD transmembrane precursor (Gray et al., 1983). The precursor possesses seven other EGF-like domains. It is also made in the kidney cortex but is not processed in this tissue (Rall et al., 1985). An extracellular portion of the precursor has 33% homology to the receptor for low density lipoprotein (LDL; Sudhof et al., 1985). The homologous regions are distinct from EGF and the LDL binding site. It has been postulated that the EGF precursor is a receptor for an unidentified ligand: this may explain its presence in kidney (Pfeffer and Ullrich, 1985).

The biological responses to EGF are mediated by a membrane receptor, a 170kD protein with an extracellular ligand binding domain and an intracellular domain possessing tyrosine kinase activity (reviewed by Hunter, 1984; Stoscheck and King, 1986; Carpenter, 1987). EGF binding promotes receptor dimerisation and activation of the kinase domain, substrates for which include lipocortin, which inhibits prostaglandin and leukotriene synthesis (phosphorylation abrogates inhibition), and the receptor itself (Yarden and Schlessinger, 1987). Other responses to EGF include enhanced Na^+/H^+ antiport activity, enhanced release of Ca^{2+} from intra- and extra-cellular stores, enhanced turnover of phosphoinositides (in A431 cells, but not in 3T3 cells) and induction of c-fos and c-myc gene expression: continued stimulation with EGF for 8 hours culminates in DNA synthesis after 12 hours (reviewed by Carpenter and Cohen, 1979; Schlessinger et al., 1983). Following binding, the ligand-receptor complex is internalised and proteolytically degraded; this processing

down-regulates the receptor and is necessary for induction of the late response of DNA synthesis (Schaudies et al., 1987). The kinase domain is crucial to the function of the EGF receptor since mutagenesis of its ATP binding site abolishes all early and late responses to EGF (Chen et al., 1987).

The product of the oncogene v-erbB, from the avian erythroblastosis retrovirus, is homologous to part of the EGF receptor and thought to be a truncated version of the chicken homologue: it lacks the N-terminal ligand binding and C-terminal autophosphorylation domains and is constitutively active (Lax et al., 1985). Elevated levels of the EGF receptor are associated with a high proportion of squamous cell carcinomas and have also been detected in other tumour types (Ozanne et al., 1986; Gullick et al., 1986). In normal epidermis, EGF receptors occur predominantly in the basal cell layer (Nanney et al., 1984).

1.3.2 Transforming growth factor-alpha

Following retroviral or chemical transformation, many cells lose their ability to bind EGF due to their secretion of a polypeptide, related to but antigenically distinct from EGF, which blocks receptor sites: transforming growth factor-alpha (TGF; De Larco and Todaro, 1978). In terms of its primary amino-acid sequence homology to EGF and its ability to bind to the EGF receptor with a similar affinity and elicit a similar response, TGF is structurally and functionally homologous to

EGF (Sporn and Todaro, 1980). TGF is a 50 amino-acid polypeptide and is cleaved from a 17kD transmembrane precursor (Teixido *et al.*, 1987). The enzyme responsible for cleavage is not that involved in processing the EGF precursor but recognises a specific alanine-valine-valine sequence (with cleavage after the alanine residue) at both ends of the mature TGF. The disulphide-bond arrangement of TGF is the same as that of EGF (Winkler, 1986).

In many biological systems, responses to EGF and TGF are identical: however, TGF is more potent at stimulating hair follicle development, bone resorption, angiogenesis and keratinocyte migration. It is synthesised early in mouse foetal development and is present throughout the epidermis of neonatal and adult human skin, where it is synthesised by keratinocytes (reviewed by Derynck, 1986 and 1988; Barrandon and Green, 1987; Coffey *et al.*, 1987). TGF is auto-induced and so must normally be subject to negative control. It may play a role *in vivo* in the response to wounding (Coffey *et al.*, 1987).

Inappropriate expression and secretion of TGF by transformed cells may act to stimulate their proliferation by an autocrine or paracrine mechanism, as has been demonstrated by artificial expression of TGF in fibroblasts *in vitro* (Rosenthal *et al.*, 1986). In agreement with this hypothesis, TGF mRNA was detected in many solid tumours that have elevated levels of the EGF receptor, such as squamous cell carcinomas (Derynck *et al.*, 1987). However, the levels of secreted TGF

do not always mirror the level of mRNA. Brachmann et al. (1989) used immunofluorescence to show the presence of unprocessed membrane forms of TGF on such cells. Site-directed mutagenesis to prevent cleavage of mature TGF and subsequent expression in cells lacking EGF receptors allowed the accumulation of membrane precursors of TGF which were shown to be able to bind and activate the EGF receptor following solubilisation. Also, co-incubation of cells expressing TGF precursors with EGF-responsive cells leads to receptor autophosphorylation, Ca^{2+}-fluxes and induction of c-fos expression in the latter in the absence of mature soluble TGF (Wong et al., 1989; Brachmann et al., 1989). Thus the TGF precursor is capable of eliciting several early responses of the mitogenic pathway. This may be relevant to the autocrine/paracrine hypothesis of tumourigenesis mentioned above and may indicate functional significance of the accumulation of unprocessed EGF precursor in some tissues (Rall et al., 1985). The possible involvement of such precursor forms in intercellular communication is discussed below (section 1.3.4).

1.3.3 Poxviral growth factors

During searches of sequence databases for entries homologous to rat TGF, a match was found to the vaccinia virus 19 kD protein (p19) (Blomquist et al., 1984; Brown et al., 1985; Reisner, 1985). The p19 gene is located within the genomic terminal repeat of vaccinia virus and encodes a 140

amino-acid polypeptide that is expressed early in infection (Venkatesan et al., 1982; see section 1.2.4). The primary sequence of p19 has potential N-terminal and internal membrane-spanning hydrophobic regions. It was subsequently shown that the conditioned medium from vaccinia-infected cells has a growth-stimulating activity that first appears 2 hours after infection (Stroobant et al., 1985; Twardzik et al., 1985). This activity was purified from the medium and its N-terminal sequence was determined, from which it was identified as a proteolytic fragment of the p19 gene product (Stroobant et al., 1985).

Based on its amino-acid composition, this vaccinia virus growth factor (VGF) is approximately 77 amino-acids long. VGF is N-glycosylated and has a molecular weight of 26 kD. Like EGF and TGF, VGF is cleaved from the external domain of a precursor inserted into the cell plasma membrane (Stroobant et al., 1985). However, the cleavage site differs from that for each of the former. EGF and TGF are smaller than VGF (50-53 amino-acids) and are not glycosylated: the VGF glyco- sylation site is outside the region homologous to EGF/TGF. VGF binds to the EGF receptor and elicits the same response in terms of receptor phosphorylation and induction of cell proliferation as does EGF itself (Stroobant et al., 1985). The affinities of EGF and VGF for the EGF receptor are comparable. Thus, VGF is structurally and functionally homologous to EGF/TGF.

During this project, Shope rabbit fibroma virus (SFV),

associated with a widespread proliferative disease in rabbits (see section 1.2.5), and the related myxoma virus, the agent of myxomatosis, were also shown to encode polypeptides analagous to VGF (SFGF and MGF respectively: Chang et al., 1987; Upton et al., 1987). SFGF and MGF are 80 and 85 amino-acids long respectively. Their sequences suggest the presence of an N-terminal signal but no transmembrane region. SFGF/MGF is expressed with a similar time-course to VGF. Lysates of cells infected with SFV or malignant rabbit fibroma virus, a recombinant between SFV and an orthopoxvirus (see section 1.2.5), possess an activity that binds to the EGF receptor and stimulates cell proliferation (Strayer and Leibowitz, 1987), although this has not been formally equated with SFGF. Recently, Lin et al. (1988) have synthesised the EGF-homologous regions of SFGF and MGF and demonstrated the same disulphide bonding as for EGF and 10% of the receptor binding and mitogenic activities of EGF, confirming that these poxviral growth factors, in addition to VGF, have the biological activities of EGF/TGF.

The possible functions of the poxviral growth factors are discussed in section 1.4.

1.3.4 Sequence homologies of EGF-like polypeptides

Table 2 shows the aligned sequences of the nine members of the EGF-like growth factor family for which sequence data are available. In addition to the overall conservation of cysteine and glycine residues there is a striking level of conservation of amino-acids within the third disulphide-loop (EGF residues 33-42; referred to in subsequent chapters as domain A), implying an important functional role for this region of the molecule (see section 1.3.5). There is also a region of the first disulphide-loop that is conserved between EGF and VGF but not TGF or SFGF/MGF (EGF residues 11-15; domain B). Although EGF, TGF and the poxviral growth factors are structurally and functionally homologous, there is no significant homology at the nucleotide level between the genes encoding them. Neither is there nucleotide homology between the VGF and SFGF genes despite both being poxviral genes, albeit from viruses in different genera: in fact the level of homology of SFGF and MGF to VGF is only 36.6% and 34.1% respectively, lower than to EGF and lower than the 65.5% homology between the vaccinia virus and SFV thymidine kinases (Upton and McFadden, 1986).

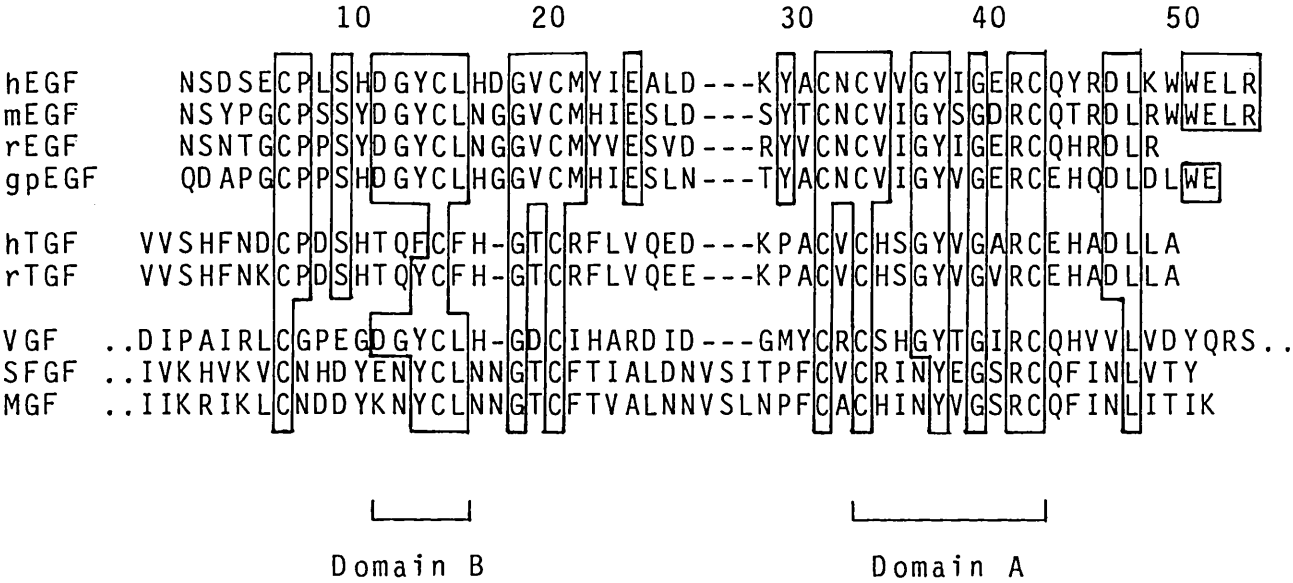


Table 2. The EGF-like growth factor family. The sequences of EGF, TGF and the growth factors from vaccinia (VGF), Shope rabbit fibroma (SFGF) and myxoma (MGF) poxviruses are aligned. Species designations are human (h), mouse (m), rat (r) and guinea-pig (gp). Numbering refers to amino-acid positions in EGF. Residues boxed are conserved amongst all EGF species. Two conserved domains (A and B, see text) are indicated. The N-terminal portions of the viral growth factors and five C-terminal amino-acids of VGF (ENPNT) are omitted. Data are taken from Simpson et al. (1985) and references therein; Chang et al. (1987); Upton et al. (1987). The single letter amino-acid code is used: A (alanine), C (cysteine), D (aspartate), E (glutamate), F (phenylalanine), G (glycine), H (histidine), I (isoleucine), K (lysine), L (leucine), M (methionine), N (asparagine), P (proline), Q (glutamine), R (arginine), S (serine), T (threonine), V (valine), W (tryptophan), Y (tyrosine).

The presence of EGF-like domains in other proteins has also been noted: these include several mammalian serum coagulation proteins (clotting factors IX, X, XII, C, S and Z, tissue plasminogen activator and the laminin B1 chain) and three non-mammalian developmental proteins thought to be involved in intercellular communication and developmental decisions (products of the Drosophila melanogaster Notch gene, the Caenorhabditis elegans lin-12 gene and the sea urchin Strongylocentrotus purpuratus uEGF gene). These sequences are aligned and referenced in table 3. The developmental proteins respectively possess 36, at least 11, and 9 tandem repetitions of this EGF-like domain, reminiscent of the structure of the EGF precursor itself: one such repeat of the Notch product contains all of the conserved residues of domain A and other residues thought to be important for structure and function and so may be able to bind to the Drosophila homologue of the EGF receptor. The observation that membrane-bound TGF precursor can bind to and activate EGF receptors on adjacent cells (see section 1.3.2) adds credence to the possible role of the developmental proteins with EGF-like domains in intercellular communication.

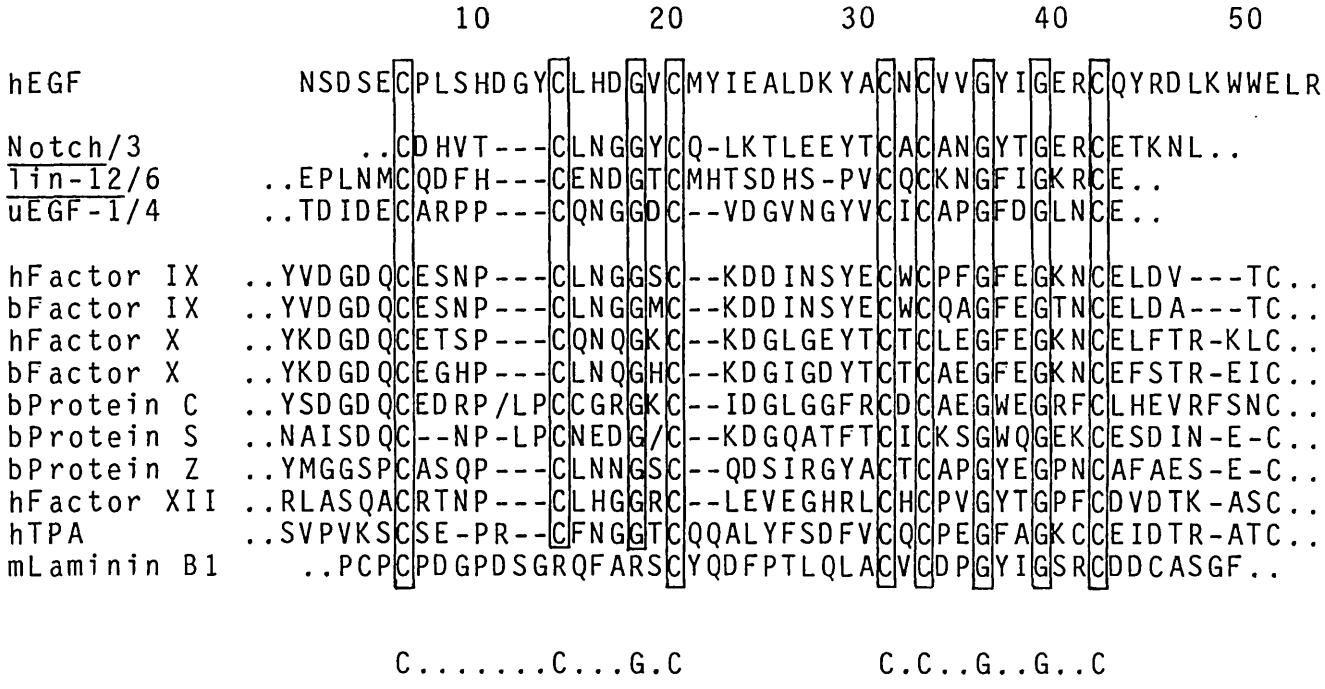


Table 3. Other EGF-like polypeptide sequences. The EGF-homologous regions of the sequences of the laminin B1 chain, tissue plasminogen activator (TPA), clotting factors IX, X, XII, C, S and Z, and the developmental genes Notch (third repeat), lin-12 (sixth repeat) and uEGF-1 (fourth repeat) from Drosophila melanogaster, Cyanorhabditis elegans and Strongylocentrotus pupuratus respectively are aligned with human EGF. For sequences possessing several EGF-like repeats, that most closely resembling the third disulphide domain of EGF is presented. Species designations for the coagulation proteins are mouse (m), human (h) and bovine (b). The symbol "/" in the sequence of bovine protein C represents an insertion of six residues relative to EGF (SGSPCD) and that in the sequence of bovine protein S an insertion of three residues (FMT). The conserved cysteine and glycine residues are boxed and indicated below as two sequence motifs. Data are taken from Blomquist et al. (1984) and references therein; Hojrup et al. (1985); McMullen and Fujikawa (1985); Greenwald (1985); Wharton et al. (1985); Dahlback et al. (1987); Sasaki et al. (1987); Hursh et al. (1987). The single letter amino-acid code is used (see legend to table 2).

1.3.5 Structure-function studies of the EGF-like polypeptides

Heath and Merrifield (1986) used a synthetic approach to study the structure-function relationship of mouse EGF. Synthetic EGF was as active as natural EGF for receptor binding and stimulation of DNA synthesis. Oxidation of methionine-21 led to a 30% reduction in activity. EGF-[15-53] (ie. residues 15-53 of EGF) was 0.01% as active as EGF, EGF-[32-53] was 0.001% as active and EGF-[43-53] was inactive. These studies suggested that the central part of the EGF molecule is important for binding. However, the low activity of all fragments implicates residues 1-14 in determining and/or forming part of the receptor binding surface. Size differences between the EGF-like growth factors at their C-termini (see table 2) and chymotryptic digestion of EGF suggests that the last five residues of EGF are not involved in binding. Also, removal of the N-terminal residues 1-3 by site-directed mutagenesis does not reduce EGF activity (Burgess *et al.*, 1988).

Synthetic peptides have also been used in attempts to identify regions of the EGF molecule responsible for receptor binding. A peptide corresponding to human EGF residues 20-31 binds the receptor with an affinity of 0.003% that of EGF itself and also elicits receptor autophosphorylation (Komoriya *et al.*, 1984). A peptide corresponding to residues 34-43 of TGF (equivalent to residues 33-42 of EGF) has an affinity of 2% that of EGF or TGF but no other activity (Nestor *et al.*, 1985). These peptides occur in the two regions of beta-sheet

(see below): both bind antibody raised to the whole molecule suggesting that they correspond to surface epitopes. Sequence comparison of these regions with the other EGF-like growth factors indicates a high level of conservation associated with the region corresponding to the latter peptide (domain A) whilst there is no conservation in the former region, which even differs in size for SFGF/MGF, with the exception of the cysteine residues (see table 2). Assuming that sequence conservation reflects functional conservation, this argues in favour of the third disulphide-loop (domain A) mediating binding of the EGF-like growth factors to the EGF receptor. Additional evidence in favour of this is that EGF cleaved with cyanogen bromide after methionine-21, which disrupts the region corresponding to the peptide synthesised by Komoriya et al., still binds with 10% of the affinity of EGF and induces some early responses, although it is not mitogenic (Yarden et al., 1982).

The solution structures of mouse EGF and human EGF have been determined by proton nuclear magnetic resonance. Both form two regions of disulphide-bridged anti-parallel beta-sheet which divide the molecule into two structural domains: residues 1-32 and 33-53 (figure 2) (Montelione et al., 1986; Cooke et al., 1987). Human TGF has a similar structure with the major differences occurring at the N- and C-termini (Montelione et al., 1988; Campbell et al., 1989). This structure could be adopted by all members of the EGF-family: all possess the six cysteine residues involved in disulphide-bridge formation and such variations in size that

occur (ie. relative insertions and deletions) are positioned in regions of the molecule not involved in beta-sheet formation (eg. the insertion of three amino-acids in SFGF relative to EGF positions 27-28 occurs in the loop between beta-sheet strands). Glycosylation of VGF also does not disrupt the structure. A hybrid growth factor consisting of the first 33 N-terminal residues of human TGF and VGF residues corresponding to the next 20 positions has biological activities comparable to TGF and VGF, indicating that structural domains can be interchanged between members of this growth factor family and are essentially equivalent (Purchio et al., 1987).

The structure-function relationships of EGF and TGF have been further studied by in vitro mutagenesis. Substitution of each of the pairs of cysteine residues involved in disulphide-bonding showed that all three disulphide-loops of human TGF are necessary for receptor binding: additionally, formation of the third disulphide-loop determines the correct formation of the other loops (Harkins and Brandis, 1988). Mutation of human EGF positions 47 (leucine), 37 (tyrosine) and 41 (arginine), all absolutely conserved in the EGF growth factor family, reduced binding to 8%, 0.6% and 0% of wild-type EGF respectively, whilst mutation of the non-conserved positions 7, 24, 27 and 29 resulted in 30-95% binding (Engler et al., 1988a and 1988b). Mutagenesis of human TGF confirmed the importance of the conserved leucine corresponding to EGF position 47 and of the preceding aspartate residue (Lazar et al., 1988).

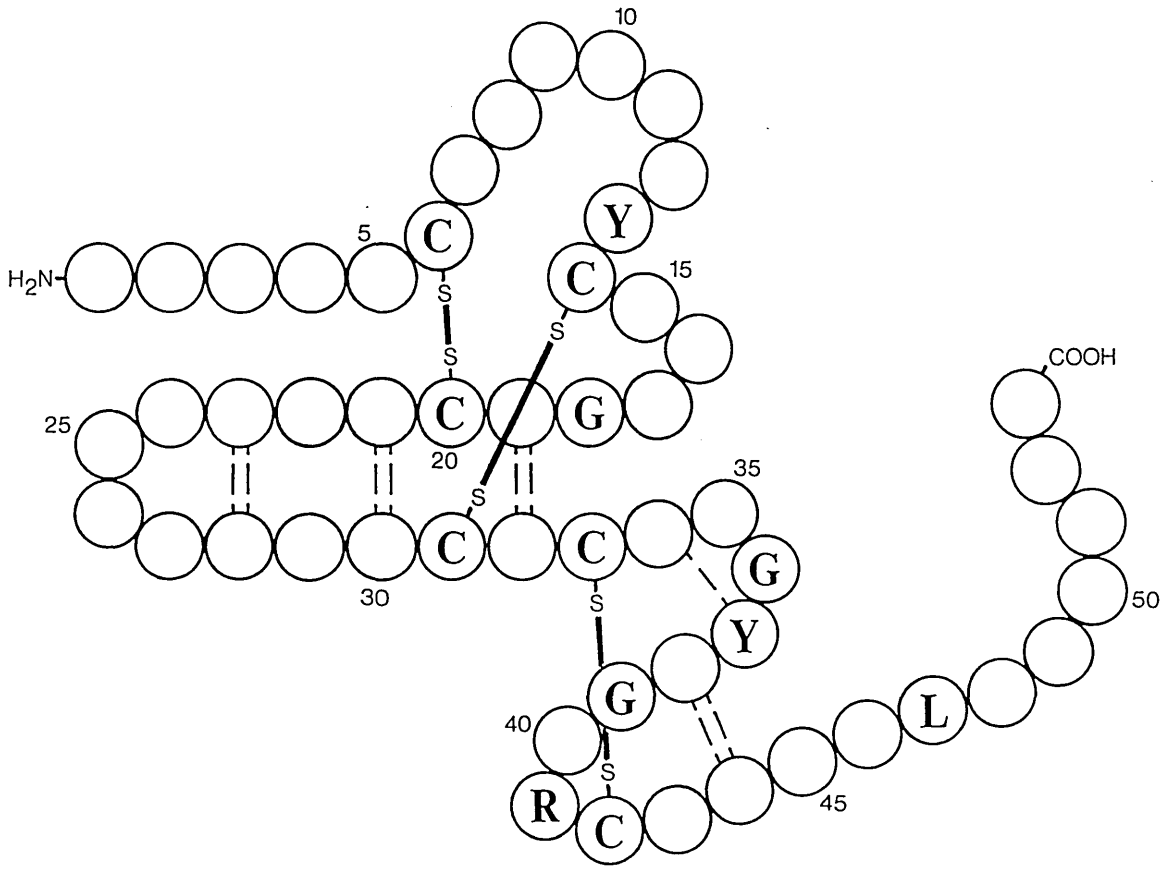


Figure 2. Schematic structure of human EGF as determined by solution nuclear magnetic resonance. Adapted from Cooke et al. (1987). The disulphide bonds (heavy lines) and hydrogen bonds (broken lines) involved in beta-sheet and beta-turn formation are indicated. The residues identified are those highly conserved amongst members of the EGF-like growth factors (refer to table 2).

Campbell et al. (1989) have considered structural information for EGF and TGF obtained by nuclear magnetic resonance in combination with sequence comparison data, arguing that conserved, or conservatively changed, residues will either have a structural role in the formation of active growth factor or will contribute directly to binding to the EGF receptor. Those conserved residues with non-structural roles as determined from the solution structures are thought to contribute to the binding surface: namely EGF residues 13, 15, 16, 41, 43 and 47. Although these fall within both beta-sheet structural domains they form one face of the molecule, consistent with a role in receptor binding. The formation of the putative binding face from non-contiguous residues explains the limited and sometimes seemingly contradictory results obtained with short synthetic peptides of EGF and TGF.

It is to be expected that other growth factors belonging to the EGF-family would possess the conserved features of the existing members. The region of greatest conservation comprises the third disulphide-loop (domain A in table 2). At the beginning of this project, without the more recent information from structure determination and mutagenesis analyses, it seemed most likely that this conserved sequence was an important determinant of EGF activity and hence the best contiguous structural feature by which to identify a new member of the EGF-family.

1.4 Function of vaccinia virus growth factor; possible mechanisms of poxvirus-induced tumourigenesis

The structural and functional homology of vaccinia virus growth factor (VGF) to EGF was discussed above (section 1.3.3). VGF appears in the medium of cells *in vitro* 2 hours after infection with vaccinia virus and reaches maximal levels after 12 hours (Twardzik *et al.*, 1985). EGF receptor autophosphorylation occurs 4.5 hours after infection (King *et al.*, 1986). Antibody against the receptor ligand binding site abolishes this effect, as well as the effects of exogenous VGF, indicating that VGF acts at the cell surface and not by intracellular association with the receptor before its appearance on the cell surface.

One suggestion of the possible role of VGF was as the structural component of the virus particle responsible for cell adsorption. However, such a role seems unlikely in the light of the following information. VGF is expressed early in infection, whereas most structural proteins are synthesised late, and the mature form is processed from a transmembrane precursor and released from infected cells (Stroobant *et al.*, 1985). Saturation of EGF receptors with EGF or complexing with antibodies to the receptor ligand binding site causes a 60% decrease in the infectivity of vaccinia virus for L cells *in vitro* (Eppstein *et al.*, 1985; Marsh and Eppstein, 1987): this decrease is too small to support a role in adsorption. Also, vaccinia virus can infect a cell line lacking EGF receptors (NR6), arguing against a role for VGF in mediating

viral attachment (Stroobant et al., 1985). Finally, no VGF was detected on purified vaccinia virus using an EGF receptor binding assay (Twardzik, personal communication to Buller et al., 1988a).

The function of VGF has been studied further by creating a vaccinia virus mutant in which tha VGF gene within both terminal repeats was partially deleted by homologous recombination (Buller et al., 1988a). As expected, this VGF-minus (VGF^-) mutant does not direct the synthesis of VGF which is present in the conditioned medium of cells infected with wild-type virus. VGF^- and wild-type viruses infect a number of cell lines, in both active and resting metabolic states, with the same efficiency: this is true regardless of the level of expression of EGF receptors. Complexing the EGF receptor ligand binding site with antibody did give a reduction in infectivity, but to the same extent for VGF^- and wild-type viruses. These data argue against a requirement for VGF and/or the EGF receptor for infection in vitro, consistent with the work discussed above.

Buller et al. (1988a) have also investigated virus yield upon infection of active or resting BS-C-1 cells, which have moderate EGF receptor levels, and Swiss 3T3 cells, which are highly responsive to EGF. VGF^- and wild-type viruses have similar yields in active cells: both give fewer progeny from resting cells, but in this case the yield from VGF^- virus is less than that from wild-type virus, the difference being greatest for the highly EGF-responsive 3T3 cells. These data

indicate that replicative efficiency (but not infectivity) is dependent upon the metabolic activity of the cells to be infected and support the idea that VGF functions as a generalised stimulator of cellular metabolic activity in order to "prime" cells for more efficient virus production.

VGF$^-$ vaccinia virus exhibits an attenuated phenotype *in vivo* (Buller *et al.*, 1988a). The mutant virus has lower infectivity and a higher lethal dose than wild-type virus following intracranial inoculation of mice, despite the formation of histologically identical lesions. Attenuation (10- to 100-fold) was also demonstrated for foot-pad inoculation of mice and intradermal inoculation of rabbits. The limited proliferation of epidermal cells that occurs at the periphery of lesions in the latter assay was lower for the VGF$^-$ virus. Attenuation presumably reflects the inability of replicating virus to activate neighbouring uninfected cells, with a resultant lower yield when these cells do become infected and thus a slower spread of virus. These data indicate that, whilst expression of VGF is not essential for virus replication *in vitro*, it does contribute to the virulence of vaccinia virus.

Growth of vaccinia virus *in vitro* and *in vivo* is associated with a limited proliferative response by cells at the lesion periphery, for which VGF may be responsible (see above). Wild-type vaccinia virus grown in A431 cells (which possess greatly elevated levels of the EGF receptor and were chosen to maximise the potential proliferative response to

VGF) rapidly induces the formation of cell foci, in contrast to the production of plaques by the VGF$^-$ mutant (Buller *et al.*, 1988a and 1988b). Antibody directed against the ligand binding site of the EGF receptor abolishes focus formation and plaques form instead. Similar differences were observed for the limited proliferation induced by infection of chorioallantoic membranes. These data are consistent with the induction of cell proliferation mediated by activation of the EGF receptor by VGF released from infected cells. Immunochemical staining of the virus-infected chorioallantoic membrane revealed that at early stages virus antigen is confined to the centre of the lesion and that the majority of proliferating cells are not infected - ie. proliferation occurs prior to infection (Buller *et al.*, 1988b). Following infection, virus replication leads to shut-down of host cell synthetic activites and the cell eventually dies. Lesions due to VGF$^-$ virus are not proliferative, contain less antigen and release fewer progeny.

Taken together, the above data suggest that the *in vivo* role of VGF is to enhance the metabolic activity of cells neighbouring those already infected and so create a more favourable basis for subsequent viral replication in the former, rather than to induce cell proliferation *per se*. The vaccinia virus infection cycle is ultimately cytolytic and so the proliferative response is localised to the advancing edge of the lesion. A similar mechanism could account for the more marked response to infection with the tumourigenic poxviruses, for which proliferation is not masked by cytolysis

(see section 1.2.5). Consistent with this suggestion, the proliferating cell types are in most cases epithelial cells or fibroblasts, which are known to possess EGF receptors. The exception to this is in the case of Yaba monkey tumour virus, for which the proliferating cell type is the histiocyte (Sproul *et al.*, 1963) of the monocyte/macrophage lineage, which is not known to express the EGF receptor. If the underlying mechanism is the same for each of these poxviruses, then presumably the tissue and cell type tropisms of the different tumourigenic poxviruses, and thus the location and nature of the tumours they induce, are determined by other species-specific factors. The degree of proliferation would be expected to depend on the responsiveness of the host cell to stimulation via the EGF receptor as well as the affinity of the virus-coded growth factor for the receptor, and on the rate of virus replication and degree of associated cell death. The tumourigenic Shope rabbit fibroma virus, and the related myxoma virus, have been shown to encode VGF homologues, consistent with the above hypothesis (see section 1.3.3).

As discussed above (section 1.1.1), the benign skin tumour induced by *Molluscum contagiosum* virus (MCV) results from the enhanced proliferation of basal cells whilst virus replication occurs only in the overlying layers of epidermis. Within the epidermis, it is the basal cells that possess the greatest numbers of EGF receptors and it is this cell layer that has the proliferative potential, the higher layers being committed to differentiation. Based on the above analysis of the role of VGF, a mechanism for MCV-induced tumourigenesis

can be advanced, in which virus growth in differentiating keratinocytes directs the secretion of an EGF-like growth factor (a VGF-homologue) which acts to stimulate proliferation of the underlying basal cells and so increase the number of cells susceptible to infection (figure 3). Enhanced proliferation is localised and reversible upon cessation of virus growth. This paracrine mechanism of tumourigenesis contrasts with that of the classical DNA tumour viruses (eg. the papillomaviruses), for which it is the infected cells themselves that show enhanced proliferation.

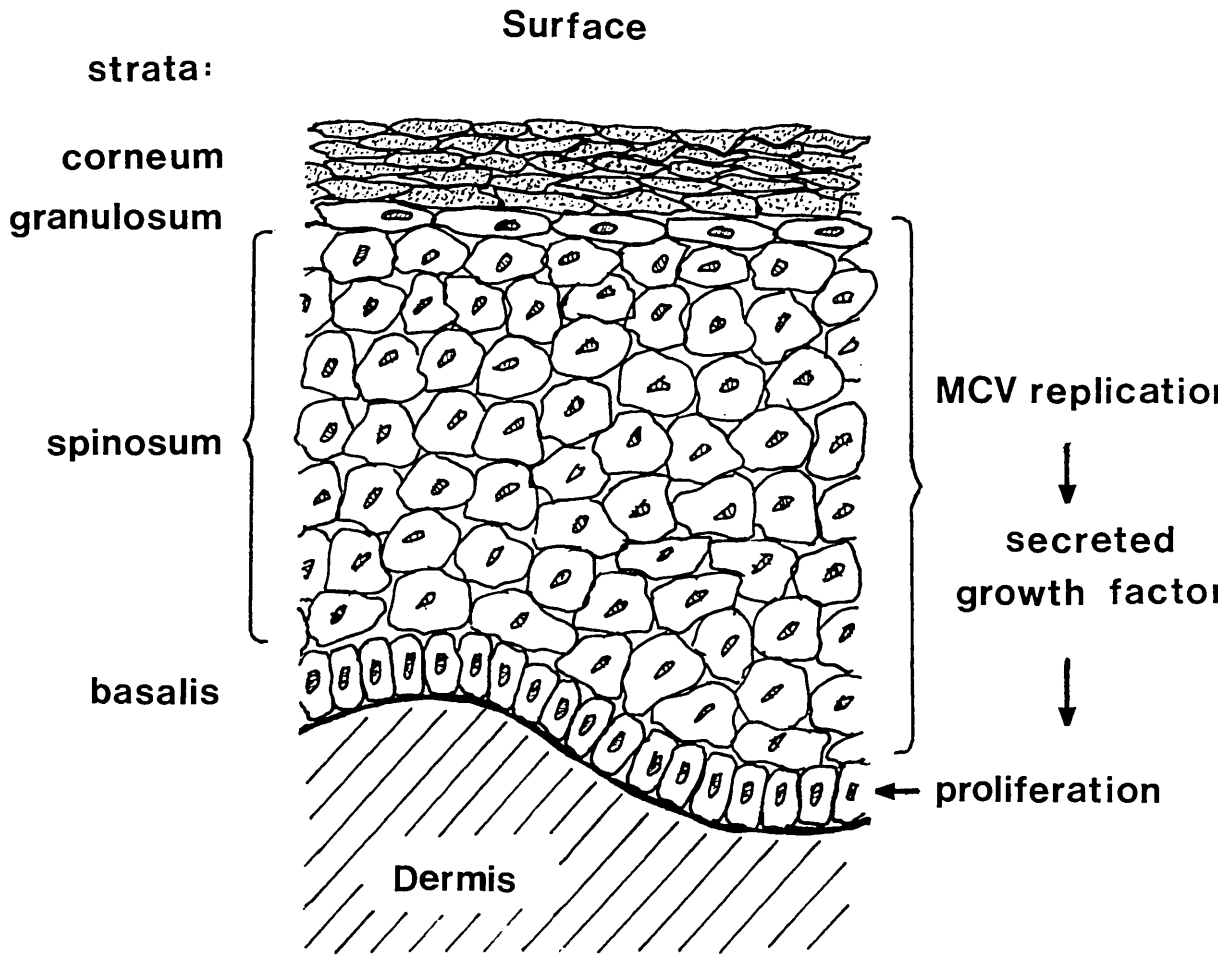


Figure 3. Proposed mechanism of Molluscum contagiosum virus-induced tumourigenesis. Virus replication occurs in the upper layers of the epidermis (stratum spinosum and stratum granulosum) and a secreted virus-encoded growth factor enhances proliferation of stem cells in the underlying stratum basalis.

CHAPTER 2

Aims of this project

A scheme was proposed in section 1.4 to account for the basal cell proliferative response to infection of the human epidermis with the poxvirus Molluscum contagiosum. The scheme postulates release from infected cells of a growth factor that stimulates the underlying stem cells. Support for the existence of such a factor comes from the recent identification and characterisation of a polypeptide released from cultured cells infected with vaccinia virus. This polypeptide (vaccinia virus growth factor) is structurally and functionally homologous to a mammalian peptide growth factor, epidermal growth factor, the prototype of a large and growing family of related growth factors.

The long-term aim, which this project was intended to initiate, was to determine whether a similar polypeptide is responsible for Molluscum contagiosum virus-induced tumourigenesis. Because the virus can not be grown in vitro and therefore such a putative factor can not be assayed directly, the nature of the investigation was to determine whether its coding potential is compatible with the proposed scheme. Thus, the intention was to clone the putative homologue of the vaccinia virus growth factor gene.

The specific aims of this project were two-fold:-

(1) The initial aim was to provide a structural characterisation of the genome of Molluscum contagiosum virus in terms of physical maps of restriction endonuclease recognition sites so as to facilitate subsequent cloning studies. This work was also expected to be a worthwhile contribution to comparative molecular poxvirology and to the investigation of the structural diversity and epidemiology of this virus.

(2) The main aim was to determine whether the Molluscum contagiosum virus genome encodes a polypeptide with structural and functional homology to epidermal growth factor. The means of addressing this problem was to use synthetic oligonucleotide probes, designed to identify DNA sequences capable of encoding conserved features of the family of growth factors related to epidermal growth factor, to identify regions of the viral genome suitable for sequence analysis.

CHAPTER 3

Materials and methods

3.1 Materials and reagents

Suppliers of materials and reagents are listed in appendix 1.

3.2 Water; sterilisation of liquids, plastics and glassware

The source of pure water used throughout this project was obtained by reverse osmosis, passage through a 0.22 um filter and de-ionisation: it had a conductivity of 0.05 micro-Sievert (20 Mega-ohm/cm). Liquids were sterilised by autoclaving for 15 min at 120 ^{o}C and 15 pounds per square inch above atmospheric pressure. Solutions and buffers whose contents were unstable to heat were sterilised by passage through a 0.22 or 0.45 um filter. Plastics (polypropylene pipette tips and reaction tubes) and glassware were sterilised by autoclaving.

3.3 Source of clinical material

Skin lesions diagnosed as Molluscum contagiosum were obtained from 105 patients presenting to the Departments of Dermatology and the Genito-Urinary Clinics at the Charing Cross Hospital and West Middlesex Hospital, and to the Hillingdon Genito-Urinary Clinic. Patients with, or at risk of, human immunodeficency virus infection were excluded from the study. The study group included 28 males and 77 females and covered the age range 2 to 70 years. Of these patients, 61 presented to the hospital departments whilst 44 (only females, for the reason mentioned above) presented to the Genito-Urinary Clinic, accounting for the sex bias in the group as a whole. Details of patients from whom lesions were obtained are given in appendix 2. Some patients presented on more than one occasion: the total number of samples processed was 124.

In the majority of cases the lesions were removed by curettage; however, in some instances the lesion core, which is rich in virus particles, was expressed. Samples consisted of between one and twenty lesions removed at a given presentation. When these were clustered and might reasonably be assumed to have arisen from the same infection the lesions were pooled before processing: when lesions were removed from separate sites at the same presentation they were processed independently. Samples were stored dry at 4 ^{o}C prior to processing.

3.4 Processing of Molluscum contagiosum lesions

Lesions were processed in one of two ways: either for preparation of total DNA (ie. viral DNA and cellular DNA) or for purification of intact virus particles prior to preparation of viral DNA. Initial processing consisted of finely chopping the lesions in either TE (for preparation of total DNA) or 1 mM Tris.Cl pH 9.0 (for purification of intact virus particles). Processing as far as the proteinase K digestion step was done in a Class 1 safety cabinet (Leec).

3.4.1 Virus purification

The chopped lesions were homogenized in 1 mM Tris.Cl pH 9.0. The homogenate was layered on 10% dextran T10 above a column of 36% sucrose in the same buffer and virus was pelleted by centrifugation in a 3 x 3 ml swing-out rotor using an MSE50 centrifuge at 16,000 rpm (16,000 g_{av}) for 1 h. The Molluscum contagiosum virus (MCV) pellet was resuspended, purified further by repeating the centrifugation step and resuspended in 100 ul TE. An equal volume of 2x lysis buffer was added, giving a final concentration of 100 mM Tris.Cl pH 7.5, 10 mM EDTA, 70 mM 2-mercaptoethanol, 0.1% SDS, 1% sodium N-lauryl sarcosinate. Proteinase K was added to a final concentration of 0.5 mg/ml and the sample was incubated at 55 ^{o}C for approximately 2 h, until clear.

3.4.2 Total DNA preparation

The chopped lesions were suspended in 30-200 ul TE (the volume depending on the amount of material) and an equal volume of 2x lysis buffer (see above) was added. Proteinase K was added to a final concentration of 0.5-1.0 mg/ml and the sample was incubated overnight at 55 ^{o}C. By morning it was essentially clear, although some undigested particulate matter often remained.

3.4.3 DNA purification

Following lysis and digestion, the purified-viral DNA or total DNA was deproteinized by extraction successively with phenol equilibrated with TE, phenol:chloroform:amyl alcohol (25:24:1), and chloroform:amyl alcohol (24:1). Samples being processed for total DNA that contained a lot of undigested particulate matter were extracted an additional two or three times at the phenol stage. All pipetting of the aqueous layer (containing the DNA) was done using pipette tips from which the end had been cut in order to create a wider bore to minimise shearing of large DNA molecules. The DNA was precipitated by the addition of 0.1 volumes of 3 M sodium acetate pH 5.2 and 2.5 volumes of ethanol, and gentle mixing by inversion. When the DNA was of good quality the precipitate was easily visible, forming a cotton wool-like structure as it collapsed on itself. It was removed on the end of a pipette tip, washed in 800 ul 70% ethanol and dried under vacuum for 5 min in a Speed Vac (Savant). The DNA was

dissolved in TE to an estimated concentration of 0.1-1 mg/ml and stored at 4 ^{o}C.

A sample of eg. 5 lesions typically yielded 2-5 ug of viral DNA as estimated from the intensity of ethidium bromide staining of DNA digests compared to markers of known concentration. In exceptional cases, up to 50 ug of DNA was obtained from a single sample. DNA obtained from a single sample is referred to as an "isolate".

3.5 Restriction endonuclease digestion of DNA

The required amount of DNA was digested with the appropriate restriction enzyme in 10 mM Tris.Cl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 100 ug/ml BSA and 0, 50, 100, or 150 mM NaCl (for low, medium, high or extra-high salt buffers respectively). One of these standard buffers was used for all enzymes with the exception of SmaI, for which 10 mM Tris.Cl pH 8.0, 10 mM $MgCl_2$, 4 mM DTT, 100 ug/ml BSA and 20 mM KCl was used. Following mixing, the reaction was incubated at the appropriate temperature for the enzyme being used. Enzyme and buffer combinations used were ApaI (low), BamHI (high), ClaI (medium), EcoRI (high), HindIII (medium), PstI (medium), SalI (extra-high), Taq I (low) and XhoI (extra-high). Typically, 2-5 units of enzyme were used per ug of DNA to be digested and incubation was for 1-2 h. The amount of enzyme added never exceeded a tenth of the final volume, in order to keep the final glycerol concentration below 5% (enzymes were supplied

in 50% glycerol). All reactions were incubated at 37 ^{o}C with the exception of those using SmaI (30 ^{o}C) or TaqI (65 ^{o}C). Sometimes double-digests were required: these were either performed simultaneously, if the buffers were compatible, or sequentially. In the latter case the first digest was extracted with phenol:chloroform:amyl alcohol (25:24:1) and the DNA was precipitated by the addition of 0.1 volumes of 3 M sodium acetate pH 5.2 and 2.5 volumes of ethanol, and recovered by centrifugation at 13,000 g prior to digestion with the second enzyme.

3.6 Agarose gel electrophoresis of DNA

Following digestion of DNA (section 3.5), 0.1 volumes of gel loading buffer (50% glycerol, 1 mM EDTA, 0.4% bromophenol blue, 0.4% xylene cyanol) was added. Restriction enzyme fragments were separated by agarose slab gel electrophoresis in a submarine electrophoresis apparatus (GNA-200 or GNA-100; Pharmacia) using an EPS 500/400 power supply (Pharmacia). The running buffer was 40 mM Tris base, 20 mM sodium acetate, 2 mM EDTA. Gels of 0.6-2% agarose in running buffer were used, the concentration depending on the required size fractionation range. Samples were run into the gels at low voltage. Electrophoresis was usually overnight, at approximately 2 V/cm for 20 x 20 cm gels; smaller checking gels were often run faster, at upto 10 V/cm. Molecular weight markers used were digests of bacteriophage lambda DNA: usually HindIII or PstI, but also ApaI or XhoI to provide high

molecular weight markers. Following electophoresis, gels were stained in running buffer containing 0.5 ug/ml ethidium bromide for 30 min; destaining, if required, was in running buffer alone. Stained DNA was visualised with ultra-violet light (wavelength 312 nm) using a transilluminator (VLTF-20M; Vilber Lourmat). Gels were photographed using FP4 film and a vertical camera set-up (Plan Polaroid MP4 Land camera) fitted with filters to exclude ultraviolet light and pass the ethidium bromide fluorescence. FP4 film was processed using Autophen or Ilfotec-HC liquid developers and Hypam fixer.

3.7 Field inversion gel electrophoresis of intact MCV DNA

Intact MCV DNA (approximately 0.5 ug in 10 ul total volume containing gel loading buffer) was electrophoresed through a 0.7% agarose gel made up and run in 0.5x TBE. A Pharmacia GNA-200 submarine electrophoresis tank, modified by replacement of the standard electrodes with 0.376 mm gauge platinum wire (Gallenkamp) was used. The gel (20 x 20 cm) was pre-electrophoresed for 30 min at 60 V. Following sample application, the gel was run for a further 30 min at 60 V before initiating field inversion; the experiment continued for 16 h at 170 V. Buffer was circulated from the cathode to the anode compartment throughout the electrophoresis run by means of a peristaltic pump: it was kept at a constant ambient temperature by the passage of water beneath the gel platform in the electrophoresis tank.

Field inversion was achieved by means of a home-made switching device wired between the power supply (EPS 500/400; Pharmacia) and gel tank, the timing being controlled by software running on a BBC microcomputer (Porter et al., 1988). The computer program provided facilities for varying the forward and reverse intervals during the electrophoresis run, as was described by Carle et al. (1986). A forward interval starting at 0.1 sec and increasing to 10 sec at the end of the 16 h run time and a constant ratio of 3:1 for forward:reverse intervals were selected, having previously been found to optimise separation in the size range 50-300 kb.

The molecular weight markers used were oligomers of bacteriophage lambda DNA, the unit size of which is 48,502 basepairs (lambda c1857ind1sam7 DNA; it was from an unfrozen preparation to minimise shearage, since good quality DNA is necessary for efficient oligomerisation). Lambda DNA oligomers, formed by virtue of the 12 basepair "sticky ends" of the unit length DNA, were made by incubating lambda DNA at a concentration of 200 ug/ml in 2x SSC, 3% Ficoll 70 for 30 min at 37 ^{o}C, followed by 24 h at room temperature, and stored at 4 ^{o}C (Anand, 1986). A loading of 0.5-1.0 ug lambda DNA was used. To minimise shearing, both sample and marker DNAs were pipetted into the gel wells using wide-bore pipette tips. Following electrophoresis, the gel was stained in 1 ug/ml ethidium bromide in 0.5x TBE for 2 h and destained in 0.5x TBE for 2 h. The gel was transilluminated and photographed as above (section 3.6).

3.8 Southern transfer of DNA

Prior to transfer by Southern blotting (Southern, 1975; Maniatis et al., 1982) gels were soaked for 1 h in 1.5 M NaCl, 0.5 M NaOH to denature the DNA, followed by 1 h in 1.5 M NaCl, 1 M Tris.Cl pH 8.0 to neutralise the gel. The gel was placed on a wick of 3MM filter paper on a support; the wick was soaked in and in contact with 20x SSC as transfer buffer. A piece of membrane, either nitrocellulose (0.45 um pore size) or nylon was placed above and in contact with the gel, ensuring the absence of air bubbles. Nitrocellulose membranes were pre-wetted by soaking in 2x SSC before application to the gel. Two pieces of 3MM paper and a stack of absorbent paper towels placed above ensured a constant upward flow of transfer buffer by capillary action, thus drawing DNA fragments from the gel and onto the membrane. Following overnight transfer, the membrane was removed, rinsed in 2x SSC and air dried. DNA was immobilised by baking for 2 h at 80 ^{o}C for nitrocellulose membranes, or by wrapping in cling-film and placing on an ultra-violet transilluminator (DNA side down) for 5 min for nylon membranes. Southern blots were stored at 4 ^{o}C until required, sandwiched between sheets of 3MM paper and wrapped in cling-film.

3.9 Preparation of oligonucleotides

Eleven different degenerate oligonucleotides were synthesised during this project. Their designs are described in chapters 6, 7 and 8, and their sequences are listed in table 4. The amino-acid sequences against which they were directed were one of two conserved domains of the EGF-family of growth factors (see section 6.2).

3.9.1 Automated oligonucleotide synthesis

Oligonucleotides (ie. oligodeoxyribonucleotides) were automatically synthesized by the phosphoramidite version of the phosphite triester method (Atkinson and Smith, 1984) using a Gene Assembler (Pharmacia Biotechnology). Synthesis was from the 3'- to the 5'-end, the 3'-most base being attached to a column support (packaged into a cassette) via its 3'-hydroxyl group. The solid support was of unswellable 10 um monodisperse plastic beads. The first deoxyribonucleoside (as its 3'-O-succinate) was attached via an amide linkage to a dodecyl spacer coupled to the support. This first base and subsequently added bases were protected at their 5'-hydroxyls by coupling to a dimethoxytrityl (DMTr) group. The reactive centres of the purine and pyrimidine bases were also protected (see below). Synthesis consisted of sequential passage of reagents through the cassette according to a standard routine programmed with the required oligonucleotide sequence. Reagents were delivered to the cassette by means of a peristaltic pump. Each synthetic cycle consisted of the

following steps, between which the support was washed with acetonitrile to remove excess reagents:

(i) The 5'-hydroxyl group of the terminal deoxyribonucleotide was deprotected (detritylated - ie. the DMTr group was removed) by treatment with 3% dichloroacetic acid in ethylene dichloride.

(ii) The deprotected 5'-hydroxyl was condensed with the phosphoramidite group at the 3' position of an incoming 5'-protected deoxyribonucleotide in the presence of tetrazole as catalyst, forming a phosphite triester bond.

(iii) The phosphite triester (oxidation state III) was oxidised to the phosphate triester (V) with iodine and collidine.

(iv) Those 5'-hydroxyls which had failed to react during the coupling reaction were capped by acylation with acetic anhydride, thus preventing their taking any further part in synthesis.

The final product thus consisted of the full-length oligonucleotide and a series of shorter capped oligonucleotides bound to the solid support and still coupled to the various base and phosphate protecting groups. However, the full-length product was unique in possessing a 5'-terminal DMTr group, a feature exploited in one of the purification procedures used (see below).

Table 4A. Oligonucleotide design for domain A

(a)

	G	Y	(T)	G	(I)	R	C	E/Q	
	Gly	Tyr	Thr	Gly	Ile	Arg	Cys	Glu Gln	
5'-	GGX	TAC T	ACX	GGX	ATA C T	AGA G CGX	TGC T	GAA G CAA G	-3'

	Oligonucleotide sequence									Number
3'-	CCG T	ATG	TGG T	CCG T	TAG T	GC T			-5'	2
3'-	CCX	ATA	TGX	CCX	TAX	TCT	AC		-5'	5
3'-	CCX	ATG	TGX	CCX	TAX	TCT	AC		-5'	6
3'-	CCC G	ATG	TGC G	CCC G	TAC G	GCC G	ACG	CTC G	-5'	8
3'-	CCC G	ATG	TGC G	CCC G	TAC G	GCC G	ACA	CTC G	-5'	9

(b)

	G	Y	(V)	G	(D/E)	R	C	E/Q	
	Gly	Tyr	Val	Gly	Asp Glu	Arg	Cys	Glu Gln	
5'-	GGX	TAC T	GTX	GGX	GAC T GAA G	AGA G CGX	TGC T	GAA G CAA G	-3'

	Oligonucleotide sequence									Number
3'-	CCC G	ATG	CAC G	CCC G	CTC G	GCC G	ACG	CTC G	-5'	10
3'-	CCC G	ATG	CAC G	CCC G	CTC G	GCC G	ACA	CTC G	-5'	11

Table 4B. Oligonucleotide design for domain B

	H/Y	D	G	Y	C	L	H/N	G	
	His Tyr	Asp	Gly	Tyr	Cys	Leu	His Asn	Gly	
5'-	CAC T TAC T	GAC T	GGX	TAC T	TGC T	TTA G CTX	CAC T AAC T	GGX	-3'

	Oligonucleotide sequence								Number
3'-		CTG	CCG T	ATG	ACG	GAG T	GT	-5'	1
3'-		CTA G	CCX	ATA G	ACA G	AA G		-5'	3
5'-	AC	GAC	GGC G	TAC	TGC	CTC G	AAC C	GG -3'	12
5'-	AC	GAC	GGC G	TAC	TGT	CTC G	AAC C	GG -3'	13

Table 4. Oligonucleotide design for the conserved domains A and B of the EGF-family of growth factors (see section 6.2). The boxes contain the single letter code (see legend to table 2) and three letter abbreviations for amino-acids and below this the possible nucleotide sequences for the coding sense DNA are given. X represents all four bases: A, C, G and T. The sequences of oligonucleotides and the sense in which they were synthesised are given below. The table is organised to show how the oligonucleotides used in this project relate to each other and to the conserved domains. Refer to chapters 6, 7 and 8 for discussions of oligonucleotide design.

A. Sequences of oligonucleotides 2, 5, 6, 8, 9, 10 and 11, designed against domain A which is conserved in all members of the EGF-family. The non-conserved amino-acids in this domain are in parentheses: oligonucleotides 2, 5, 6, 8 and 9 were designed with those present in VGF at these positions (part a), whilst the remainder were designed with the amino-acids for which some preference was shown (part b).

B. Sequences of oligonucleotides 1, 3, 12 and 13, designed against domain B which is conserved between EGF and VGF.

The protected bases used in the syntheses were β-cyanoethyl phosphoramidites: 5'-O-dimethoxytrityl N6-benzoyl 2'-deoxyadenosine 3'-O-(β-cyanoethyl -N,N- diisopropylamino) phosphoramidite, 5'-O-dimethoxytrityl N4-benzoyl 2'-deoxy-cytidine 3'-O-(β-cyanoethyl-N,N-diisopropylamino) phosphor-amidite, 5'-O-dimethoxytrityl N2-isobutyryl 2'-deoxyguanosine 3'-O-(β-cyanoethyl-N,N-diisopropylamino) phosphoramidite, and 5'-O-dimethoxytrityl 2'-deoxythymidine 3'-O-(β-cyanoethyl-N,N-diisopropylamino) phosphoramidite.

3.9.2 Oligonucleotide work-up

Before oligonucleotides could be used as hybridisation probes, a certain amount of work-up and purification was required, namely that they had to be released from the solid support, the 5'-hydroxyl group had to be deprotected (ie. detritylated, to enable subsequent attachment of a radioactive phosphate group with T4 polynucleotide kinase), the bases had to be deprotected (to allow them to base pair and so form a duplex with the target DNA) and the full-length product had to be separated from the contaminating aborted (ie. shorter) products.

Base and phosphate deprotection and oligonucleotide cleavage from the solid support were accomplished in the same step by treatment with ammonia as follows. The cassette was removed from the machine, placed in a screw capped Eppendorf tube and spun dry of liquid (acetonitrile remaining from the final stages of synthesis). 1 ml of 25% ammonia was added,

the tube was spun to wet the column and then sealed. The tube was placed in a glass vial containing 5 ml of 25% ammonia (this was necessary in order to maintain equal pressure inside and outside of the tube) and incubated for 16 h at 50 ^{o}C. Following cooling, the cassette was removed to a fresh tube and the liquid was spun from it and pooled with that still in the first tube, resulting in approximately 1 ml of 25% ammonia containing the released oligonucleotides and free cyanoethyl base protecting groups. This was applied to a Sephadex G-25 column (NAP-10; Pharmacia) and eluted, by gravity feed, with 20% ethanol in order to separate the ammonia by exclusion chromatography. Twenty-five 100 ul fractions were collected. A small portion (1-2 ul) of each fraction was diluted to 400 ul with TE and its absorbance at 260 nm was measured, using a Pye-Unicam SP8/400 spectrophotometer and quartz cuvettes. Those fractions containing significant amounts of absorbing material were pooled (approximately 1 ml total volume, corresponding to fractions 10-20) and evaporated to dryness in a Speed Vac (Savant). This method was used for all the oligonucleotides made in this project.

Three different methods of detritylation were used. The initial method (used for oligonucleotides 1, 2 and 3) was to dissolve the dried oligonucleotide in 100 ul water, add 400 ul glacial acetic acid and leave for 15 min at room temperature. This was then evaporated to dryness and dissolved in 500 ul TE before separation through a Sephadex G-25 column (NAP-10) as described above. The liberated DMTr group came off the column after the oligonucleotide and also

absorbed at 260 nm (its presence could also be monitored by acidifying these later fractions and observing the absorbance of the $DMTr^+$ cation at 496 nm). Those fractions in the first peak of 260 nm absorbance were pooled, dried and dissolved in 500 ul TE.

For oligonucleotides 5 and 6, detritylation was accomplished whilst they were still attached to the support, using the program method normally employed by the machine during synthesis. Oligonucleotides 8-13 were purified and detritylated in the same procedure, using a NENSORB-PREP column (see below).

3.9.3 Purification of full-length oligonucleotides

Two different methods were employed to purify the full-length product from shorter oligonucleotide by-products. The first made use of size and the fine resolution of polyacrylamide gels whilst the second relied on the fact that only the full-length product possessed a 5'-DMTr protecting group (used for oligonucleotides 8-13).

Oligonucleotides 1, 2, 3, 5 and 6 were purified by ultraviolet shadowing of a polyacrylamide gel as follows. Approximately half of the detritylated product was applied to a 25 mm well of a 1.5 mm thick 25% polyacrylamide, 7 M urea vertical gel after addition of 0.25 volumes of formamide. Some formamide containing 0.1% bromophenol blue and 0.1% xylene cyanol was loaded in an adjacent well to act as a

marker for the progress of the electrophoresis. The gel was made up in TBE with a 19:1 ratio of acrylamide to N,N'-methylene bisacrylamide, filtered through a Whatman GF/C filter and degassed under vacuum. 180 ul 10% ammonium persulphate and 35 ul N,N,N',N'-tetramethylethylene diamine (TEMED) were added and the gel was poured and allowed to set at room temperature. The inner surfaces of the glass plates had previously been siliconised with Sigmacote to ease separation of the gel from the glass after electrophoresis. The gel was run for about 3 h at 40 mA or overnight at 5 mA (after running the samples in at 40 mA for 20 min), until the bromophenol blue dye marker had run approximately 10 cm. The electrophoresis buffer was TBE. Both glass plates were removed and the gel was wrapped in cling film. It was placed on a thin layer chromatography silica plate and illuminated from above with short wavelength ultraviolet light (254 nm; UV Products). The silica fluoresced green except where the presence of absorbing oligonucleotide cast a shadow. The major shadow was of the largest size and its position was marked. This region of the gel was excised, cut into four pieces and soaked overnight in 1 ml TE at room temperature to elute the oligonucleotide from the gel slice. A faint shadow of a ladder of smaller oligonucleotides was also observed in most cases, corresponding to the aborted sequences. The eluted full-length product was applied to a Sephadex G-25 column (NAP-10), in order to remove any polyacrylamide gel contaminants, and the absorbing fractions were pooled and dried as above.

Oligonucleotides 8-13 were purified and detritylated by affinity chromatography using NENSORB-PREP columns (Du Pont NEN). The principle of this procedure is that only full-length product possesses a terminal DMTr group. The oligonucleotide was released from the solid support by treatment with ammonia and separated on a Sephadex G-25 column (NAP-10) as above; the only difference to the procedure already described was that the elution buffer was 20% ethanol that had been adjusted to pH >8.0 to ensure stability of the acid-labile DMTr group. The NENSORB-PREP column was activated with methanol and washed with 0.1 M triethylammonium acetate pH 7.0 (TEAA). All solutions were drawn through the column by means of a syringe. The oligonucleotide was dissolved in 1 ml TEAA and half of this was further diluted to 4 ml with TEAA and applied to the column. 10% acetonitrile in TEAA was used to wash the failure sequences, which lacked a DMTr group, from the column. Subsequently, 0.5% trifluoroacetic acid was used to release the DMTr group, leaving only the full-length product attached to the column. This was eluted with 35% methanol and six 1 ml fractions were collected, 10 ul of each of which was diluted in TE and measured for absorbance at 260 nm. The oligonucleotide occurred in two of the fractions; these were pooled and dried as above.

The dried full-length oligonucleotide obtained from either procedure was dissolved in 200 ul TE and 1-2 ul was diluted to 400 ul with TE and analysed spectrophotometrically. The absorbance at 260 nm and the ratio of that at 260 nm to that at 280 nm were used to calculate the concentration and

assess the purity of the oligonucleotide (see below). A 50 ng/ul stock solution of oligonucleotide in TE was made for subsequent use and stored at -20 ^{o}C.

3.9.4 Calculation of oligonucleotide concentration

Given that single-stranded DNA has an absorbance at 260 nm of approximately 1.0 unit for a concentration of 30 ug/ml and a path length of 1 cm, an approximate concentration for each oligonucleotide was calculated from its absorbance. However, because of the short lengths of oligonucleotides, their relative base compositions may be significantly different from those of longer DNA, which might reasonably be expected to have approximately equal base composition. Such a difference would affect both the absorption coefficient and the molecular weight of the oligonucleotide, since these values for the four bases differ. An accurate calculation of the molecular weight (M_r) was obtained by summing those of the constituent bases, using values of 331.2, 307.2, 347.2 and 322.2 for dAMP, dCMP, dGMP and dTMP respectively, and subtracting (18 x (n-1) + 80) for an oligonucleotide of length n bases to account for the molecules of water lost during phosphodiester bond formation and the lack of a 5'-terminal phosphate. The absorption coefficient at 260 nm (E_{260}) was obtained by summing those of the constituent bases, using values of 15.3, 7.4, 11.8 and 9.3 cm^2/umol for dAMP, dCMP, dGMP and dTMP respectively (Sproat and Gait, 1984; data taken from Dawson et al., 1986). For a path length of 1 cm, the absorbance (A) is equal to the coefficient (E) multiplied by

the concentration in units of umol/ml. Thus, the oligonucleotide concentration (in ug/ml) = $A_{260} \times M_r / E_{260}$ multiplied by the dilution factor used to determine the absorption. This formula gives an absorbance value of 1.00 for single-stranded DNA of equal base composition at a concentration of 30 ug/ml.

3.9.5 Assessment of oligonucleotide purity

The purity of the resulting oligonucleotide, as well as its ability to be phosphorylated at its 5'-hydroxyl, was assessed by labelling with ATP. 20 pmol was incubated in a 20 ul volume containing 50 mM Tris.Cl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, 10 uCi [γ^{32}P]ATP (3,000 Ci/mmol) and 4 units T4 polynucleotide kinase for 1 h at 37 ^{o}C. 5 ul of formamide containing 0.1% bromophenol blue and 0.1% xylene cyanol was added and the sample was electophoresed through a 25% polyacrylamide, 7M urea gel as described above. One glass plate was removed and the gel, still attached to the other plate, was wrapped in cling film. An autoradiograph was obtained by direct contact of XAR-5 film with the wrapped gel for 2-5 min. The presence of a single band, corresponding in size to that of the full-length product, indicated purity.

Purity was further assessed by measuring the absorbance of the oligonucleotide at 260 nm and at 280 nm and determining the ratio. This was then compared to a theoretical value calculated from the known absorption coefficients for the nucleotides at 260 nm (data given above) and the known ratios

(E_{260}/E_{280}) of 6.67, 1.01, 1.49 and 1.37 for dAMP, dCMP, dGMP and dTMP respectively, from which their absorption coefficients at 280 nm were calculated (data taken from Dawson et al., 1986). This calculation assumes that there is no hyperchromic effect for oligonucleotides or that, if there is, it affects the absorbances at both wavelengths similarly. A close approximation of the observed ratio to the calculated ratio was taken as an indication of purity. This was the case (within 5%) for all oligonucleotides synthesised with the exception of number 3, the smallest (14-mer): see table 5.

	n	D	n_A	n_C	n_G	n_T	M_r	E_{260}	E_{260}/E_{280} Th.	Exp.
1	17	4	3	4	6	4	5226	183.5	1.66	1.64
2	17	32	2	5	4.5	5.5	5165	171.9	1.51	1.54
3	14	64	5.25	4.25	2.25	2.25	4237	159.3	2.09	1.66
5	20	256	5	7	2	6	6012	207.7	1.74	1.66
6	20	256	4	7	3	6	6028	204.2	1.64	1.60
8	24	64	3	10	7	4	7291	239.7	1.48	1.54
9	24	64	4	10	6	4	7275	243.2	1.55	1.58
10	24	64	3	12	6	3	7236	233.4	1.44	1.48
11	24	64	4	12	5	3	7220	236.9	1.51	1.53
12	22	8	4.5	7.5	7	3	6733	234.9	1.67	1.71
13	22	8	4.5	6.5	7	4	6748	236.8	1.69	1.70

Table 5. Parameters for the series of oligonucleotides synthesised in this project. The first column contains the number identifying each of the eleven oligonucleotides. Subsequent columns give the length in bases (n), degeneracy (D) and number of each constituent base (n_A, n_C, n_G and n_T for the bases A, C, G and T respectively). Fractional numbers within these latter four columns are due to degeneracy: the average composition of all the oligonucleotides present in a degenerate mixture is given. The next two columns give the molecular weight (M_r) and absorption coefficient at 260 nm (E_{260}, in units of $cm^2/umol$), again for the average oligonucleotide in the mixtures: these values are theoretical and were obtained as discussed in the text. The final two columns give values for the ratio of the absorbances at 260 nm and 280 nm (E_{260}/E_{280}), both theoretical (Th.) and experimental (Exp.): the similarity of these values was taken as an indication of purity.

3.9.6 Theoretical assessment of oligonucleotide specificity

As a guide to the specificity of the oligonucleotide sequences chosen, a calculation was made of how frequently matches with a small number of mismatches would occur on a random basis. For a non-degenerate oligonucleotide of length n bases, the probability of a target sequence matching with x mismatched bases is:-

$$P(x) = 3^x \cdot C_{n,x} / 4^n$$

An approximate extension of this formula, to apply to oligonucleotides which have 4-fold degeneracy at each of y positions (overall degeneracy, $D = 4^y$), is:-

$$P(x) = 3^x \cdot C_{n-y,x} / 4^{n-y}$$

since now mismatches are only possible at n-y positions.

On average, a target sequence with x mismatches is expected to occur once in 1/P(x) bases. This formula was tested by searching for matches with test oligonucleotides in test sequences, using the computer program FIND (section 3.19.3), and was found to make a reasonable prediction of these findings.

3.10 Radiolabelling of DNA

3.10.1 Random oligonucleotide primer extension

Radiolabelled probes from double-stranded DNA were constructed by the method of Feinberg and Vogelstein (1984). For labelling of individual restriction fragments, digests were electrophoresed in 0.6-1% low melting-point agarose gels and the required fragment(s) were excised after staining with ethidium bromide and visualisation by transillumination. The gel slices containing the DNA fragments were melted for 10 min at 65 ^{o}C, diluted with water at a ratio of 2-3 ml/g of gel, boiled for 10 min to denature the DNA and stored at -20 ^{o}C. Preparatory to the labelling reaction, the gel was re-boiled for 10 min.

The labelling reaction contained water to a final volume of 50 ul, the required quantity of molten gel containing the DNA (typically 10-20 ng DNA), 10 ul "oligo-labelling buffer", 2 ul 10 mg/ml BSA, 10-50 uCi [α^{32}P]dCTP (3,000 Ci/mmol) and 2 units of E. coli DNA polymerase I "Klenow" fragment. The final buffer concentration was 48 mM Tris.Cl pH 8.0, 4.8 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 20 uM each of dATP, dGTP, dTTP, 200 mM N-2-hydroxyethyl piperazine N'-2-ethane sulphonic acid (HEPES) and 160 ug/ml random hexamer oligonucleotide primers ($p(dN)_6$). The reaction was incubated at 20 ^{o}C for 4-16 h.

Unincorporated nucleotides were removed by exclusion

chromatography through Sephadex G-50 either by gravity feed, collecting sixteen 100 ul fractions from a column packed in a glass Pasteur pipette, or by spun-column chromatography, using a column packed in a 1 ml syringe and held in a centrifuge tube (Maniatis et al., 1982). To reduce probe losses, the glass pipette and glass wool used to plug the pipette/syringe were siliconized with Sigmacote. The elution buffer was 5x SSC, 0.1% SDS. In the former case, fractions were counted and those containing the labelled DNA were pooled to give a total volume of 4-500 ul. In the latter case, the reaction volume was made up to 100 ul with elution buffer before application to the column, and the labelled DNA was collected in a single fraction of 100 ul following centrifugation for 4 min at 1,600 g_{av}. The amount of label incorporated into the DNA was counted by measuring Cerenkov radiation, using an LKB 1214 Rackbeta scintillation counter (with an efficiency of approximately 30%). Probe specific activities were 10^8-10^9 dpm/ug, depending on the amount of substrate DNA.

3.10.2 5' end-labelling of oligonucleotides

10-20 pmol oligonucleotide was incubated in a 20 ul reaction volume containing 50 mM Tris.Cl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, 50 uCi [γ^{32}P]ATP (3,000 Ci/mmol) and 4-10 units of T4 polynucleotide kinase for 1-2 h at 37 ^{o}C. Unincorporated nucleotides were separated by exclusion chromatography through Sephadex G-25, collecting sixteen 100 ul fractions from a column packed in a glass Pasteur pipette, as described above. The elution buffer was

6x SSC. Fractions were counted by measuring Cerenkov radiation as above and those containing the labelled DNA were pooled to give a total volume of 4-500 ul. Probe specific activities were 1-5 x 10^6 dpm/pmol.

3.11 Filter hybridisation

3.11.1 Hybridisation with probes constructed by random oligonucleotide primer extension

Filters used with probes constructed by random oligonucleotide primer extension (section 3.10.1) were Southern blots onto nitrocellulose or nylon membranes (section 3.8). They were pre-hybridised for 4-16 h before probe addition. Pre-hybridisation solution was 6x SSC, 0.5% SDS, 5x Denhardt's solution, and 100 ug/ml denatured sonicated salmon sperm DNA. 1x Denhardt's solution is 0.02% Ficoll 70, 0.02% polyvinylpyrrolidone, 0.02% BSA. Denatured sonicated salmon sperm DNA was obtained by sonicating a 10 mg/ml solution using an MSE Soniprep 150 to a size of approximately 500 bp (as assessed by agarose gel electrophoresis), denaturing by boiling for 10 min and quenching in a solid CO_2/ethanol bath. For pre-hybridisation, the filters were sealed in plastic bags containing the solution described and incubated at 65 ^{o}C, with shaking, by submerging in water at that temperature. Nitrocellulose filters required wetting in 2x SSC prior to this treatment.

Probes were denatured by boiling for 10 min, quenched on ice and added to the pre-hybridisation solution in the bags containing the filters. Hybridisation continued at 65 ^{o}C, with shaking, for 16-20 h. Following hybridisation, filters were removed from the plastic bags, rinsed in 6x SSC, 0.1% SDS at 65 ^{o}C and washed at that temperature in several 30 min stages at successively lower salt concentrations, finishing with 0.1x SSC, 0.1% SDS for a stringent wash. Filters were then removed, air dried, wrapped in cling-film and auto-radiographed (section 3.12).

3.11.2 Hybridisation with end-labelled oligonucleotide probes

Filters used with oligonucleotide probes were Southern blots onto nitrocellulose membranes (section 3.8). Nylon membranes were not used because of difficulties experienced with blocking of non-specific sites during pre-hybridisation (due to the greater binding capacity of nylon and the lower temperatures employed with oligonucleotide probes). Filters were wetted in 2x SSC and pre-hybridised for 4-6 h at the hybridisation temperature, which varied according to the melting temperature of the oligonucleotide in use (see below). Pre-hybridisation solution was 6x SSC, 5x Denhardt's solution (described above) and 100 ug/ml denatured *E. coli* tRNA as blocking agent in place of the salmon sperm DNA used above. Denatured tRNA was obtained by boiling a 10 mg/ml solution for 10 min and quenching in a solid CO_2/ethanol bath. tRNA was found to be a more reliable block to non-specific binding with oligonucleotide probes (ie. it gave a consistently lower

background hybridisation signal), presumably because of its more defined sequence composition compared to genomic salmon sperm DNA: the latter would be more likely to contain close matches to oligonucleotide probes. Probes were added to the pre-hybridisation solution in the bags containing the filters and hybridisation was continued at the selected temperature for 16-20 h.

In initial experiments with each oligonucleotide, the temperature chosen was estimated to be 5-10 ^{o}C lower than the melting temperature. Subsequent experiments modified this in the light of experience. The theoretical melting temperature (Tm, the temperature at which 50% of a perfectly matched oligonucleotide is melted off) for each oligonucleotide was calculated, using the following formula (the "Wallace rule"):-

$$T_m = 4 \times (G + C) + 2 \times (A + T)\ ^{o}C$$

where letters represent the number of times the corresponding bases occur in the oligonucleotide sequence (Thein and Wallace, 1986). An oligonucleotide forming a hybrid containing mismatched base-pairs would melt off at a lower temperature, being destabilised by approximately 5 ^{o}C per mismatch.

Following hybridisation, filters were removed from the plastic bags, rinsed in 6x SSC, 0.1% SDS at the hybridisation temperature and washed in the same solution initially at the hybridisation temperature for 30 min and subsequently at increasing temperatures, with solution changes, to the

required level of stringency. The filters were then removed and autoradiographed (section 3.12). If the filters were to be re-washed then it was important not to allow them to dry (lest the probe became irreversibly bound) and so they were enclosed in an air-tight wrapping of cling-film. The final washing stringency was determined by assessing the autoradiographic result and re-washing at a higher temperature, if required, in order to optimise the signal-to-noise ratio. The hybridisation temperatures and washing protocols for each oligonucleotide used are given in the appropriate positions of the text below.

3.12 Autoradiography

Filters wrapped in cling-film were autoradiographed by placing in contact with X-ray film in an autoradiography cassette. Exposure was at -70 ^{o}C. Two intensifying screens (Fast tungstate; Cawo, Genetic Research Instruments) were used and the film was pre-flashed to provide a linear response. Film was pre-flashed using an electronic flash gun masked with red filters to reduce the light intensity and calibrated such that the background film density was raised by 0.15 absorbance units when measured at 540 nm (Laskey, 1984). X-ray film used was Rx-100, X-Omat XAR-5 or Hyperfilm MP. Autoradiographs were processed with Phenisol liquid developer and Hypam fixer, or were processed automatically in a Kodak RP X-Omat Processor.

3.13 DNA clone construction

During the course of this project, clones of MCV DNA in both plasmid pBR322 (Bolivar et al., 1977) and phage M13 (Messing, 1983) vectors were constructed.

3.13.1 Fragment isolation from low melting-point agarose

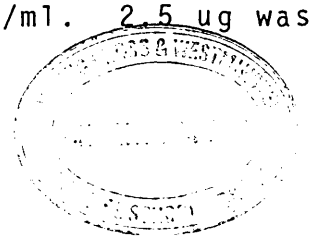
For cloning specific restriction fragments of MCV DNA, as well as for subcloning from these primary clones, gel electrophoresis was used to separate the desired fragments from the other fragments generated. Gels were made up and run as described above (section 3.6), except that 0.7-1% low melting-point agarose was used. The gel slice containing the required DNA fragment was removed. However, in order to prevent damage of this DNA by exposure to ethidium bromide and ultraviolet light transillumination, the following method was used. An aliquot of the digest (the "marker digest") was removed and run in a gel track adjacent to that containing the remainder. After electrophoresis, the gel was cut in two and only that portion containing the marker track was stained for visualisation of the DNA. The position of the required DNA fragment was marked on this portion and subsequent realignment of the two gel halves allowed its position to be determined on the unstained portion, from which the gel slice containing it was removed. When the required DNA fragment had a mobility similar to that of one of the dyes normally present in the gel loading buffer, the buffer was used without dyes (consisting of 50% glycerol, 1 mM EDTA only) to prevent contamination of

the DNA.

The gel slice containing the required DNA fragment was heated at 65 ^{o}C for 15 min to melt the agarose. TE was added to adjust the volume to 500 ul (diluting the agarose 2-3 fold) and the tube was heated for a further 15 min to ensure complete melting of agarose and release of the DNA from its matrix. This was extracted 2 or 3 times with phenol equilibrated with TE, once with phenol:chloroform:amyl alcohol (25:24:1) and once with chloroform:amyl alcohol (24:1) in order to remove the agarose. The DNA was then precipitated overnight at -20 ^{o}C with 0.1 volumes of 3 M sodium acetate pH 5.2 and 2 volumes of ethanol, and recovered by centrifugation at 13,000 g for 15 min. The pellet was washed with cold 70% ethanol, re-centrifuged and dried under vacuum using a Speedvac (Savant). The DNA was dissolved in 6-10 ul TE, of which an aliquot was analysed by agarose gel electrophoresis to assess the efficiency of its recovery (typically 30-50%).

3.13.2 Addition of synthetic EcoRI linkers to blunt-ended DNA

For cloning blunt-ended fragments (eg. those generated with SmaI) synthetic oligonucleotides were ligated to the blunt ends in order to create an EcoRI site at each end of the fragment. Linkers used were 8-mers (5'-GGAATTCC-3') and were not phosphorylated. Phosphorylation was required prior to ligation. The ability of the linkers to be ligated was determined by self-ligation, as follows. Linkers were dissolved in TE at a concentration of 500 ug/ml. 2.5 ug was



incubated in 50 mM Tris.Cl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA, 1 mM ATP in a 50 ul reaction volume with 30 units T4 polynucleotide kinase for 1 h at 37 ^{o}C. 10 uCi [γ^{32}P]ATP (3,000 Ci/mmol) was included in the reaction to allow partial labelling of the linkers. A 5 ul aliquot was removed, 0.5 ul 10 mM ATP and 3 units of T4 DNA ligase were added, and the reaction was incubated for 5 h at 12 ^{o}C. In order to assess the extent of ligation, the reaction was electrophoresed in a 10% polyacrylamide gel: an equivalent aliquot of unligated kinased linkers was also loaded. The gel was run for 40 min at 60 mA and autoradiographed as described in section 3.9.5. The ligated reaction showed a loss of signal corresponding to the monomer size and the appearance of a new signal of much larger size, indicating that the linkers were successfully kinased and self-ligated in this system.

Blunt-ended DNA in a volume of 1 ul was combined with 20 ul kinased linkers, 2 ul 10 mM ATP and 10 units T4 DNA ligase were added and the reaction was incubated for 16 h at 12 ^{o}C. The concentrations used were approximately 7 nM blunt-ended DNA and 19 uM linker DNA. The ligation reaction was heated for 15 min at 65 ^{o}C, BSA and high salt restriction buffer were added (section 3.5) and the ligation products were digested with 30 units EcoRI for 4 h in a reaction volume of 50 ul. The reaction was extracted with phenol:chloroform:amyl alcohol (25:24:1) and the DNA fragment, now possessing linkers, was separated from linker fragments by exclusion chromatography through a column of Sephadex G-50 packed in a

glass Pasteur pipette, eluting with 0.3 M NaCl in TE. Seventeen 100 ul fractions were collected and 15 ul of each was analysed by agarose gel electrophoresis to determine which contained the DNA fragment. Two fractions contained the DNA: these were pooled, precipitated overnight at -20 ^{o}C and recovered as above (section 3.13.1). The efficiency of recovery of insert DNA was estimated to be 25%.

3.13.3 Preparation of vector DNA for cloning

Vector DNA (either plasmid pBR322 or M13 RF) was digested with the required restriction enzyme(s) in the usual way. When only one enzyme was used, the DNA was subsequently phosphatased to prevent the possibility of circularisation of the vector during ligation. This was achieved by adding 10 units of calf intestinal phosphatase to the digest reaction and continuing incubation at 37 ^{o}C for a further 30 min. The digest was then extracted with phenol:chloroform:amyl alcohol (25:24:1), precipitated with ethanol and recovered as described above (section 3.13.1). Recovery of vector was monitored by agarose gel electrophoresis, which was also used to established complete digestion of the DNA.

One experiment involved using pBR322 cut with both BamHI and HindIII, generating a fragment of 346 bp and the required vector fragment of 4,016 bp. These were separated by electrophoresis in low melting-point agarose and the vector was recovered as described in section 3.13.1.

3.13.4 Ligation of insert and vector DNAs

Solutions of insert and vector DNAs in TE were combined and ligated in 66 mM Tris.Cl pH 7.5, 5 mM $MgCl_2$, 5 mM DTT, 1 mM ATP with 1-2 units of T4 DNA ligase. The reaction volume was typically 8-10 ul and was incubated for 16 h at 12-20 ^{o}C. The concentrations of insert and vector DNAs used were typically in the ratio of 3:1 (insert:vector) and in the range 1-10 nM. The actual concentrations and reaction temperatures used in particular experiments are given at the appropriate positions in the text below.

3.14 Growth, storage and transformation of E. coli

Cloned DNA was propagated in *E. coli*, either in pBR322 conferring ampicillin resistance, or in an M13 vector with productive bacteriophage infection.

3.14.1 Host strains of E.coli

Two strains of *E. coli* were used, the choice depending on the use of plasmid or M13 vectors. Strain DH1 (Hanahan, 1983) was used in the former case whilst strain JM101 (Yanisch-Perron *et al.*, 1985) was used in the latter. The genotypes of these strains are:-

DH1: F^-, endA1, hsdR17(r_k^-, m_k^+), supE44, thi-1, lambda$^-$, recA1, gyrA96, relA1

JM101: Δ(lac-proAB), supE, thi, [F', traD36, proAB, $lacI^{q}ZΔM15$]

3.14.2 Growth media and antibiotics

Several different broths and agars are referred to below. Their recipes are as follows:-

L. broth: 1% bacto tryptone, 0.5% bacto yeast extract, 0.5% NaCl, 0.1% glucose.

SOB: 2% bacto tryptone, 0.5% bacto yeast extract, 10 mM NaCl, 2.5 mM KCl; with addition 1:100 of 1 M $MgCl_2$, 1 M $MgSO_4$.

SOC: SOB with addition 1:100 of 2 M D-glucose.

2x TY: 1.6% bacto tryptone, 1% bacto yeast extract, 0.5% NaCl.

Glucose/minimal medium agar:
1x M9 salts, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mM thiamine-HCl, 0.2% glucose, 1.5% agar.
(10x M9 salts is 6% Na_2HPO_4, 3% KH_2PO_4, 1% NH_4Cl, 0.5% NaCl.)

L. agar: 1% bacto tryptone, 0.5% bacto yeast extract, 1% NaCl, 1.5% agar.

H. agar: 1% bacto tryptone, 0.8% NaCl, 1.2% agar.

H. top agar: 1% bacto tryptone, 0.8% NaCl, 0.8% agar.

All of the above were sterilised by autoclaving. The additions noted for SOB and SOC were made just before use.

Ampicillin was made up as a 25 mg/ml stock from its sodium salt and sterilised by filtration. When required, it was added to broth or agar to a final concentration of 50 ug/ml. Molten agar was cooled to <55 ^{o}C before addition.

Cells were incubated in liquid culture at 37 ^{o}C with shaking (200 rpm). Those grown on solid media were incubated at 37 ^{o}C in an inverted position.

3.14.3 Transformation with plasmid- and phage-based DNA clones

Host E. coli was transformed using the standard high efficiency protocol of Hanahan (1985). The same procedure was used for clones constructed in pBR322 and M13 vectors but there were some variations in the treatment after transformation. The method given below is that used with pBR322-based clones, for which ampicillin resistance was the selection mechanism employed. The differences for M13-based clones are then indicated.

A glycerol stock of the host strain was streaked on

agar and incubated overnight at 37 ^{o}C. L. agar was used for strain DH1 and glucose/minimal medium agar was used for strain JM101. The reason for the latter was that minimal medium selects for the presence of the F' plasmid in this strain (since the F' plasmid carries genes for proline synthesis). A single bacterial colony was used to inoculate 5 ml SOB and grown overnight. This culture was diluted 1:100 in 100 ml SOB and grown for 100-120 min to a density of approximately 5 x 10^{7} cells/ml, as determined by an optical density at 550 nm of 0.2 absorbance units. The viable cell density was monitored by plating dilutions of these cells on L. agar plates and counting the number of colonies. The required volume of this culture was removed (2.5 ml per transformation), incubated on ice for 10 min and the cells were harvested by centrifugation for 2 min at 1,000 g. The cell pellet was resuspended in TFB (1/3 x the original volume) and incubated on ice for 10 min. TFB is 100 mM KCl, 45 mM $MnCl_2$, 10 mM $CaCl_2$, 3 mM $[Co(NH_3)_6]Cl_3$, 10 mM sodium 2[N-morpholino]ethone sulphonic acid pH 6.3 (Na-MES). Care was taken during resuspension and subsequent mixing steps to prevent damaging the cells. The cell suspension was re-centrifuged, and the pellet was re-suspended in TFB (1/12.5 x the original volume) and incubated on ice for a further 5 min.

DMSO was added to the suspension of cells in TFB to a concentration of 3.5% (7 ul per 200 ul) and the cells were incubated on ice for 10 min. DTT was added to a concentration of 3.5% (7 ul per 200 ul of a 2.2 M solution in 10 mM potassium acetate pH 7.5) and incubation was continued for 10

min. A second addition of DMSO was made, to 7% final concentration, and incubation was continued for a further 10 min. 220 ul aliquots (approximately 10^8 competent cells) were removed to pre-chilled 1.5 ml Eppendorf tubes and the DNA to be used for transformation was added and mixed. Typically, 10-50 ng total DNA from a ligation reaction was used. A control for the efficiency of transformation, consisting of 1 ng of uncleaved vector DNA, was also done. The cell/DNA mix was incubated on ice for at least 40 min. The cells were then heat-shocked for 2 min at 42 ^{o}C and quenched on ice for 2 min. SOC was added to a final volume of 1 ml and the cells were incubated at 37 ^{o}C for 60 min without shaking.

Usually it was desired to plate the entire volume of transformed cells. This required concentration of the cells by centrifugation at 500 g for 90 sec and resuspension by gentle pipetting in 100 ul SOC. These were then plated by spreading on L. agar containing 50 ug/ml ampicillin and dried at room temperature. Only a fraction of the cells transformed with control DNA was plated. Plates were inverted and incubated overnight at 37 ^{o}C. Transformants were evident as bacterial colonies. The efficiency of transformation was typically 1-5 x 10^7 transformants per ug of control DNA.

For M13-based clones the same protocol was used as far as the heat-shock and quench step. 0.5 ml SOC was then added and the cells were incubated as above at 37 ^{o}C for 30 min. They were added to 270 ul of a mix of fresh JM101, IPTG (isopropyl β-D-thiogalactoside; a non-physiological inducer of

the lac operon, and hence of β-galactosidase) and XGal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside; a substrate of β-galactosidase yielding a blue hydrolysis product) made up in the ratio 200 ul of a 1:50 dilution of a static culture of JM101 grown for 2 h: 40 ul 100 mM IPTG: 40 ul 2% XGal in dimethylformamide. 3 ml molten H. top agar (at a temperature of 42 ^{o}C) was added and mixed before plating by pouring onto H. agar plates. These were allowed to set, inverted and incubated overnight at 37 ^{o}C. Transformants were evident as plaques on the lawn of JM101, since such cells produced infectious phage. Recombinant clones gave clear plaques, whilst non-recombinant and control clones gave blue plaques due to the uninterrupted B-galactosidase coding sequence present on the vector. The efficiency of transformation was typically 5-10 x 10^{6} transformants per ug of control DNA.

3.14.4 Storage of plasmid and phage clones

Colonies of E. coli strain DH1 harbouring pBR322-based clones were picked and inoculated into a small volume of broth containing 50 ug/ml ampicillin and grown overnight. For long-term storage, glycerol was added to a concentration of 20% and the cells were frozen at -20 ^{o}C. Host strains were also stored in this way. Short-term storage of liquid cultures and agar plates was at 4 ^{o}C.

Plaques of M13-based clones were inoculated into a small volume of broth containing fresh JM101 cells, grown for 5 h and phage particles were purified as described below

(section 3.16.1). The suspension of phage particles was stored at -20 ^{o}C with no pre-treatment necessary.

3.15 Preparation and restriction mapping of plasmid DNA

3.15.1 Small-scale isolation of plasmid DNA by alkaline lysis

Plasmid DNA was isolated on a small-scale using the rapid alkaline lysis method (Maniatis et al., 1982). Individual colonies of E. coli containing pBR322-based clones were inoculated into 2 ml L. broth containing 50 ug/ml ampicillin and incubated overnight at 37 ^{o}C. 1.5 ml was poured into an Eppendorf tube and centrifuged for 1 min at 13,000 g. The cell pellet was resuspended by vortexing in 100 ul 50 mM glucose, 10 mM EDTA, 25 mM Tris.Cl pH 8.0, 4 mg/ml lysozyme and incubated for 5 min at room temperature. 200 ul ice-cold 0.2 M NaOH, 1% SDS was added, mixed by rapid inversion and incubated for 5 min on ice. 150 ul ice-cold 3 M potassium acetate pH 4.8 was added, vortexed briefly and incubated for 5 min on ice. This was centrifuged for 5 min at 13,000 g and the supernatant, containing plasmid DNA, was removed. This was extracted with an equal volume of phenol:chloroform:amyl alcohol (25:24:1), after which the DNA was precipitated with 2 volumes of ethanol for 5 min at -20 ^{o}C. The DNA was pelleted by centrifugation for 5 min at 13,000 g, washed with cold 70% ethanol, re-centrifuged and dried under vacuum in a Speedvac (Savant). The pellet was dissolved in 50 ul TE containing 20 ug/ml bovine pancreatic

ribonuclease IA. 8 ul was removed and digested with the appropriate enzyme(s) to analyse the nature of the insert.

This method was used to screen the products of a cloning exercise. It was also used to produce plasmid DNA suitable for labelling by random oligonucleotide primer extension, for restriction mapping cloned inserts and for subcloning into M13 vectors.

3.15.2 Large-scale isolation of plasmid DNA by caesium chloride density centrifugation

High quality plasmid DNA was isolated on a large-scale using alkaline lysis and caesium chloride density centrifugation (Maniatis et al., 1982). 2 ml of a static culture of the required clone was inoculated into 1 litre of L. broth and incubated for 8 h at 37 ^{o}C. 5 ml 34 mg/ml chloramphenicol was added and incubation was continued overnight to amplify the plasmid DNA. The bacterial cells were collected by centrifugation in a GSA fixed-angle rotor using an RC-5B centrifuge (Sorvall) at 5,000 rpm (2,500 g_{av}) for 5 min and resuspended in 20 ml 50 mM glucose, 10 mM EDTA, 25 mM Tris.Cl pH 8.0, 4 mg/ml lysozyme. This was incubated for 10 min at room temperature. 40 ml ice-cold 0.2 M NaOH, 1% SDS was added, gently mixed and incubated on ice for 10 min. 20 ml ice-cold 3 M potassium acetate pH 4.8 was added, vortexed thoroughly and incubated on ice for 30 min. Cell debris and chromosomal DNA were removed by centrifugation in a GSA rotor at 7,000 rpm (5,000 g_{av}) for 10 min. 50 ml isopropanol was

added and left for 5 min at room temperature to precipitate the nucleic acid, which was collected by centrifugation in a GSA rotor at 10,000 rpm (10,000 g_{av}) for 10 min. The pellet was resuspended in 31 ml TE, and 31.9 g CsCl and 0.5 ml 10 mg/ml ethidium bromide were added. This was cleared of dense particulate matter by centrifugation in an HB-4 swing-out rotor at 10,000 rpm (10,000 g_{av}) for 5 min. The specific gravity was checked and adjusted with TE or CsCl, if necessary, to a value of 1.54-1.58. The solution was spun in a 38.5 ml crimp-top tube in a TFT50.38 rotor using an OTD-65B ultracentrifuge (Sorvall) at 42,000 rpm (163,000 g_{av}), 19 ^{o}C for >36 h.

After ultracentrifugation, one main band of fluoresence was visible (corresponding to closed circular super-coiled DNA) with a fainter one above (relaxed DNA). Bands were usually visible in daylight. The lower band was removed as follows. A hyperdermic needle was used to pierce the tube at the top to allow air to enter. A second needle, attached to a syringe, was inserted through the side of the tube just below the band containing the plasmid DNA, which was slowly drawn off in a volume of 3-4 ml. This was extracted with an equal volume of isopropanol saturated with TE/CsCl, to remove the ethidium bromide: this step was repeated several times. The solution was then dialysed against a large volume of cold TE, with several changes of the latter, to remove the CsCl. The DNA was precipitated overnight at -20 ^{o}C by the addition of 0.1 volumes of 3M sodium acetate pH 5.2 and 2 volumes of ethanol. The DNA was recovered by centrifugation in an HB-4

rotor at 12,000 rpm (15,000 g_{av}) for 15 min, washed with cold 70% ethanol, re-centrifuged and dried under vacuum.

Plasmid DNA was dissolved in TE and its concentration was determined spectrophotometrically, given that DNA at a concentration of 50 ug/ml has an absorbance of 1.0 unit at 260 nm. Typical yields of plasmid DNA from a litre of culture were 1-1.5 mg. The purity of the DNA was determined from the ratio of its absorbance at 260 nm to that at 280 nm and the clone construct was verified by restriction enzyme digestion and agarose gel electrophoresis.

3.15.3 Restriction mapping of cloned DNA

Restriction maps of cloned DNA were obtained by digestion with various endonucleases, both individually and in pairs. Most restriction sites could be mapped by analysis of the sizes of the resultant fragments (with reference to the position of sites within the vector DNA) and by comparison of double-digests with the two respective single-digests.

3.16 Preparation of M13 DNA

3.16.1 Isolation of M13 phage particles and single-stranded DNA for use as sequencing templates

Individual plaques of M13 clones growing on an agar lawn of *E. coli* strain JM101 were picked (by taking an agar plug surrounding the plaque) and inoculated into 2 ml 2x TY broth containing fresh JM101 (a 1:100 dilution of a static culture). Cultures were shaken vigourously at 37 ^{o}C for 5-6 h. 1.5 ml was poured into an Eppendorf tube and centrifuged for 5 min at 13,000 g to pellet the cells. The supernatant, containing the extruded phage particles, was removed and re-centrifuged to ensure complete removal of cells. If required, the bacterial cell pellet was processed for double-stranded RF DNA (see below). Supernatants were either processed immediately or stored at 4 ^{o}C until needed, in which case they were re-centrifuged before use. 1.3 ml supernatant was added to 200 ul 20% polyethylene glycol (PEG), 2.5 M NaCl, mixed and left at room temperature for 15 min in order to precipitate the phage particles. This was centrifuged for 5 min at 13,000 g and the supernatant was discarded. The tube was re-centrifuged for 2 min and final traces of supernatant were removed. The phage pellet was resuspended in 100 ul TE and extracted with phenol equilibrated with TE followed by chloroform:amyl alcohol (24:1) (the latter waw necessary to remove traces of PEG). Single-stranded phage DNA was precipitated by adding 10 ul 3 M sodium acetate pH 5.2 and 250 ul ethanol and leaving overnight at -20 ^{o}C. The DNA was

pelleted by centrifugation for 10 min at 13,000 g, washed with cold 70% ethanol, re-centrifuged and dried under vacuum in a Speedvac (Savant). The DNA was dissolved in 16 ul TE prior to use for sequencing. The presence of an insert was determined by electrophoretic retardation of clone DNA with respect to control DNA.

3.16.2 Isolation of double-stranded M13 RF DNA

For screening the products of a cloning exercise using M13 vectors, the alkaline lysis method for the rapid isolation of plasmid DNA was applied. Plaques were picked and grown as above (section 3.16.1) and the bacterial cell pellet was processed as described in section 3.15.1, yielding double-stranded replicative form (RF) phage DNA. This DNA could be digested to analyse the clone construction.

Large-scale preparation of RF DNA involved application of the caesium chloride density gradient method used for plasmid DNA. 10 ml fresh JM101 (1:100 dilution of a static culture) in 2x TY broth was inoculated with 100 ul of a frozen stock of phage of the required clone and grown for 5 h as above. The cells were pelleted and 5 ml supernatant was used to inoculate 1 litre 2x TY containing fresh JM101. This culture was grown overnight and the cells were pelleted and processed as described in section 3.15.2. Typical yields of RF DNA from a litre of culture were 120-160 ug.

3.17 Unidirectional digestion of M13 clones with exonuclease III for the production of a nested set of deletion clones for sequencing

The 2.5 kb SmaI fragment of the clone pMBH-II was subcloned into M13mp8, using synthetic EcoRI linkers. A clone in each orientation was obtained and named mIIa and mIIb (section 6.5). In order to sequence through this 2.5 kb region, an ordered set of clones with various extents of deletion of the insert from the end nearest the sequencing primer site was obtained for the production of single-stranded templates for sequencing. This was done for each of mIIa and mIIb, allowing sequence determination from both strands of the insert DNA. The method used was that described by Henikoff (1984) which is based on the property of exonuclease III that it will digest double-stranded DNA in a 3' to 5' direction from 5'-overhangs but not from 3'-overhangs. Double-stranded RF DNA was obtained for clones mIIa and mIIb by caesium chloride density centrifugation (section 3.16.2) and linearised by simultaneous digestion with BamHI and PstI. Neither of these enzymes cut within the insert DNA. The orientation of these sites within the multiple cloning site of M13mp8 is such that the vector DNA nearest the sequencing primer site possesses a 3'-overhang generated by PstI (which protects the vector from digestion) whilst the insert DNA possesses a 5'-overhang generated by BamHI: thus, digestion with exonuclease III is unidirectional.

3.17.1 Establishing the conditions for exonuclease III digestion

It was necessary to establish both the rate of digestion with exonuclease III and that deletion respectively did and did not occur from ends generated by BamHI and PstI (5'- and 3'-overhangs respectively). 4 ug bacteriophage lambda DNA digested with BamHI or PstI was dissolved in 66 mM Tris.Cl pH 8.0, 6.6 mM $MgCl_2$ at a concentration of 100 ug/ml. 260 units of exonuclease III (an excess of enzyme over susceptible ends) was added, mixed and incubated at 37 ^{o}C. 2 ul aliquots were removed at 30 sec intervals and added to 7.5 ul 0.2 M NaCl, 5 mM EDTA pH 8.0 to stop the reaction. The enzyme was inactivated by heating to 80 ^{o}C for 10 min and the products were precipitated with 30 ul ethanol in a solid CO_2/ethanol bath. The DNA was pelleted by centrifugation for 10 min at 13,000 g, washed with cold 70% ethanol, re-centrifuged and dried under vaccum in a Speedvac (Savant). The DNA was dissolved in 50 ul 0.25 M NaCl, 30 mM potassium acetate pH 4.6, 1 mM $ZnSO_4$, 5% glycerol, 67 units/ml S1 nuclease and incubated for 30 min at room temperature. 6 ul 0.5 M Tris.Cl pH 8.0, 0.125 M EDTA was added to stop the reaction, 6 ul gel loading buffer was added and the products of these reactions were electrophoresed. The results were as follows. All BamHI fragments, each possessing two susceptible ends, were deleted in proportion to the exonuclease III digestion time, at a rate of 450-500 bp/min. The only PstI fragments that showed deletion were those resulting from the termini of the bacteriophage lambda genome, since these

possessed susceptible 5'-overhangs (the cos ends), deleting at a rate of approximately 250 bp/min. The bands due to all deleted fragments remained sharp even after deletion of 4 kb, indicating that the rate of digestion was well regulated. This experiment determined (i) that BamHI ends are susceptible to digestion with exonuclease III whilst those generated by PstI are not, and (ii) that the rate of digestion by exonuclease III under the conditions used (ie. the processivity of the enzyme, the rate of progression of a single enzyme molecule along a single substrate DNA molecule) was approximately 230-250 bases/min/end.

3.17.2 Construction of deletion clones

11 ug mIIa and mIIb RF DNA (obtained by caesium chloride density gradient centrifugation) was simultaneously digested with BamHI and PstI. Additionally, 1 ug was digested with BamHI or PstI. The digested DNA was purified and dissolved in 66 mM Tris.Cl pH 8.0, 6.6 mM $MgCl_2$ at a concentration of 100 ug/ml. Prior to deletion clone generation the above test system was applied, as follows. 5 ul of each digest was incubated with 30 units exonuclease III at 37 ^{o}C. 2.5 ul was removed after 5 min and stopped by addition to 7.5 ul 0.2 M NaCl, 5 mM EDTA pH 8.0. The remaining 2.5 ul was stopped after 10 min. The reactions were processed and analysed by gel electrophoresis as above (section 3.17.1). There was no deletion of the PstI digest, the BamHI digest deleted at a rate of approximately 500 bp/min (since both ends of the linearised DNA were susceptible), and

the BamHI/PstI double-digest deleted at a rate of approximately 250 bp/min (only one susceptible end). This confirmed that digestion of the clone DNA with both BamHI and PstI had been successful and that the rate of digestion was similar to that determined above.

Having established that the BamHI/PstI double-digest was satisfactory, processing continued as follows. 5 ug of this digest was incubated at 37 ^{o}C with 325 units exonuclease III (an excess of enzyme over susceptible ends). Fourteen 2.5 ul aliquots were removed at 1 min intervals and added to 7.5 ul 0.2 M NaCl, 5 mM EDTA pH 8.0 to stop the reaction. The enzyme was inactivated by heating to 80 ^{o}C for 10 min and the products were precipitated with 30 ul ethanol in a solid CO_2/ethanol bath. The DNA was pelleted by centrifugation for 10 min at 13,000 g, washed with cold 70% ethanol, re-centrifuged and dried under vaccum in a Speedvac (Savant). Each DNA pellet was dissolved in 50 ul 0.25 M NaCl, 30 mM potassium acetate pH 4.6, 1 mM $ZnSO_4$, 5% glycerol, 67 units/ml S1 nuclease and incubated for 30 min at room temperature. 6 ul 0.5 M Tris.Cl pH 8.0, 0.125 M EDTA was added to stop the reaction. Portions from the 1, 7 and 14 min time-points were analysed by gel electrophoresis to establish that deletion had occurred as expected. The samples were extracted with 25 ul phenol equilibrated with TE followed by 25 ul chloroform:amyl alcohol (24:1) and the DNA was precipitated with 140 ul ethanol and recovered as before. Each DNA pellet was dissolved in 10 ul 20 mM Tris.Cl pH 8.0, 7 mM $MgCl_2$, 10 units/ml *E. coli* DNA polymerase I "Klenow" fragment and

incubated for 5 min at 37 ^{o}C. 1 ul 0.125 mM each of dATP, dCTP, dGTP, dTTP was added and incubation was continued for a further 5 min: this step produced ligatable blunt ends. 40 ul 66 mM Tris.Cl pH 7.5, 5 mM $MgCl_2$, 5 mM DTT, 100 ug/ml BSA, 1 mM ATP and 25 units/ml T4 DNA ligase was added and the samples were incubated for 4 h at room temperature to circularise the deleted clones. 20 ul of this ligation reaction was used to transform E. coli strain JM101 as described in section 3.14.3. Typically, 50 plaques were obtained from each ligation reaction, some of which were picked for the preparation of single-stranded sequencing templates (section 3.16.1).

The extent of the deletion for individual clones could be determined by preparation of RF DNA (section 3.16.2) and double-digestion with EcoRI and HindIII: these enzymes cut either side of the insert in the multiple cloning site of M13mp8 and not within the insert, allowing the insert to be sized directly. This was done for some clones of interest. Normally, the extent of deletion was subsequently determined from the position of the sequence data obtained within the total sequence of the insert of mIIa/b (section 3.19.2). Such analyses showed that approximately two-thirds of the clones generated had deleted to an extent expected from the time of digestion with exonuclease III. This proportion is similar to that reported by Henikoff (1984). Some clones deleted much further than expected, presumably due to the presence of single-stranded "nicks" beyond the position reached by exonuclease III digestion which rendered the other strand susceptible to digestion by S1 nuclease. Some clones deleted

to a lesser degree, presumably due to the presence of "nicks" in one strand causing exonuclease III digestion of the other strand to cease prematurely. The clones showing unexpected sizes of deletion were still useful for sequencing purposes.

3.18 DNA sequencing

DNA sequencing was performed on single-stranded M13 templates according to the dideoxy method of Sanger *et al.* (1977). Templates were prepared as described in section 3.16.1. The vectors used were M13mp8, 9, 18 and 19 and sequencing was primed using the standard "universal" 17-mer primer (dGTAAAACGACGGCCAGT). Initial reactions used *E. coli* DNA polymerase I "Klenow" fragment but subsequently the modified T7 DNA polymerase (Sequenase™; Tabor and Richardson, 1987) was used, with separate elongation and termination reactions.

3.18.1 Sequencing with DNA polymerase I "Klenow" fragment

4 ul template DNA and 1 ul primer (1-2 ng) were combined with 3.5 ul water and 1.5 ul reaction buffer (100 mM Tris.Cl pH 8.5, 50 mM $MgCl_2$) and annealed either by incubation at 55-60 ^{o}C for 2 h or by slow cooling from 65 ^{o}C to room temperature over 30 min. 1.5 ul [α^{35}S]dATPαS (500 Ci/mmol, 10 uCi/ul) and 1 ul Klenow (1 unit/ul) were added and mixed. 2.5 ul of this primer/template/label/enzyme mix was placed in each of four Eppendorf tubes labelled A, C, G, T and 2 ul of the

following was added:-

Tube A: 62.5 uM dCTP, 62.5 uM dGTP, 62.5 uM dTTP, 50 uM ddATP

Tube C: 4.1 uM dCTP, 82 uM dGTP, 82 uM dTTP, 25 uM ddCTP

Tube G: 82 uM dCTP, 4.1 uM dGTP, 82 uM dTTP, 20 uM ddGTP

Tube T: 82 uM dCTP, 82 uM dGTP, 4.1 uM dTTP, 250 uM ddTTP

The concentrations of dideoxynucleotides used were determined by experiment to allow sequences to be read upto >300 bases from the primer. These sequencing reactions were incubated for 20 min at room temperature. 2 ul 0.5 mM each of dATP, dCTP, dGTP, dTTP ("chase solution") was added and incubation was continued for 15 min to elongate any non-specifically terminated extensions. 4 ul 95% formamide, 20 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol was added and the reactions were stored at -20 ^{o}C until required.

3.18.2 Sequencing with T7 DNA polymerase (Sequenase^{TM})

7 ul template DNA and 1 ul primer (0.5 pmol, 2.8 ng) were combined with 2 ul reaction buffer (200 mM Tris.Cl pH 7.5, 100 mM $MgCl_2$, 250 mM NaCl) and annealed by slow cooling from 65 ^{o}C to room temperature over 30 min. 1 ul 0.1 M DTT, 2 ul 1.5 uM each of dCTP, dGTP, dTTP, 1 ul [α^{35}S]dATPαS (500 Ci/mmol, 10 uCi/ul) and 2 ul Sequenase^{TM} enzyme (14 units/ul) were added, mixed and incubated for 5 min at room temperature

to elongate from the primer. The limiting factor in the elongation step was the amount of "cold" nucleotides and not the amount of labelled dATPaS. 3.5 ul of this elongation reaction was placed in each of four Eppendorf tubes labelled A, C, G, T and 2.5 ul of the following termination mixtures (each of which contained 50 mM NaCl) was added:-

Tube A: 80 uM each of dATP, dCTP, dGTP, dTTP, 8 uM ddATP

Tube C: 80 uM each of dATP, dCTP, dGTP, dTTP, 8 uM ddCTP

Tube G: 80 uM each of dATP, dCTP, dGTP, dTTP, 8 uM ddGTP

Tube T: 80 uM each of dATP, dCTP, dGTP, dTTP, 8 uM ddTTP

These termination reactions were incubated for 5 min at 37 ^{o}C and stopped by the addition of 4 ul 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol. Reactions were stored at -20 ^{o}C until required.

When the quantity of template DNA was less than optimal, the concentration of nucleotides in the elongation reaction was reduced from 1.5 uM to 0.75 uM to enable sequence data close to the primer site to be obtained.

3.18.3 Polyacrylamide gel electrophoresis of sequencing reactions

Sequencing reactions were electrophoresed through 6% polyacrylamide containing 7M urea using a model S2 system (Bethesda Research Laboratories; gel size 40 x 30 cm). The gels were either of uniform thickness (0.4 mm) or were "wedge" gels (0.4 mm at the top, 1.2 mm at the bottom). Gels were made up in TBE with a 19:1 ratio of acrylamide to N,N'-methylene bisacrylamide. The acrylamide/urea solution was filtered through a Whatmann GF/C filter and degassed under vacuum. 1 ml 10% ammonium persulphate and 30-50 ul N,N,N',N'-tetramethylethylenediamine (TEMED) were added and the gel was poured. The inner surfaces of the glass plates had previously been siliconised with Sigmacote to ease separation of the gel from the glass after electrophoresis. The flat edge of a sharkstooth comb was inserted and the gel was left to set at room temperature in a near horizontal position. Once set, the combs were reversed to provide "wells" for sample application and the gel was pre-electrophoresed for 30 min at 1,000-1,200 V, using TBE as running buffer.

Sequencing reactions were heat denatured at 80-90 ^{o}C in a dry block for 2 min and 2-3 ul was loaded. The gel was run at 1,000-1,200 V for times between 2-16 h, depending on whether sequence close to or distant from the primer site was required. After electrophoresis, one glass plate was removed and the gel was submerged in 10% glacial acetic acid, 10% methanol for 15 min (0.4 mm gels) or 1 h (wedge gels). The

gel, still supported by the second glass plate, was removed and a sheet of Whatman 3MM filter paper was laid on top. This was carefully peeled back with the gel attached. The gel was dried under vacuum at 80 ^{o}C onto 3MM filter paper using a gel-drier (Bio-Rad) and autoradiographed at room temperature using Hyperfilm-MP (Amersham). Autoradiography times varied from overnight to 1 week depending on the signal strength.

3.18.4 Sequencing through "compressions"

Some regions of DNA were problematic to sequence and resulted in "compressions" on sequencing gels due to secondary structure formation during gel electrophoresis. Efforts to overcome this problem involved either the use of more strongly denaturing electrophoresis conditions or the use of analogues of dGTP: the inosine analogue, dITP, or 7-deaza-dGTP. dITP was used with Sequenase^{TM} in place of and at twice the concentration of dGTP. 7-deaza-dGTP was used with Klenow in place of dGTP. Both these analogues form less stable base pairs with dCTP compared to those formed by dGTP and so secondary structure formation is less likely. A consequence of this weaker base pairing is that secondary structure formation by the template can displace the elongating strand and cause non-specific termination. Thus, whenever these analogues were used their reactions were electrophoresed alongside reactions using dGTP.

Another method used to overcome secondary structure formation during electrophoresis was to increase the

denaturing conditions by including 40% deionised formamide in the polyacrylamide gel in addition to 7M urea. The voltage was increased to 1,500 V to compensate for the lower current drawn.

In practice, a combination of the use of dGTP analogues and more strongly denaturing electrophoresis conditions was necessary to overcome and read through the more stubborn compressions.

3.19 Computer analysis of DNA sequence data

3.19.1 Reading of sequencing autoradiographs

Sequences were read from autoradiographs by recording the order of bands in the four tracks corresponding to reactions terminated with the four dideoxynucleotides. This was done with the aid of a sonic digitiser (Graf/Bar GP-7) connected to a BBC microcomputer running the program GELTALK (obtained from R. Findlay, AFRC Computing Centre). This program makes use of a speech synthesiser to talk the sequence back to the operator as it is recorded, thus reducing the possibility of reading errors. The program was modified to produce a computer sequence file in a format compatible with the Staden programs (see below).

3.19.2 Sequence data compilation

Sequence data files were transferred over the Joint Academic Network (JANET) to a VAX computer at Edinburgh University, using the file transfer program KERMIT. All subsequent sequence compilation and analysis was performed on this computer with terminal emulation running on the BBC microcomputer.

Individual gel readings that were part of a larger sequencing exercise (eg. those generated from the overlapping set of deletion clones used in chapter 6) were compiled into a single contiguous sequence with the Staden programs (Staden, 1982; Staden, 1987). The screening program SCREENV was used to detect the presence of vector (M13) sequence within readings: readings containing such sequences (ie. those running off the end of the insert) were edited appropriately. The database programs DBUTIL and DBAUTO were used to assemble the contiguous sequence. Sequence data were derived from both DNA strands and often more than once from each strand. Discrepancies were resolved by consulting the autoradiographs and by re-sequencing if necessary. A consensus sequence was derived from the database and used for subsequent analysis.

3.19.3 Sequence data analysis

Programs used to analyse sequence data were ANALYSEQ (Staden, 1984) and the package of programs produced by the University of Wisconsin Genetics Computer Group (UWGCG;

Devereux, 1984). ANALYSEQ was used to translate the sequence in the six possible reading frames, for searching for short sequence motifs and for the determination of potential coding sequences using the algorithms of McLachlan, Staden and Boswell.

UWGCG programs were used to search for restriction sites (MAPSORT), count bases (COMPOSITION), identify open reading frames (orfs) (FRAMES), identify potential coding sequences using the Fickett algorithm (TESTCODE), translate orfs (TRANSLATE), study codon usage (CODONFREQUENCY), compare two sequences (GAP), search for short sequence motifs (FIND), and compare individual sequences with groups of sequences (ie. database searches, WORDSEARCH and SEGMENTS). GAP uses the algorithm of Needleman and Wunsch, inserting gaps to optimise sequence alignment. WORDSEARCH uses the algorithm of Wilbur and Lipman to find matches to a query sequence from a database of sequences (DNA or protein). For searching with DNA sequences the "mask" option of WORDSEARCH was selected whereby homology in a repeated ++- pattern (where + refers to a sequence match and - may or may not match) was sought, thus allowing for sequence divergence at the third position of codons. SEGMENTS uses the algorithm of Smith and Waterman to illustrate the segments of homology found by WORDSEARCH.

Database searches used sequences from release 54.0 of the Genbank nucleic acid sequence database and release 15.0 of the National Biomedical Research Foundation (NBRF) protein sequence database (see Bishop et al., 1987).

CHAPTER 4

Molluscum contagiosum virus: an epidemiological survey

The work presented in this chapter is an epidemiological survey of 105 patients with Molluscum contagiosum lesions presenting to hospital departments or to Genito-Urinary clinics (section 3.3). 124 samples were obtained (usually by curettage) from this group of patients. The majority of lesions were processed for total DNA (ie. viral and cellular): however, in some cases Molluscum contagiosum virus (MCV) was first purified (section 3.4). DNA was analysed by restriction enzyme digestion and agarose gel electrophoresis (sections 3.5 and 3.6).

4.1 There are three subtypes of Molluscum contagiosum virus

When digested with restriction enzymes, MCV DNA gives a discrete, easily recognisable pattern of DNA fragments. Even in those samples processed for total DNA the banding due to viral DNA (which predominated) was easily visible. Studies with the restriction endonucleases BamHI, ClaI and HindIII revealed that there are three independent subtypes of MCV whose DNA possesses one of three different cleavage patterns. Figure 4 illustrates this observation for the enzymes BamHI and ClaI. The subtypes are referred to as MCV I, MCV II and MCV III. Further discussion of these subtypes is given in

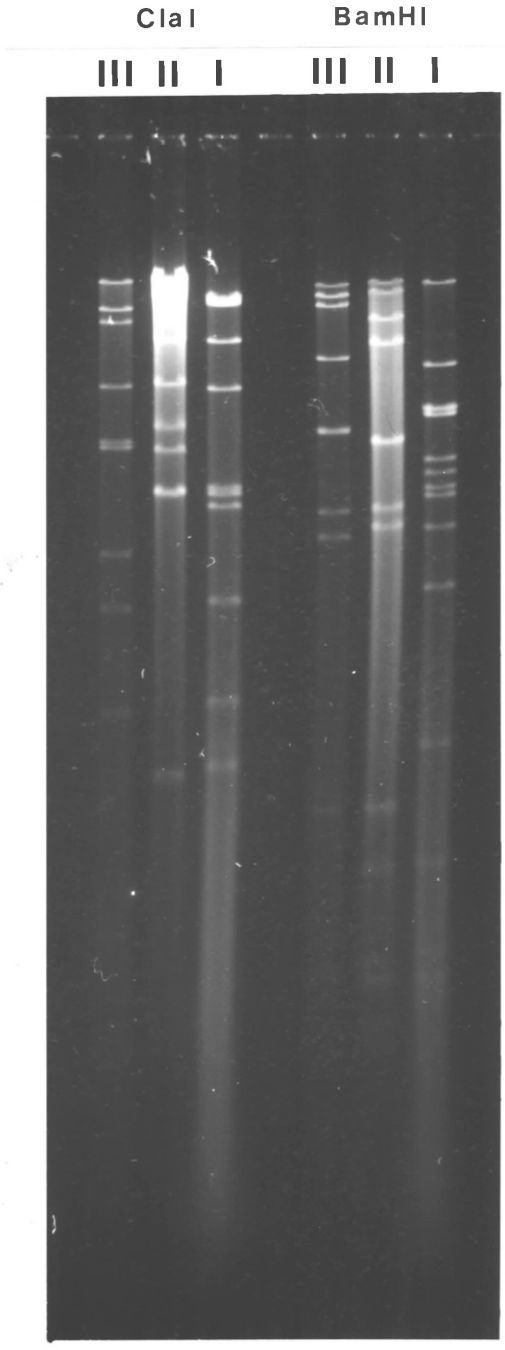


Figure 4. Comparative restriction digests of MCV I, MCV II and MCV III DNAs for the enzymes BamHI and ClaI. The isolates used were numbers 4, 26 and 8 respectively. Digests were electrophoresed in adjacent tracks of the same gel to facilitate comparison.

chapter 5, in which their genomic restriction maps are determined and compared, and DNA hybridisation studies demonstrating their close relationship are described. There were no detectable differences in the phenotypes of lesions containing each of the three MCV subtypes.

Within the 124 samples studied the ratio of the incidence of MCV I:II:III was 92:31:1. Parr *et al.* (1977) and Darai *et al.* (1986) have also reported multiple MCV subtypes (see section 1.2.6). Parr *et al.* described three subtypes, two of which were closely related and correspond to MCV I as described here, and the third of which corresponds to MCV II. Darai *et al.* reported two subtypes, corresponding to MCV I and MCV II as described here. Thus, the rare subtype III described in this study has not been reported previously.

4.2 The study group of patients; occurrence of MCV subtypes

During the course of this study, 124 *Molluscum contagiosum* isolates were obtained from 105 patients. The MCV subtype was determined in each case by DNA restriction analysis with the enzyme BamHI and by reference to the digest patterns obtained for three isolates chosen to represent the MCV subtypes (isolates 1, 26 and 8 respectively, illustrated in figure 4: these three isolates were studied further in the following chapter). Appendix 2 lists all 124 isolates with their patient details, grouped according to MCV subtype. The ratio of the incidences of MCV I:II:III was 92:31:1 in this

selection of isolates. In five other cases, determination was not possible due to the poor quality or insufficient quantity of DNA obtained. Four other patients had lesions removed and processed that proved negative for MCV DNA: these patients were subsequently re-diagnosed, upon histological examination of the lesions, as having viral warts, proliferative keratosis, angioma and possible histiocytoma.

Of the 105 patients, two independent isolates were obtained from each of 11 patients and three were obtained from each of 4 patients. In 11 of these 15 cases, the different isolates were obtained when the patients made two or more presentations: in the remaining cases, lesions were removed at the same presentation but from different sites (refer to appendix 2). In only one of these 15 cases were different MCV subtypes observed in the different isolates from the same patient; namely, that isolates 12, 13 and 91, obtained on three different occasions from a female patient with genital lesions, were determined to be subtypes I, II and I respectively. The later presentations were 7 weeks (isolate 13) and 16 months (isolate 91) after the initial presentation (isolate 12). This indicates a dual infection, or successive infections, with MCV I and MCV II in this patient. There was no case in which a single isolate indicated mixed infection with different MCV subtypes.

The study group consisted of 28 males (age range 3-70 years) and 77 females (age range 2-62 years). This sex bias reflects patient selection in that the 44 patients who

presented to the Genito-Urinary clinic were all female (see section 3.3). The 61 patients presenting to the hospital departments comprised 28 males and 38 females and were not subject to any selection procedure. There was, nevertheless, a sex/age bias in these patients: of the 16 patients aged less than 15 years, 11 were male. Some patients suffered from other medical problems that might have been relevant to the progression of Molluscum contagiosum: eg. immuno-suppressive disorders, known to exacerbate infection (reviewed in Brown et al., 1981). These included atopic dermatitis (8 cases), ichthyosis (1 case), systemic lupus erythematosis (1 case), erythema (1 case) and pancytopenia (1 case).

Two Molluscum contagiosum lesion variants have been described: the intrafollicular lesion, occurring in the epidermal cells of the hair follicle (Krishnamurthy, 1974) and the so-called "giant" lesion, which is much larger (>15 mm diameter) than normal lesions (2-5 mm) and usually solitary (Brown et al., 1981). Two intrafollicular lesions and three giant lesions were amongst those studied. There was nothing unusual about the viral DNA isolated from such lesions when examined as above.

Molluscum contagiosum lesions occurred at most sites of the body. A large proportion of isolates studied were from genital lesions (61 isolates; 49%), reflecting sexual transmission as a common route of infection amongst adults. Lesions were also common on the trunk (23 isolates; 18.5%), limbs (17 isolates; 14%) and head (23 isolates; 18.5%) but did

not occur on the palms or soles, as noted by Postlethwaite et al. (1967).

The youngest incidence of MCV II was in a patient aged 15 years (isolate 70, from a female with a facial lesion). There were 17 isolates from 16 patients who were younger than 15 years of age but all of these were MCV I. The oldest incidences of MCV I and MCV II were 62 and 70 years of age respectively.

4.3 Statistical analysis of the occurrence of MCV subtypes

As noted above, lesions induced by the three subtypes of MCV were phenotypically indistinguishable. There was only a single incidence of MCV III, in a 16 year old female with lesions on her back. Although her lesions were unusually large, there was nothing peculiar about the clinical presentation. Since no other cases of MCV III exist, this subtype is not included in the following statistical analysis. Of the remaining 123 isolates from 104 patients studied, the ratio of incidence of MCV I:II was 92:31, approximately 3:1. Although the progression of infection with MCV I and MCV II is apparently the same, the possibility exists that these subtypes are associated with other clinical features or have a preferential site or mode of infection. The latter would be analogous to the situation with herpes simplex virus subtypes I and II, which show a preference for facial and genital sites respectively (Chaney et al., 1983) and to the association of

human papillomavirus subtypes with particular sites of infection (Gissman et al., 1983).

In the following analysis, numbers refer to numbers of patients and not isolates to avoid bias due to those patients who presented on more than one occasion for which the outcome was the same. Thus, if a given patient has more than one isolate falling in the same category, only one is counted. The totals of those included in the categories discussed below are not always consistent since sometimes a patient is included in two categories (eg. a patient presenting with genital and non-genital lesions on different occasions, or with MCV I and MCV II on different occasions). The total number of patients with MCV I and MCV II were 80 and 25 respectively. The results of the relative subtype incidences in various categories discussed below were analysed using the X^2 test with Yates' correction (X^2_y). These data are presented in table 6.

Amongst male patients the ratio of the occurrence of MCV I to MCV II is 25:3, whilst the ratio amongst females is 55:22 (P=0.10). As mentioned above, there is a sex/age bias amongst these patients. Considering only those 15 years of age and above, the ratio for males is 14:3 and that for females is 50:22, not significantly different (P=0.44). Thus, there is no sex bias in the incidence of MCV I and MCV II.

Of patients with atopic dermatitis, the ratio of MCV I to MCV II incidence is 7:1, whilst that in other patients is

not significantly different: 73:24 (P=0.73).

The relative incidence of MCV I and MCV II in patients below 15 years of age is 16:0, whilst that in patients at or above this age is 64:25. This observation is statistically significant (P=0.035). Since the majority of lesions from patients in the older age group were genital whilst only one lesion at this site was amongst the younger age group, this statistic was re-calculated for only those patients presenting with non-genital lesions. The relative incidences were again significantly different: 15:0 for those younger than 15 years of age, and 27:10 for the remainder (P=0.063). Thus, the lack of MCV II incidence in patients younger than 15 years of age is not a reflection of the association of this subtype with genital lesions of presumed sexual transmission.

The relative incidence of MCV I and MCV II in patients with genital lesions is 38:16, whilst that in patients with non-genital lesions is 42:10 (P=0.31). Considering only the group of patients aged 15 years and above, the incidence ratio of MCV I:II amongst those with genital lesions is 37:16, compared to 27:10 for those with non-genital lesions (P=0.93). These ratios are not significantly different, confirming that there is no restriction of MCV II to genital lesions.

	MCV I	MCV II	X^2_y	P
All patients	80	25		
Males	25	3	2.69	0.10
Females	55	22		
Atopic dermatitis	7	1	0.12	0.73
Non-atopic dermatitis	73	24		
<15 years	16	0	4.45	0.035
15 years and above	64	25		
Genital lesions	38	16	1.04	0.31
Non-genital lesions	42	10		
Non-genital lesions only				
<15 years	15	0	3.43	0.063
15 years and above	27	10		
15 years and above only				
Genital lesions	37	16	0.008	0.93
Non-genital lesions	27	10		
Males	14	3	0.59	0.44
Females	50	22		

Table 6. Incidence of MCV I and MCV II in various patient categories. Numbers refer to patients and not isolates. If a given patient had more than one isolate in the same category, only one was counted. The totals of those included in the categories are not always consistent since occasionally patients were included in both (eg. because they presented once with a genital lesion and once with a non-genital lesion). Data were analysed with the X^2 test, applying Yates' correction (X^2_y). The value of X^2_y and P, the probability, are given in the table.

4.4 Summary of chapter

DNA prepared from Molluscum contagiosum lesions was analysed by restriction endonuclease digestion and agarose gel electrophoresis, revealing three independent subtypes of MCV. The relative incidences of the three subtypes in a group of 124 isolates was 92:31:1. MCV III is a unique subtype, not previously described. Lesion sites included the trunk, limbs, head and genitalia. The latter site was common in adults, reflecting sexual transmission of the virus. Lesions induced by different subtypes were phenotypically indistinguishable.

There was no correlation of the occurrence of either MCV I or MCV II with patient sex, those with atopic dermatitis, or the site of infection (genital versus non-genital). However, there was a significiant correlation with patient age, MCV II being present only in the group whose ages were 15 years and above (P=0.035). This correlation is not a reflection of patient sex or of sexual transmission as the mode of MCV-infection.

CHAPTER 5

Molluscum contagiosum virus genome structure and restriction maps; the relationship of viral subtypes I, II and III; subtype genetic variation

Mention has been made in the preceding chapter of the existence of three subtypes of Molluscum contagiosum virus (MCV). In this chapter the characterisation of viral DNA and construction of restriction maps for all three subtypes is described. This work was originally performed for an isolate of MCV I before becoming aware of the existence of other MCV subtypes, and so this is presented first. It is followed by an analogous account for an isolate of each of the two remaining subtypes. The three subtypes are compared by cross-hybridisation and, finally, genetic variation amongst different isolates of a given subtype is described.

5.1 Preparation and preliminary characterisation of MCV I DNA

Molluscum contagiosum lesions were obtained by curettage and stored at 4 ^{o}C prior to virus purification. Lesions from a single site on a single patient were pooled and homogenised. Virus particles were pelleted by centrifugation through 36% sucrose everlaid with Dextran T10 and then lysed with proteinase K in the presence of SDS and sarcosyl

detergents (section 3.4.1). After extraction with phenol and precipitation with ethanol the resulting purified viral DNA was dissolved in TE and examined by restriction endonuclease digestion and agarose gel electrophoresis (sections 3.5 and 3.6). One isolate of MCV DNA (isolate 1, subtype I) was of particularly good quality and abundance and was selected for characterisation of the viral genome. All experiments described below used this isolate, with the exception of the double-digests, for which a different isolate that gave the same cleavage profiles was used.

In the initial experiments, MCV DNA was digested with several different restriction endonucleases in order to allow an estimation of the size of the genome by summation of the sizes of individual restriction fragments. The enzymes BamHI, ClaI, HindIII, PstI, SmaI, and XhoI were selected for this purpose: the recognition sequences for these enzymes are GGATCC, ATCGAT, AAGCTT, CTGCAG, CCCGGG and CTCGAG respectively. DNA fragments were separated by gel electrophoresis in 0.6% agarose and visualised by ethidium bromide staining (figure 5). Fragments are identified alphabetically, with "A" referring to the largest. Digests of bacteriophage lambda DNA with ApaI, HindIII or PstI were used as size markers to construct a calibration curve of distance migrated against $\log_{10}$(size in kb), from which the sizes of MCV DNA fragments were determined. The enzymes PstI, SmaI and XhoI gave many fragments, some of which were difficult to resolve and many more of which were probably lost from the gel due to their small size. They were thus not useful enzymes for

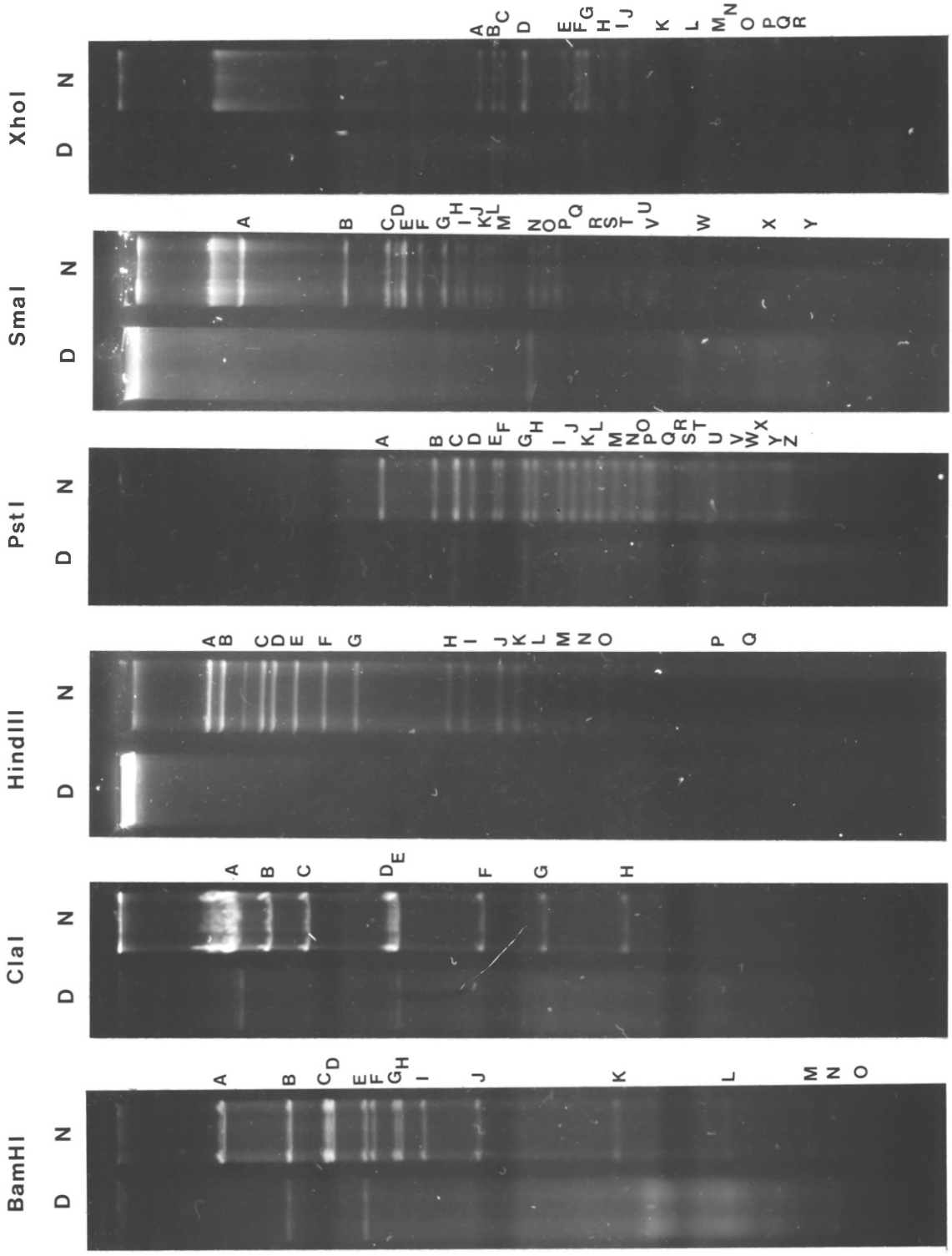


Figure 5. Restriction digests of MCV I DNA (isolate 1) and identification of rapidly renaturing terminal fragments. Digests with the enzymes BamHI, ClaI, HindIII, PstI, SmaI and XhoI are illustrated in the tracks labelled N (native) and fragments are identified alphabetically as in table 7. Fragments that are 2-molar are evident by their enhanced staining with ethidium bromide (eg. BamHI fragments C and D). Digests in the tracks labelled D (denatured) were denatured by the addition of formamide and heating at 65 ^{o}C and then cooled on ice before loading in order to identify the rapidly renaturing terminal hairpin fragments (eg. BamHI fragments B and E).

	BamHI	ClaI	HindIII	PstI	SmaI	XhoI
A	49	36,35,34	54	10.2	22.5	6.4
B	18.3	24	28	7.8	8.7	6.1
C	13.8 (2)	16.5	16.5	7.1 (2)	7.0 (2)	5.9
D	13.2 (2)	9.9	15.2	6.7	6.8	5.4 (2)
E	11.2	9.6 (2)	11.9	6.1	6.5 (2)	4.6
F	10.4	6.6	9.7	5.9	6.2	4.4 (2)
G	9.4	5.1	8.2	5.4	5.5 (2)	4.3 (2)
H	9.2	3.7	5.3	5.2	5.2	4.0
I	8.2		5.0	4.6	5.1	3.7
J	6.4		4.3 (2)	4.4	4.8	3.6
K	3.8		4.0 (2)	4.2	4.6	3.2
L	2.5		3.8	4.0	4.5	2.9
M	1.8		3.3	3.8	4.4	2.6 (2)
N	1.6 (2)		3.1	3.6	3.9 (2)	2.5 (2)
O	1.4		2.9 (2)	3.4	3.7 (2)	2.4 (2)
P	0.9		1.9	3.3	3.5 (2)	2.0
Q			1.3	3.2	3.3	1.9
R				3.0	3.0	1.8
S				2.9	2.9 (2)	
T				2.8	2.7 (2)	
U				2.7	2.5	
V				2.5	2.4 (3)	
W				2.3	2.0	
X				2.2	1.5	
Y				2.1	1.3	
Z				2.0		
kb	189.7	190.0	189.6	>118.5	165.0	>89.3

Table 7. Sizes (kilobase pairs) of restriction fragments of MCV I DNA (isolate 1) generated by BamHI, ClaI, HindIII, PstI, SmaI and XhoI. Fragment sizes were determined from their mobilities as calibrated against bacteriophage lambda DNA markers. However, because this is inaccurate for fragments greater than 30 kb in size, figures given for the larger fragments were derived from the restriction maps subsequently determined. Fragments are identified alphabetically, with "A" referring to the largest, and relative fragment molarities greater than one are indicated in parentheses. The components of the 3-molar ClaI fragment A were not resolved by gel electrophoresis although their sizes as estimated from the maps are slightly different, as given here. For BamHI, ClaI, HindIII and SmaI all fragments visualised are listed. For PstI and XhoI large numbers of small (<2 kb) fragments were generated which were not resolved but which account for a significant proportion of the genome; listing stops at this limit of resolution. The total genome size as estimated by summation of the listed fragment sizes for each enzyme is given at the bottom of the columns.

sizing the viral genome. In contrast, ClaI generated few fragments. BamHI and HindIII were the most useful enzymes in that they generated a reasonable number of separable fragments. Some fragments were identified as being 2- or 3-molar relative to the majority of other fragments (ie. the component fragments had very similar or identical mobilities and were not resolved) due to their enhanced staining relative to fragments immediately above and below; otherwise a gradual decrease in staining intensity with increasing mobility was observed. Additionally, the 3-molar ClaI fragment A co-migrated with the small amount of residual cell DNA in the preparation which was not cleaved with ClaI (or SmaI or XhoI; see figure 5) due to the presence of a CpG dinucleotide in the recognition site for this enzyme and the fact that such sequences are under-represented and often methylated (rendering them insensitive to cleavage) in the human genome (Bird, 1986). However, one other isolate was exceptionally pure and showed partial resolution of the three components of this fragment (see figure 4). Viral DNA fragment sizes and relative molarities are given in table 7. Due to the limited ability of agarose gel electrophoresis to resolve fragments of size greater than 30 kb the sizes of the largest fragments in the BamHI, ClaI and HindIII digests as determined from their mobilities were only approximate. The figures given for these fragments in table 7 are those obtained with reference to the restriction maps, whose derivation is described below. All other fragment sizes given are those obtained by direct measurement. There is good agreement between the totals obtained by summation of the sizes of restriction fragments

generated by BamHI, ClaI and HindIII and the genome size of this isolate of MCV was thus determined to be 190 kb.

SmaI was expected to cut MCV DNA infrequently due to the high G+C content of its recognition sequence and the fact that most other previously analysed poxvirus genomes are AT-rich (G+C contents of 26-40%) and possess few such sites (Andrews et al., 1978; Mackett and Archard, 1979). However, SmaI generated many (>25) fragments. This is in general agreement with the observations by Parr et al. (1977), whose studies on the denaturability of the MCV genome using electron microscopy of DNA spreads demonstrated the need for an extra 10 ^{o}C equivalents to achieve levels of denaturation comparable to those of vaccinia virus DNA (36% G+C content). From this observation the G+C content of MCV DNA can be estimated as 60 %. From sequence data subsequently derived, covering approximately 5 kb of the genomes of MCV I and MCV II, a figure of 58% was observed.

The genomes of other poxviruses possess covalently cross-linked termini and this property has been used previously to identify terminal restriction fragments by their ability to renature rapidly upon release from denaturing conditions (Geshelin and Berns, 1974). MCV DNA was digested as above with the enzymes BamHI, ClaI, HindIII, PstI, SmaI and XhoI in a total reaction volume of 24 ul. Following digestion, 16 ul was removed and added to 32 ul deionised formamide. This was heated at 65 ^{o}C, in order to denature the DNA fragments, and then quenched on ice for 30 min to

inhibit renaturation of all except the terminal fragments which were able to fold back on themselves and renature intramolecularly by virtue of their being a single strand of self-complementary DNA (ie. "to renature rapidly"). The native and denatured digests were electophoresed in adjacent tracks. Comparison of the two showed that either one or two significant bands were present in the denatured digest and that these corresponded to bands present in the native digest which were thus identified as the genomic terminal fragments (figure 5). The other fragments did not give bands since single-stranded DNA stains poorly with ethidium bromide. One of the ClaI terminal fragments was one of the 3 components of fragment A. This experiment allowed it to be visualised in the absence of the other components; it was sized as approximately 36 kb. Terminal fragments identified by rapid renaturation were BamHI fragments B and E (18.3 kb and 11.2 kb respectively), ClaI fragments A and E (36 kb and 9.6 kb), HindIII fragment O (2.9 kb, 2-molar), PstI fragments C and G (7.1 kb and 5.4 kb), SmaI fragment N (3.9 kb, 2-molar) and XhoI fragments B and C (6.1 kb and 5.9 kb).

Another feature of previously characterised poxviral DNA is the presence of inverted terminal repeats of several kilobases (Mackett and Archard, 1979). That MCV DNA also possesses this feature is suggested by the observation that both HindIII and SmaI generated single, 2-molar (assumed identical) rapidly renaturing fragments. It is likely that these two enzymes cut within the terminal repeat. In contrast, the other four enzymes generated two terminal

fragments of different sizes since they cut outside of the repeat. The largest 2-molar terminal fragment was 3.9 kb in size (generated by SmaI) and the smallest 1-molar terminal fragment generated by an enzyme that does not cut within the repeat was 5.4 kb (PstI), placing lower and upper limits on the size of the repeat. That a terminal repeat does exist was subsequently confirmed by cross-hybridisation of the fragments identified here as occurring at the termini (see section 5.2).

The sizes of the terminal fragments varied in different MCV isolates. This is also a previously characterised feature of orthopoxviruses in which it is due to the presence of variable numbers of short (70 bp) tandem repetitions within the terminal inverted repeat (Wittek and Moss, 1980). This observation is evidence for an analogous situation in MCV and is substantiated by the significantly higher signal strength resulting from hybridisation between terminal fragments than that resulting between other fragments (see below): the most likely explanation for such enhanced hybridisation kinetics is the presence of multiple repeat sequences. Indeed, promotion of rapid annealing of DNA strands during poxviral DNA replication is the presumed function of this structure (Wittek and Moss, 1980).

These initial experiments revealed a genome size of 190 kb for this isolate of MCV, a relatively high G+C content, the presence of terminal covalent cross-links, and the possession of a terminal inverted repeat of 3.9-5.4 kb, which itself contains multiple tandem repetitions.

5.2 Hybridisation studies for restriction map determination

In order to determine the relative positions of restriction fragments within the viral genome individual fragments were isolated, labelled and used as probes to hybridise to Southern blots of heterologous and homologous (control) digests of the same DNA. This method has been used previously for the determination of orthopoxviral restriction maps (Mackett and Archard, 1979). 1-2 ug of MCV DNA was digested with BamHI, ClaI or HindIII, electrophoresed in a 1% low melting-point agarose gel and gel slices containing individual fragments were excised. These were used for construction of radiolabelled probes by random oligonucleotide primer extension, using 10 uCi [α^{32}P]dCTP per fragment (section 3.10.1). Unincorporated nucleotides were removed by exclusion chromatography. Further samples of the same DNA were digested with BamHI, ClaI, HindIII, PstI, SmaI or XhoI, electrophoresed in wide (eg. 5 cm) gel slots, and Southern blotted onto nitrocellulose (section 3.8). Following transfer and fixation of the DNA, the filters were cut into strips 1.5-2 mm in width such that a large number of identical strips was obtained. Before cutting, the filter was marked by ruling lines across it to facilitate later realignment of strips. Groups of strips, one from each digest, were sealed in plastic bags and pre-hybridised overnight (section 3.11.1). Thus for each probe there was a bag containing a single strip of filter from each digest. Probes were denatured and added to the pre-

hybridisation solution. Hybridisation was continued overnight, after which the bags were opened and all strips, which had initially been carefully identified, were washed stringently. The strips were air dried and mounted on paper in such a way that the original filter was "reconstructed" (ie. all strips derived from the same filter were placed adjacently, successive ones having been hybridised to successive probes). This array of filter strips was autoradiographed for various exposure times. Figure 6 shows an autoradiograph illustrative of those obtained using this methodology.

The autoradiographs were "read" by noting which fragments generated by one enzyme ("probes") hybridised to which fragments generated by a second. Only the strongest signals were noted since often faint signals to other bands occurred. This appears to have been due to the fact that fragments isolated from gels in the manner described are not pure but contain small quantities of the other fragments generated, since hybridisations with cloned fragments in later experiments did not give this effect. Signal strength is proportional to the extent of the overlap between probe and target sequences for probes generated by random priming. Thus the danger here was that of ignoring significant, though weak, hybridisations rather than using misleading data.

The target of interpretation of these hybridisation experiments was to derive a linear map of restriction sites along the viral genome from the data indicating which DNA

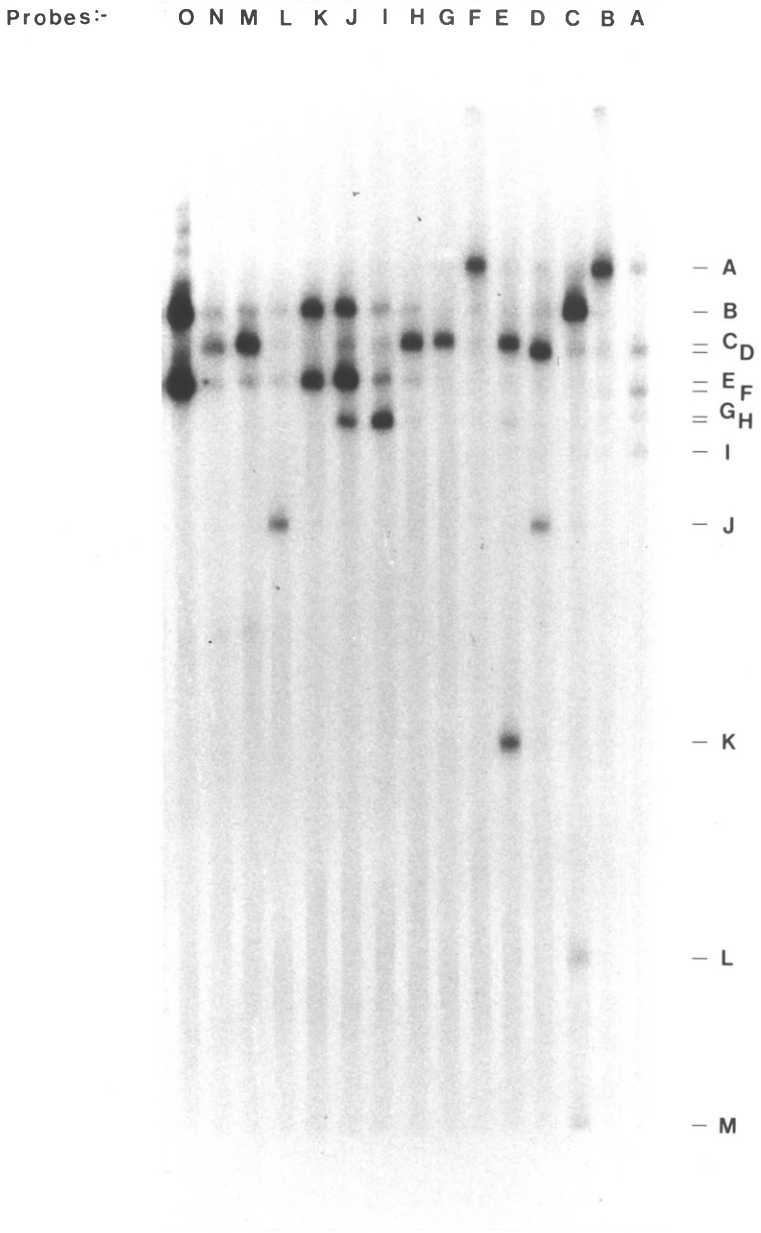


Figure 6. An autoradiograph illustrative of those obtained during the restriction mapping analysis described in section 5.2. This example was obtained by probing strips of a Southern blot of a BamHI digest of MCV I DNA with HindIII fragment probes A-O and realigning the strips prior to autoradiography. The probe fragment is indicated at the top of each strip whilst the hybridising BamHI fragments are labelled at the side. The data from this experiment are tabulated in appendix 3A (i). In practice, a number of varying autoradiographic exposures were necessary to extract all of the information from such an experiment.

fragments generated by one restriction enzyme hybridised to which fragments in a second digest, the rationale being that hybridisation indicates that fragments overlap (ie. co-map) in the genome. The only hybridisation of fragments for which this interpretation would be expected not to be true is in the case of terminal fragments: such fragments from one end of the genome were expected to hybridise to fragments derived from the other end because of the terminal inverted repeat whose presence was inferred above. This and other instances of repetition, leading to hybridisation of fragments which do not co-map, should be apparent from control experiments in which the labelled fragments generated by a given enzyme were hybridised to digests with the same enzyme.

5.3 Construction of MCV I restriction maps: interpretation of hybridisation data

The mapping data discussed below were obtained from the following experiments, performed as described above:-

(1) HindIII fragments A-O used as probes to Southern blots of DNA digested with BamHI, ClaI, HindIII (control), PstI, SmaI and XhoI.

(2) ClaI fragments B, C, F and G used as probes to Southern blots of DNA digested with BamHI, ClaI (control), HindIII, PstI, SmaI and XhoI.

(3) BamHI fragments A-O used as probes to Southern blots of DNA digested with BamHI (control), ClaI and HindIII.

Hybridisation data obtained from these experiments are consistent and are given in appendix 3A. They were interpreted with reference to fragment sizes and molarities (table 7). Two main problems in data interpretation were caused by the inability to resolve hybridisation signals to closely spaced fragments on the autoradiographs and due to the fact that some fragments are 2-molar. Particularly problematic were BamHI fragments C and D which are both 2-molar and which were not resolved sufficiently to allow their individual use as probes; thus the combined C and D fragments were used. Signals resulting from hybridisation to these fragments could however be resolved when BamHI digests were probed with other

fragments. Also problematic was the 3-molar ClaI fragment A which accounts for approximately half of the genome size; consequently the ClaI fragments did not provide many useful linkages during interpretation.

Control hybridisations indicated that all labelled fragments were correctly identified. These controls also confirmed the presence of a terminal repetition since the genomic termini cross-hybridised. Thus eg. BamHI fragments B and E cross-hybridised; also, both these fragments hybridised to HindIII fragment O, ClaI fragments A and E, PstI fragments C and G, SmaI fragment N, and XhoI fragments B and C. All of these fragments were previously identified as terminal fragments on the basis of their ability to renature rapidly due to terminal cross-linking (section 5.1). The only instance of non-terminal fragments generated by the same enzyme cross-hybridising was that of HindIII fragments J and K (both 2-molar). In addition to cross-hybridising, these fragments also hybridised to both terminal fragments generated by BamHI, ClaI, PstI, SmaI and XhoI, all of which are larger than the terminal HindIII fragment O. Since the size of the terminal repeat is less than that of the sum of the sizes of Hind/O and Hind/J or K, the only possible interpretation of this observation is that one component of each of HindIII fragments J and K maps one at each end of the genome, adjacent to the terminal O fragments and partially within the terminal repeat: they thus cross-hybridised by virtue of sharing repeat sequences and the situation arose because there is a HindIII site within the repeat. There was no other instance of

hybridisation of a given fragment generated by one enzyme to other fragments generated by the same enzyme, indicating the absence of internal repetitions.

Restriction maps were constructed by forming linkage groups of overlapping fragments generated by different enzymes. These were extended from the two genomic ends, identified as discussed above, and from hybridisations centred on HindIII fragment B. Eventually the groups merged to create a map for the entire genome. This is complete for the restriction endonucleases BamHI, ClaI and HindIII but only partial for PstI, SmaI and XhoI since these enzymes generated large numbers of fragments. An account of the process of deriving and extending linkage groups follows. The enzymes mentioned above are abbreviated to Bam, Cla, Hind, Pst, Sma and Xho respectively and restriction fragments are identified as eg. Bam/E for BamHI fragment E. Refer to the completed map shown in figure 7. It should be noted that the terms "left" and "right" as applied to the relative orientation of fragments within linkage groups and ultimately to the orientation of the genome were chosen arbitrarily, although for ease of presentation they are used here as they refer to the completed maps.

Hind/B hybridised to Bam/A and Cla/A (3-molar) and no other fragments generated by these enzymes. Bam/A also hybridised to Hind/A and Hind/F, of which Hind/A hybridised to Cla/A but Hind/F hybridised to Cla/C. Bam/A also hybridised to Cla/F, as did Hind/A, but the latter did not hyridise to

Cla/C. These data established a linkage group based on Bam/A; Hind/A, B and F (left to right); and Cla/F, A and C. Hybridisations to Xho/A confirmed that Hind/A and B map adjacently, and those to Xho/M confirmed that Hind/B and F are adjacent. (Although Xho/M is 2-molar, one component could be mapped elsewhere on the basis of hybridisation to Hind/J and K; see below.) The following analysis extended the linkage group leftwards. Hind/A hybridised to Cla/A (3-molar), D or E (2-molar) and G in addition to F, and to Bam/D (2-molar), F, G, I and N (2-molar) in addition to A. Sma/H linked Cla/G to the left of Cla/F and Bam/F linked Cla/D to the left of Cla/G. Of the remaining BamHI fragments hybridising to Hind/A (namely Bam/D, G, I and N), Bam/G, I and N did not hybridise to Cla/D indicating that Bam/D, which did hybridise to Cla/D, maps to the left of Bam/F. (Although Bam/D could only be used as a probe in combination with Bam/C for the reason mentioned above, the fragment mapped here is definitely D and not C as determined by the hybridisation of Bam/C+D to Cla/D and the hybridisation of Hind/A to Cla/D and Bam/D but not Bam/C.) The relative order of the fragments Bam/G, I and N was not determined but they extended the linkage by placing Cla/A to the left of Cla/D. Since no other fragments hybridised to Hind/A, Bam/I and G hybridised to no HindIII fragments other than Hind/A, Bam/N is 2-molar, and Cla/A is 3-molar, the progress of linkage to the left was halted at this point.

The following analysis extended the linkage group developed above to the right. Hind/G maps to the right of Hind/F since both these fragments hybridised to Cla/C.

Hind/G, E, I and J (2-molar) were linked by Sma/A; and Hind/E, I, J and N were linked by Cla/B. Hind/E maps to the right of Hind/G since both hybridised to Bam/C (2-molar) whilst Hind/I and J did not. This was confirmed by the hybridisation of Hind/E, F and G to Pst/A. Cla/C also linked Bam/C adjacent to Bam/A. Hind/E located Bam/K and O to the right of Bam/C, but their order was not determined. Hind/I and J are adjacent since both hybridised to Bam/H. Their order was determined by hybridisation of Hind/E and I to Xho/D. (Although the latter is 2-molar, one component maps to Hind/A; and whilst the second component may not be used to establish a linkage, it can be used to determine the relative orientation given the linkage.) Cla/B linked Hind/N to the right of Hind/J which was confirmed by the hybridisation of Hind/J and N to Xho/F. (Although Xho/F is 2-molar, one component maps to Hind/A.) Hind/N linked Cla/H to the right of Cla/B, and Cla/H linked Hind/M to Hind/N. Hind/M linked Cla/E rightwards. (This linkage must be to Cla/E (2-molar) and not D, although these fragments were not resolved in this experiment, since Cla/D has already been mapped by Bam/F as described above.) Linkage of Hind/H, M and N by the second Bam/C and hybridisation of Hind/H to Cla/E placed Hind/H to the right of Hind/M and Bam/C to the right of Bam/H. Sma/P confirmed the linkage of Hind/M and N. (Although Sma/P is 2-molar, one component was mapped elsewhere on the basis of hybridisation to Hind/J and K, either partially within the terminal repeat, in which there is a site for SmaI, or close to the right terminus - see below). Since no other HindIII fragments hybridised to Bam/C, and Cla/E is 2-molar, progress rightwards was halted at this

point.

The result of the above analysis was to create a linkage group of length approximately 140 kb, extending from Bam/G, I and N, Hind/A and Cla/A at the left end to Bam/C, Hind/H and Cla/E at the right. Linkage groups were also derived from the genomic termini, as follows. Hind/O (2-molar) is the HindIII fragment at each terminus, as determined by the observation that this fragment alone in a HindIII digest has the ability to renature rapidly and thus possesses the covalent cross-link (section 5.1). Similarly, Bam/B and E, and Cla/A (3-molar) and E (2-molar) are terminal fragments. Hind/J and K are present just internal to and adjacent to Hind/O, one at each end (see above). It was not possible at this stage to determine which of Hind/J and K occurred at which terminus due to their cross-hybridisation and size similarity. Cross-hybridisation occurred by virtue of each containing part of the terminal repeat, the boundary of which falls within these fragments. Bam/B and Cla/A must occur at the same terminus (with Bam/E and Cla/E at the other) since Bam/B hybridised to Hind/C which hybridised to Cla/A and not E (the third Cla/A and second Cla/E are the only ClaI fragments still not mapped). The observation of both Hind/J and K (as well as O), and both Cla/A and E hybridising to Bam/B is due to the presence of all or part of the terminal repetition within these fragments. The terminus consisting of Hind/O, Bam/B and Cla/A was arbitrarily designated as the left terminus. Hind/Q maps to Bam/B and hybridisations with Hind/C linked Bam/L, M and N (the order of which was not determined)

to the right of Bam/B. Sma/J linked Hind/L to the right of Hind/C and Bam/J linked Hind/D to the right of Hind/L. The latter linkage was confirmed by hybridisation to Pst/K and to Xho/O. (Although the latter is 2-molar, one component maps to Hind/A and Cla/F.) Hind/D linked Bam/D to the right of Bam/J. Hind/C, D and L, and Bam/D and J all hybridised to Cla/A confirming that Cla/A is the left-terminal ClaI fragment. Since Bam/D and Hind/D hybridised only to Cla/A but overhang this terminal Cla/A fragment then this linkage group extended from the "left" genomic terminus must join to the left end of the linkage group derived earlier, which has a protruding Cla/A. (The third Cla/A maps internally to this latter group.) The sizes for such a join fit well and the linkage was confirmed by hybridisation of Hind/A and D to Sma/E. (Although the latter is 2-molar, one component was mapped elsewhere in the same manner as for Sma/P, mentioned above.)

Apart from Hind/P and Bam/P, for which no hybridisation data were obtained, the only fragments still to be mapped (which must occur at the right end of the genome) are Hind/K, the second Hind/J or K, the second Hind/O, Bam/E and the second Cla/E. The right terminal fragments are Bam/E, Hind/O and Cla/E (see above). Thus Bam/E maps to the right of Bam/C at the right end of the linkage group derived above, and Cla/E similarly maps to the right of the first Cla/E fragment. As mentioned above, Hind/J or K occurs adjacent to the terminal Hind/O leaving the second Hind/K mapping to the right of Hind/H and to the left of Hind/J or K. This analysis of these fragments is in agreement with the hybridisation data obtained

for them. Which of the remaining Hind/J and K fragments occurs at which terminus was resolved on the basis that both hybridised to Xho/M. (Although the latter is 2-molar, one component could be mapped elsewhere due to hybridisation to Hind/B and F.) Since both terminal XhoI fragments (Xho/B and C) are larger than the terminal repeat and thus Xho/M does not contain repeat sequences, this hybridisation with Xho/M places one component of each of Hind/J and K adjacent to each other. Thus Hind/K occurs at the left terminus whilst Hind/J occurs at the right terminus, adjacent to the second Hind/K fragment already mapped there.

The end result of the above analysis is the ordering of all BamHI fragments, all ClaI fragments and all HindIII fragments into a linear restriction map for the MCV I genome (see figure 7), with the following exceptions: Hind/P and Bam/P were not placed and the relative orientations of Bam/G, I and N, of Bam/L, M and N, and of Bam/K and O are unknown.

A number of PstI, SmaI and XhoI fragments were mapped above since they formed useful linkages. Many more were now mapped by virtue of their hybridisation to HindIII and ClaI fragment probes and with reference to the derived restriction maps for these enzymes. The fragments so mapped are included in figure 6. Since the fragments generated by PstI, SmaI and XhoI were numerous and mostly small, the majority of them were not mapped precisely. Which of Pst/C and G occurs at the left or right terminus was deduced from the presence of a PstI site within a clone of the near left-terminal component of Hind/K

(clone pH/K_L-I; see section 8.3).

5.4 Other data used in MCV I restriction map construction

As mentioned above, hybridisation signal strength was proportional to the length of overlap between probe and target (with the notable exception of hybridisation between fragments containing multiple repeat sequences from the terminal inverted repeat). Thus, for example, Hind/M hybridised to Cla/H more strongly than to Cla/E, indicating a greater overlap with the former although this fragment is the smaller of the two. Also, Hind/G hybridised strongly to Pst/A while Hind/E and F hybridised comparatively weakly, consistent with Pst/A spanning Hind/G entirely and Hind/E and F only partially. The data given in appendix 3 and used in the above deduction of the MCV I restriction maps refer to those hybridisation signals that were strong enough to be relied upon. Weak signals were ignored lest they be due to probe contamination (restriction fragments isolated from gels are not completely free of other fragments) or "background" hybridisation. However, weak signals could have resulted from significant hybridisations for which fragment overlap was small. Autoradiographs were reviewed at this point in order to "fine tune" the maps - ie. in regions of the maps where restriction sites for two enzymes had been placed near coincidently the question was asked as to whether fragments either side of the site for the first enzyme hybridised weakly or not at all to those on the opposite side in digests with

the second enzyme. Reference was made throughout to control hybridisations to confirm probe purity. The only modifications were slight adjustments of the relative positions of BamHI and HindIII sites to allow slight overlap of Bam/H with Hind/E and Bam/C with Hind/J, and of the relative positions of HindIII and ClaI sites to allow slight overlap of Hind/F and Cla/A. The following sites are very nearly coincident, as judged by a lack of hybridisation as described: the BamHI site between Bam/J and L, M or N, and the HindIII site between Hind/C and L; the BamHI site between Bam/F and A, and the ClaI site between Cla/G and F; the BamHI site between Bam/A and C, and the HindIII site between Hind/F and G; and the BamHI site between Bam/C and E, and the HindIII site between Hind/H and K. The question could not be answered for the BamHI site between Bam/D and G, I or N, and the HindIII site between Hind/A and D since Hind/A contains the other Bam/D fragment and so a strong hybridisation signal already existed.

During the initial set of hybridisations with HindIII fragment probes, a control set of hybridisations to a Southern blot of a HindIII digest was used. This digest was unintentionally incomplete to the extent that, although it was still adequate as a control to verify the probes, it also provided some mapping information which corroborated the maps derived above. The following partial digest products were observed: Hind/J+K, J+K+O, J+O, K+O, C+L, E+I, E+G, F+G, H+M, M+N, H+M+N, C+L+D. Their identities were determined by hybridisation to their constituent fragments and comparison with their expected sizes. Hind/C+L was an abundant partial

digest product. The HindIII site between Hind/C and L appears to be particularly refractory since in HindIII digests visualised in agarose gels a sub-molar band of this size (which was not present in more extensive digests) was not uncommonly observed, despite the absence of other partial digest products. In the hybridisation experiment with BamHI probes the HindIII digest had such a band which hybridised to Bam/B, J, L, M and N, again confirming linkage of Hind/C and L.

Another source of data pertaining to restriction mapping was variability observed between different viral isolates. Genetic heterogeneity occasionally resulted in isolates lacking restriction sites or possessing novel sites relative to the isolate mapped here. This is discussed more fully elsewhere (section 5.9) but suffice to say that instances of the former were loss of the BamHI site between Bam/D and F, loss of the BamHI site between Bam/D and I, and loss of the HindIII site between Hind/E and G. This was deduced by the absence in viral DNA digests of the fragments mentioned and the presence of a novel fragment equal in size to the sum of the two. All cases mentioned are consistent with the maps as derived above. However, the case of loss of the BamHI site between Bam/D and I actually provides useful new data since the relative order of Bam/G, I and N was not determined from the hybridisation analysis. Thus Bam/I can be mapped adjacent to Bam/D, whilst the relative positions of Bam/G and N remain unknown.

The final source of data used in restriction map construction was that obtained from double-digests of MCV I DNA. The three double-digests with BamHI and ClaI, ClaI and HindIII, and BamHI and HindIII were done (using high, medium and high salt buffers respectively and incubating with both enzymes concurrently). Single-digests were done at the same time and electrophoresed on the same gel, allowing comparison of each double-digest with the two relevant single-digests. This allowed identification of which fragments generated with one enzyme were or were not further digested by the second enzyme, and vice versa, and allowed the sizes of novel digest products to be obtained. The latter were of value in determining the extents of overlap between large fragments whose sizes were not accurately known due to the limited resolving power of agarose gel electrophoresis. Bam/A, B, C (both), E, H, J and L were cleaved by HindIII; Hind/A, C, D, E, J (one only), K (one only) and L were cleaved by BamHI. Cla/A (all three), B, C, E (both) and H were cleaved by HindIII; Hind/A, D, F, G, K (one only), M and N were cleaved by ClaI. Bam/A, C (both), D (both), E and F were cleaved by ClaI; Cla/A (two only), B, C, D and E (one only) were cleaved by BamHI. Other fragments appeared not to be cleaved by the second enzyme; however, only for those fragments known to map well within fragments generated by the second enzyme was it possible to state so definitively since a comigrating fragment could result from cleavage of an initially larger fragment and because removal of short sequences would not significantly alter mobility. These data are consistent with the hybridisation data discussed above in that fragments that hybridised

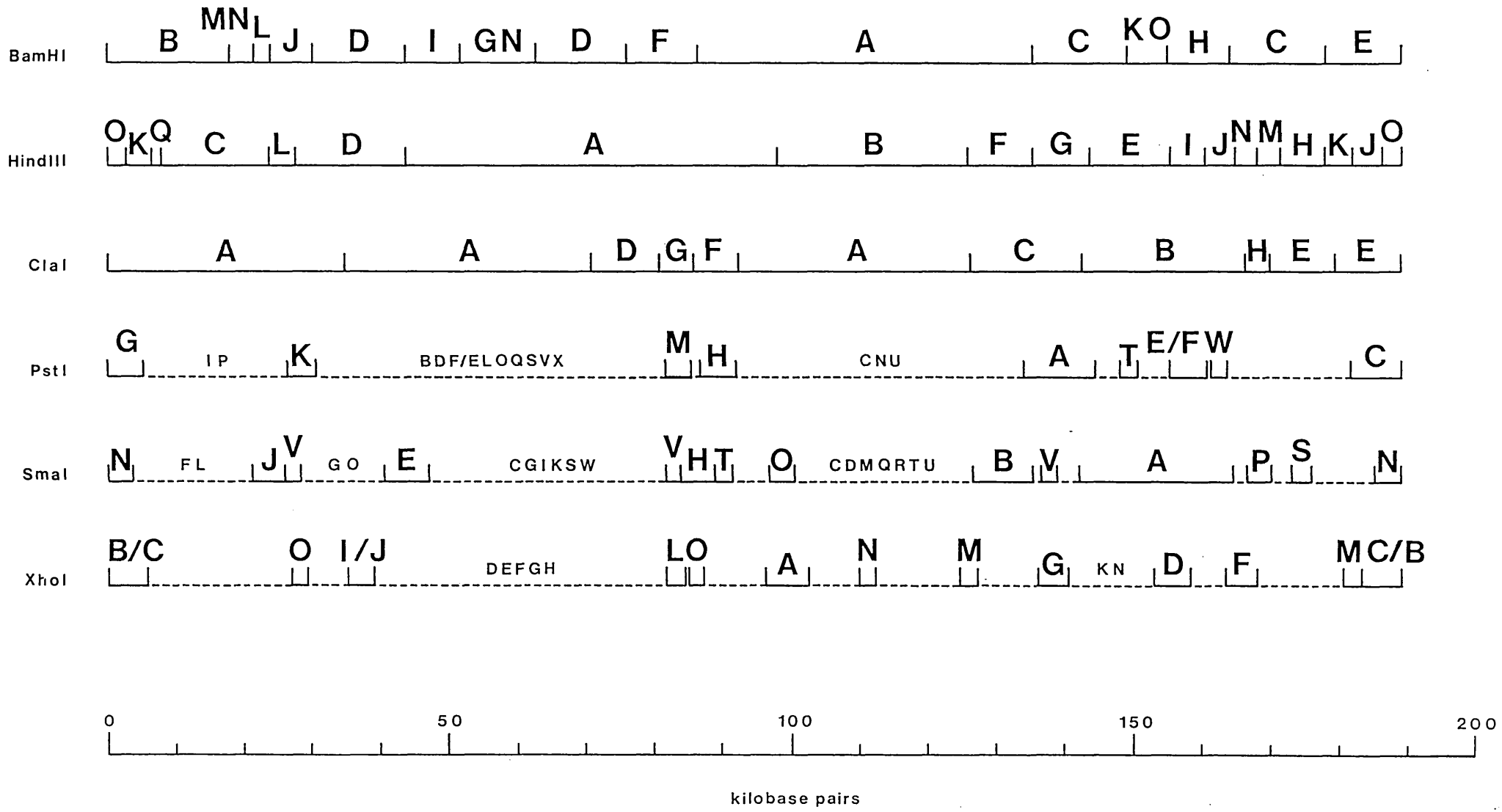
to at least two fragments in digests with the second enzyme alone were cleaved by that enzyme. Several novel fragments were present in double-digests. In general, their sizes were in accord with those expected on the basis of the maps generated above. The two largest novel fragments in the BamHI/HindIII double-digest were 11.5 kb and 9.7 kb. The latter was known to be the overlap of Bam/B and Hind/C since a clone of this region of the genome was obtained (clone pMBH-I; see section 6.4): the former can only have been derived from the overlap between Bam/A and Hind/A since Hind/B was not cleaved with BamHI, Hind/D contains Bam/D, and all other overlaps are too small. This datum is useful because the sizes of Bam/A and Hind/A could not be obtained accurately by agarose gel electrophoresis: with this new information, along with the mapping information already described, their sizes were determined to be 49 kb and 54 kb respectively. Similarly, the largest novel fragment in the ClaI/HindIII double-digest was 28 kb (co-migrating with Hind/B) which can only have been derived from the overlap of Hind/A and Cla/A. Again, this was useful in establishing the extent of overlap between two large fragments whose sizes were only known approximately. As expected from the maps, one of the three Cla/A fragments was not cleaved with BamHI. Its size was estimated as 36 kb from the BamHI/ClaI double-digest. (This figure is more accurate than that obtained from the ClaI digest alone because in the latter case three viral fragments and uncleaved cell DNA have very similar mobilities.) Another point of value to be extracted from the double-digest data was the observation that Bam/L, but not M or N (the relative map

positions of these three fragments were not determined in the above analysis), was cleaved by HindIII. Thus Bam/L maps at the right of this grouping, to the left of Bam/J and containing the HindIII site between Hind/C and L (both of which were cleaved by BamHI). Finally, the cleavage of one of the Hind/K fragments by BamHI determined that the BamHI site between Bam/C and E is to the right of the HindIII site between Hind/H and K.

The restriction maps for MCV I DNA derived with reference to all of the foregoing data are presented in figure 7.

(Next page)

Figure 7. Restriction maps of MCV I DNA (isolate 1) for the enzymes BamHI, ClaI, HindIII, PstI, SmaI and XhoI. The viral genome is a linear double-stranded DNA molecule of 190 kb. The positions of restriction sites for a given enzyme are indicated by vertical lines and the letters between such sites identify the fragments, which are labelled as in figure 5 and table 7. The maps are complete for BamHI, ClaI and HindIII although there are some instances when the relative orientation of two fragments is unknown (eg. BamHI fragments G and N). In such a case the restriction site is not marked although it is known to occur at one of two possible positions. The restriction maps for PstI, SmaI and XhoI are only partial since there are many sites for these enzymes: those fragments that could be mapped are indicated as above whilst others that could only be assigned to a more general region of the genome are indicated by their identifying letter only. When a fragment is known to be one of two possibilities which were not distinguished it is identified as eg. XhoI fragment B/C.


BamHI
B MN L J D I GN D F A C KO H C E
HindIII
O K Q C L D A B F G E I J N M H K J O
ClaI
A A D G F A C B H E E
PstI
G IP K BDF/ELOQSVX M H CNU A T E/F W C
SmaI
N FL J V GO E CGIKSW V H T O CDMQRTU B V A P S N
XhoI
B/C O I/J DEFGH LO A N M G KN D F MC/B
0
50
100
150
200
kilobase pairs

5.5 Characterisation of MCV II DNA and construction of restriction maps

In the same way as for the isolate of MCV I already discussed, an isolate of MCV II (isolate 26) was selected and studied further. Because BamHI, ClaI and HindIII had proved most useful for examining MCV I DNA these same enzymes were used in the analysis of MCV II DNA. Restriction digests are illustrated in figure 8 and the sizes and relative molarities of fragments generated by these enzymes are given in table 8. The sizes of Bam/A, Cla/A and Hind/A were determined more accurately from the restriction maps derived below. The totals obtained from summation of fragment sizes for the three enzymes used were in agreement and gave a genome size of 200 kb, compared to 190 kb for that of MCV I.

MCV II DNA was shown to possess covalently closed termini in the same manner as for MCV I DNA. The rapidly renaturing terminal HindIII fragments were identified as Hind/I (5.3 kb) from each end of the genome. Similarly, the SmaI terminal fragments were 7.2 kb at each end, indicating the presence of a terminal inverted repeat of >7.2 kb. The relative molarity of Hind/I (which was known to be >1 on the basis of the intensity of ethidium bromide staining) was determined to be 3 according to the following analysis. Three HindIII fragments showed size variation between different isolates of MCV II, namely those identified for isolate 26 as Hind/H and I and a fragment that co-migrated with I in this isolate. Thus, Hind/I in other isolates is 2-molar and

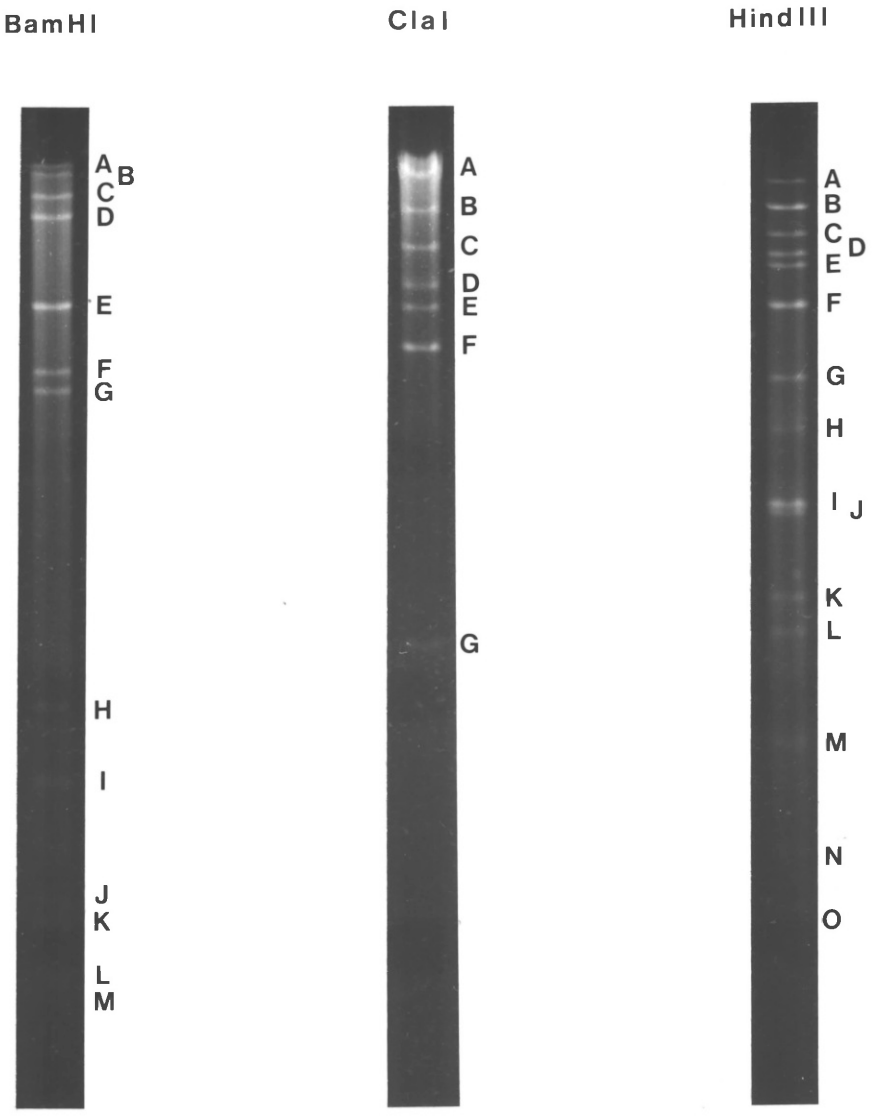


Figure 8. Restriction digests of MCV II DNA (isolate 26) with BamHI, ClaI and HindIII. Fragments are labelled alphabetically as in table 8.

	BamHI	ClaI	HindIII
A	57	38,37,37	33.5
B	41	22	23 (2)
C	28.5	16.5	18
D	21	13.3	15.5
E	11.5 (2)	11.5	14.5
F	9.0	9.7 (2)	11.5 (2)
G	8.2	3.6	8.4
H	3.1		6.9
I	2.5		5.3 (3)
J	1.8		5.1
K	1.6 (2)		3.9
L	1.4		3.5
M	1.3		2.5
N			1.7
O			1.4
P			
Q			
R			
S			
kb	200.5	198.3	199.3

Table 8. Sizes (kilobase pairs) of restriction fragments of MCV II DNA (isolate 26) generated by BamHI, ClaI and HindIII. Refer to table 7 for details. All fragments visualised by gel electrophoresis are listed and the estimated genome size is given at the bottom of the columns.

consists of the two terminal fragments, since HindIII cuts within the terminal repeat. The third component of Hind/I in the isolate used here for mapping (now referred to as Hind/I') showed size variation between isolates similar to that of Hind/H: in fact, for the six isolates discussed below (section 5.9), the size of Hind/H was always 1.6 kb greater than that of Hind/I' although both fragments varied in size between isolates by up to 1 kb. This is suggestive of both Hind/H and I' containing part of the terminal repetition and mapping one at each end of the genome, adjacent to the terminal Hind/I fragments. Size variations of the terminal Bam/D and Cla/D fragments reflected the variation in the sum of the sizes of Hind/I and H/I'. An explanation for size variability of Hind/H, I and I' would be the possession by these fragments of variable numbers of tandem repetitions, as was inferred above for the MCV I terminal fragments. Consistent with this mapping of Hind/H and I' is the observation that Hind/H and I (including I') cross-hybridised (see below), analagous to the cross-hybridisation of MCV I Hind/J and K, and that Hind/H and I hybridised to both terminal BamHI and ClaI fragments. This situation differs from that for MCV I, however, in that MCV I Hind/J and K were of constant size and only the terminal HindIII fragment (Hind/O) showed size variation between isolates. This implies that the HindIII site within the terminal repeat of MCV II occurs between two blocks of tandem repetitions, whilst all such repetitions occur on the terminal side of the HindIII site within the MCV I terminal repeat. It would be expected that MCV II Hind/H and I' would cross-hybridise to Hind/I in addition to each other if the

repetitions either side of the HindIII site were of similar sequence. However, this could not be answered from the hybridisation data obtained in this study due to the co-migration in gels of Hind/I and I' for this isolate. MCV I Hind/J and K would not be expected to cross-hybridise with MCV I Hind/O (and they did not) since they do not contain such repetitions. The occurrence of two blocks of tandem repeats within the terminal repeat of the vaccinia virus genome has been reported (Wittek and Moss, 1980) and provides a precedent for the observations made here for MCV II. Since variations in the size of Hind/I and of Hind/H and I' are independent, the number of tandem repetitions in the blocks either side of the HindIII site is shown to vary independently. The maximum size of the terminal repeat of MCV II DNA is given by the sum of the sizes of Hind/I and I' (ie. the smaller of Hind/H and I'). Thus this isolate of MCV II DNA has a terminal inverted repeat of 7.2-10.6 kb, compared to 3.9-5.4 kb for that of MCV I.

Using the same approach as above for MCV I, restriction fragments were isolated and used to probe digests of MCV II DNA. Those used were:-

(1) HindIII fragments A-M used as probes to Southern blots of DNA digested with BamHI, ClaI and HindIII (control).

(2) ClaI fragments B-G used as probes to Southern blots of DNA digested with BamHI, ClaI (control) and HindIII.

(3) BamHI fragments C-K used as probes to Southern blots of DNA digested with BamHI (control), ClaI and HindIII.

Hybridisation data obtained from these experiments are consistent. The data are given in appendix 3B and were interpreted according to the same procedures described for MCV I (section 5.3), as outlined below. Control hybridisations confirmed the identification of terminal fragments since these cross-hybridised. The only instance of non-terminal fragments cross-hybridising to other fragments generated by the same enzyme was that of Hind/H and I, which also hybridised to the BamHI and ClaI terminal fragments. The situation arose by virtue of there being a HindIII site within the terminal repeat, as discussed above. As for mapping the MCV I isolate, use was made of data obtained from double-digests (BamHI and ClaI, ClaI and HindIII, and BamHI and HindIII) in addition to hybridisation data. Bam/A, B, C and H were cleaved by ClaI; Cla/A (all three), C and F (one only) were cleaved by BamHI. Cla/A (all three), B, C, D and F (one only) were cleaved by HindIII; Hind/A, C, E, G and H were cleaved by ClaI. Bam/A, B, C and D were cleaved by HindIII; Hind/A, B (both), C and D were cleaved by BamHI. Consistent with these data, those fragments generated by one enzyme that were further digested with the second hybridised to two or more fragments generated by the latter. Other fragments were apparently not cleaved by the second enzyme.

Hind/C hybridised to Bam/E (2-molar), G and H, and to Cla/A (3-molar) and C. Of these, Bam/G hybridised to Cla/C,

Bam/E to Cla/A and Bam/H to Cla/A, establishing a linkage group based on Hind/C, Bam/E, H and G, and Cla/A and C (left to right). As before, the terms "left" and "right" were chosen arbitrarily. This linkage group was extended rightwards by the additional hybridisation of Cla/C to Hind/G and Bam/A, the latter being the right terminal BamHI fragment. The right terminal HindIII and ClaI fragments are Hind/I (the same at both ends) and Cla/A or D. Bam/A also hybridised to Hind/F (both), H, J, K and M, of which Hind/F (2-molar) hybridised to Cla/B, F and G. Thus, in order to accommodate these other fragments, the right terminal ClaI fragment must be Cla/D. Hind/K linked Cla/F to the left of Cla/D and, since Hind/F but not Hind/H, J or M was cleaved by ClaI, Hind/F maps to the left of Hind/K. The relative orientations of the second Hind/F, Hind/J and M, and of Cla/B and G were not determined. Hind/H or I (ie. I') lies between Hind/K and I, Hind/H being favoured on the basis of the size available.

Bam/E extended the linkage leftwards to either Hind/A or B (2-molar). Hind/A hybridised to Bam/B and E (2-molar), and to Cla/E and F (2-molar). Cla/E hybridised to Bam/B but not E, whilst Cla/F hybridised to Bam/E but not B, establishing a linkage group based on Hind/A, Bam/E and B, and Cla/F and E (left to right). Thus Hind/B must be to the left of Hind/C since there is insufficient room to accommodate Cla/E and F (in addition to Cla/A already mapped there) if it were Hind/A. Hind/B linked Bam/B to the left of Bam/E and thus joined to the linkage group developed from Hind/A. The linkage of Hind/A and B by Bam/B was confirmed from the

BamHI/HindIII double-digest data, discussed below.

The only ClaI fragments not mapped are the other two Cla/A fragments which therefore extend the linkage to the left terminus. The left terminal BamHI fragment is Bam/D. Hind/D linked Bam/I, J and K (one) to the right of Bam/D but their relative order was not determined. Hind/H or I (ie. I') lies between Hind/I and D, Hind/I being favoured on the basis of the size available. The only fragments remaining to be mapped are Bam/C, F, the second K, L and M, and the second Hind/B, Hind/E and Hind/L, of which only Bam/C and Hind/E were cleaved by ClaI. Thus these latter fragments must span the ClaI site between the two Cla/A fragments, placing Bam/C to the right of Bam/I, J and K, and Hind/B to the left of Hind/A (confirmed by the overhang of Bam/E to the left of Hind/A and its hybridisation to Hind/B and not Hind/E or L). The relative orientation of Hind/E and L could not be determined from the hybridisation data or those obtained from the BamHI/HindIII double-digest. However, the HindIII fragment corresponding to Hind/L was subsequently cloned from another isolate of MCV II (clone pH/M-IIB; see section 7.3) and found to contain a BamHI site very close to the HindIII site at one end, indicating that Hind/L maps to the right of Hind/D, with Hind/E to its right. The positions of the second Bam/K, Bam/L and Bam/M, all of which hybridised to Hind/B, could not be unequivocally determined since the latter is 2-molar, both components of which are entirely contained within Cla/A fragments. From the hybridisation data, any of Bam/K, L and M could lie between Bam/B and the Bam/E fragment to its right, Bam/C and F, or

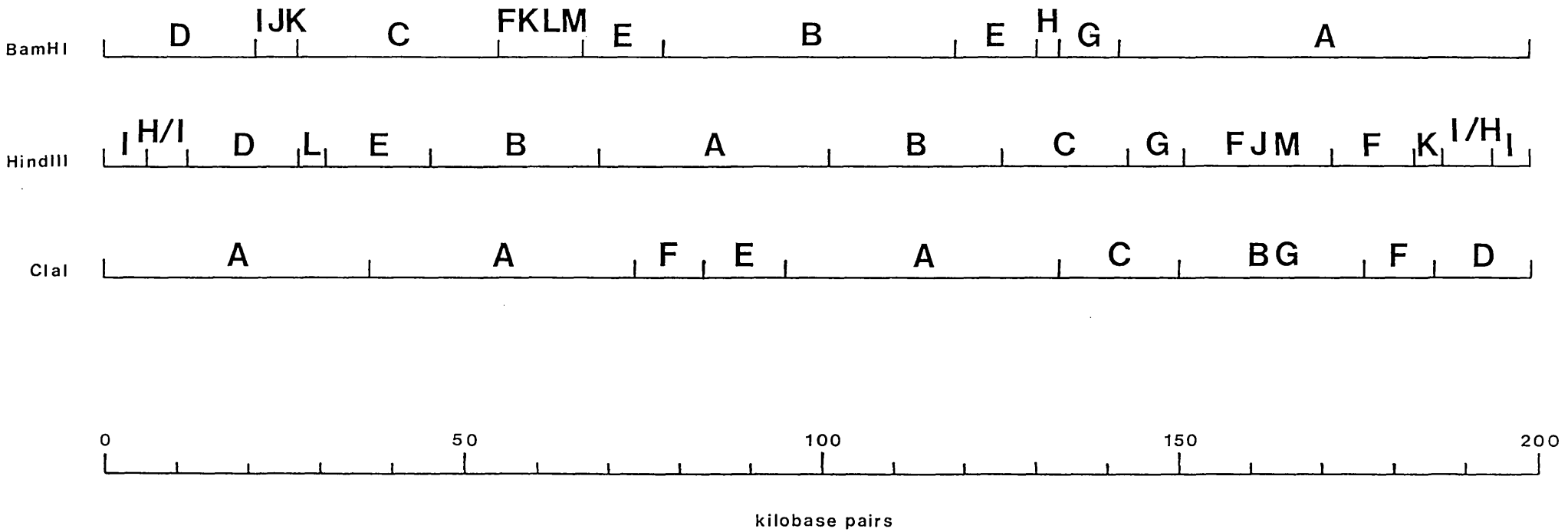
Bam/E and F. However, they could be accommodated most easily at either the second or third of these possible loci. Thus the relative ordering of Bam/F, K, L and M was not determined. Hind/N and O were not mapped, no hybridisation data being obtained for these fragments.

As before, the sizes of the largest novel fragments in double-digests were of value in mapping for determining the extents of fragment overlaps. Particularly useful were two novel fragments of 18 and 23 kb in the BamHI/HindIII double-digest. The latter must have been derived from Hind/A since it is bigger than all other HindIII fragments and is thus the overlap of Hind/A with Bam/B. The 18 kb fragment by a similar argument must have come from one of the Hind/B fragments and is thus the overlap of Hind/B with Bam/B. Other novel fragment sizes were in accord with those expected from the derived maps although they could not be identified with certainty. The overlap of Bam/D and Hind/D was deduced to be 9.5 kb since this region of the genome was subsequently cloned (clone pMBH-II; see section 6.4). Finally, two isolates lacking Hind/C and one of the Hind/B fragments, and possessing a novel fragment equal in size to the sum of the two provided additional evidence that Hind/B and C map adjacently. For one of these isolates there was a novel fragment of 31 kb in a ClaI/HindIII double-digest, indicating an overlap of 8 kb for Hind/C and Cla/A since Hind/B is 23 kb long and does not contain a ClaI site.

The restriction maps for MCV II DNA derived from the above analysis are presented in figure 9.

(Next page)

Figure 9. Restriction maps of MCV II DNA (isolate 26) for the enzymes BamHI, ClaI and HindIII. Fragments are labelled as in figure 8 and table 8. Refer to figure 7 legend for details.


BamHI
D
IJK
C
FKLM
E
B
E
H
G
A
HindIII
I
H/I
D
L
E
B
A
B
C
G
FJM
F
K
I/H
I
ClaI
A
A
F
E
A
C
BG
F
D
0
50
100
150
200
kilobase pairs

5.6 Characterisation of MCV III DNA and construction of restriction maps

Only a single instance of MCV III was observed (isolate 8). MCV III DNA was subjected to the same analysis described above for MCV I and MCV II, except that fewer experiments were performed due to limiting amounts of DNA: neither renaturation nor double-digest experiments were possible. Digests with BamHI, ClaI and HindIII are illustrated in figure 10 and fragment sizes and molarities are given in table 9. Accurate sizes of Bam/A and B, and of Cla/A were obtained from the restriction maps derived below. The totals obtained from summation of fragment sizes for the three enzymes used were in agreement and gave a genome size of 196 kb, similar to those for the isolates of MCV I and MCV II studied above (190 and 200 kb respectively). Hybridisation experiments performed were:-

(1) ClaI fragments A-H used as probes to Southern blots of DNA digested with BamHI, ClaI (control) and HindIII.

(2) BamHI fragments D-I used as probes to Southern blots of DNA digested with BamHI (control), ClaI and HindIII.

Hybridisation data obtained from these experiments are consistent and are given in appendix 3C. Terminal fragments were identified from control experiments by cross-hybridisation which, in the cases of MCV I and MCV II, had confirmed identification based on the ability to renature rapidly.

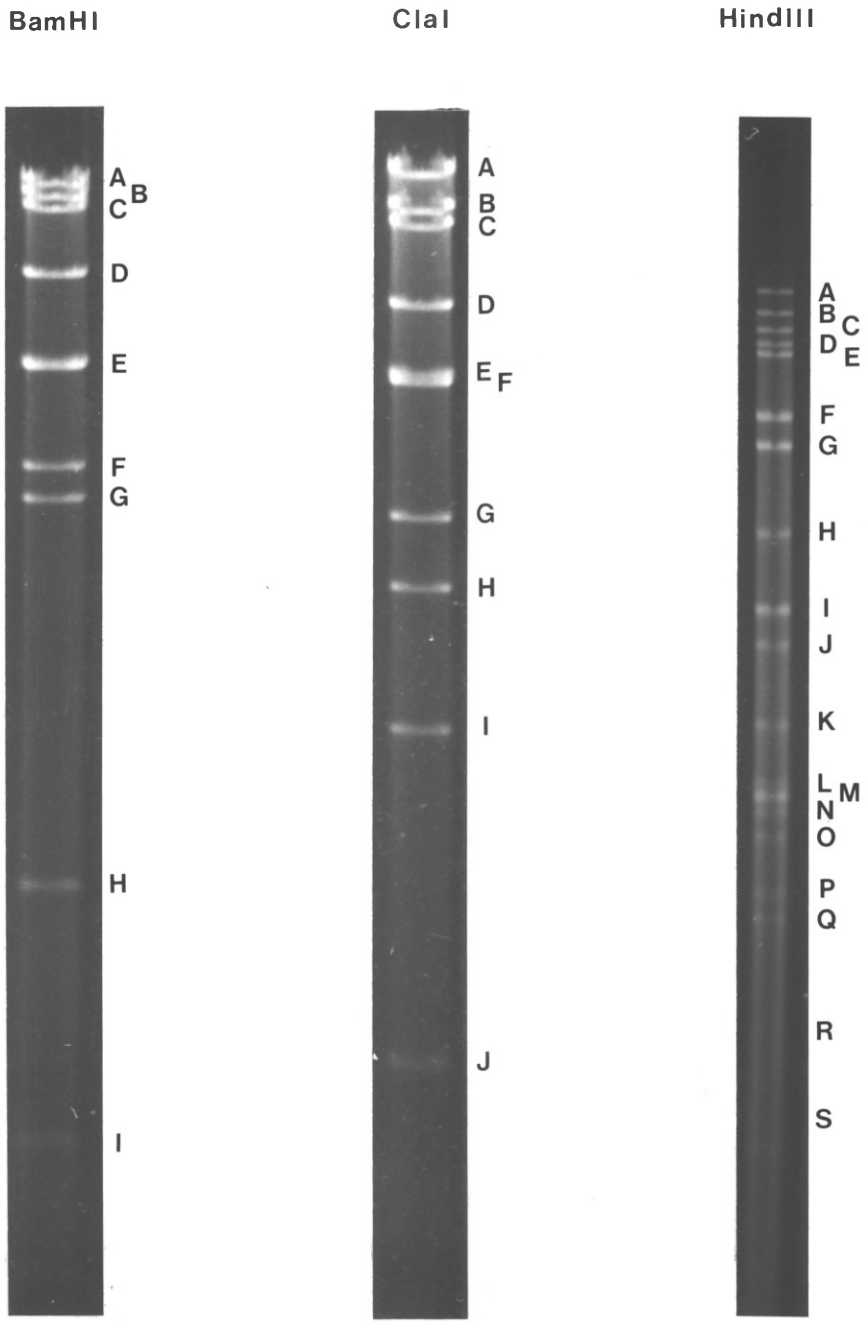


Figure 10. Restriction digests of MCV III DNA (isolate 8) with BamHI, ClaI and HindIII. Fragments are labelled alphabetically as in table 9.

	BamHI	ClaI	HindIII
A	53	73	25
B	42	36	21
C	33	28	19
D	20	16.3	17
E	12.5 (2)	11.6	15.9
F	8.9	11.3	11.3 (2)
G	8.0	7.4	10.1 (2)
H	3.2	6.2	7.4
I	1.7	4.3	5.9 (2)
J	1.2	2.0	5.4
K			4.5
L			3.8
M			3.7 (2)
N			3.5
O			3.3
P			2.8
Q			2.6
R			1.9
S			1.5
kb	196.0	196.1	196.6

Table 9. Sizes (kilobase pairs) of restriction fragments of MCV III DNA (isolate 8) generated by BamHI, ClaI and HindIII. Refer to table 7 for details. All fragments visualised by gel electrophoresis are listed and the estimated genome size is given at the bottom of the columns.

These were Bam/D and E, Cla/A and E or F, and Hind/M (2-molar). The latter was determined by hybridisation with the terminal fragments generated by BamHI and ClaI and indicates an inverted terminal repeat of >3.7 kb. Since the terminal BamHI and ClaI fragments all hybridised to Hind/K, M (2-molar) and O, Hind/M is the terminus with Hind/K and O mapping adacently, one at each end of the genome. Thus the terminal repeat of MCV III DNA contains a HindIII site, as does that of MCV I and MCV II. The maximum size of the MCV III terminal repeat is given by the sum of the sizes of Hind/M and O (ie. the smaller of Hind/K and O). Thus MCV III DNA has a terminal inverted repeat of 3.7-7.0 kb, compared to 3.9-5.4 kb and 7.2-10.6 kb for those of MCV I and MCV II respectively. There were no other instances of cross-hybridisation of fragments generated by the same enzyme.

Construction of linkage groups from the hybridisation data was initiated from the Cla/A terminus, arbitrarily designated the "left" terminus. Cla/A hybridised to Hind/A, as did Cla/G and H. Cla/H linked Hind/F (2-molar) to the right of Hind/A. Cla/G linked Bam/A and E (2-molar) and Cla/H hybridised to Bam/A but not E whilst Cla/A hybridised to Bam/E but not A, establishing that Bam/E is to the left of Bam/A. Cla/C and E or F also hybridised to Bam/A of which Cla/C but not E or F hybridised to Hind/F, placing Cla/C to the right of Cla/H. Cla/C also located Hind/D, one component of Hind/I (2-molar), and Hind/J and L, of which Hind/D linked Cla/E or F to the right of Cla/C. The BamHI and HindIII fragments mapping to Cla/A at the left end of the genome were ordered as

follows. Cla/A hybridised to Bam/C, D, E, F and the second I, of which D or E is at the terminus, and to Hind/A, B, E, G (2-molar), the second Hind/I, and Hind/K, M, O and P, of which M is at the terminus. Of these, Bam/C and D are linked by Hind/G and so the left terminal BamHI fragment must be Bam/D. Hind/B linked Bam/F and I to the right of Bam/C and to the left of Bam/E, and so Hind/E maps between Hind/G and B. Bam/D additionally hybridised to Hind/I and P, and also to Hind/K and O, as did Bam/E, the right terminal BamHI fragment. One of Hind/K and O occurs at each end of the genome adjacent to Hind/M, as discussed above: on the basis of the size available, Hind/K is favoured at the left terminus. The relative orientations of Hind/I and P, Hind/I, J and L, and of Bam/F and I were not determined.

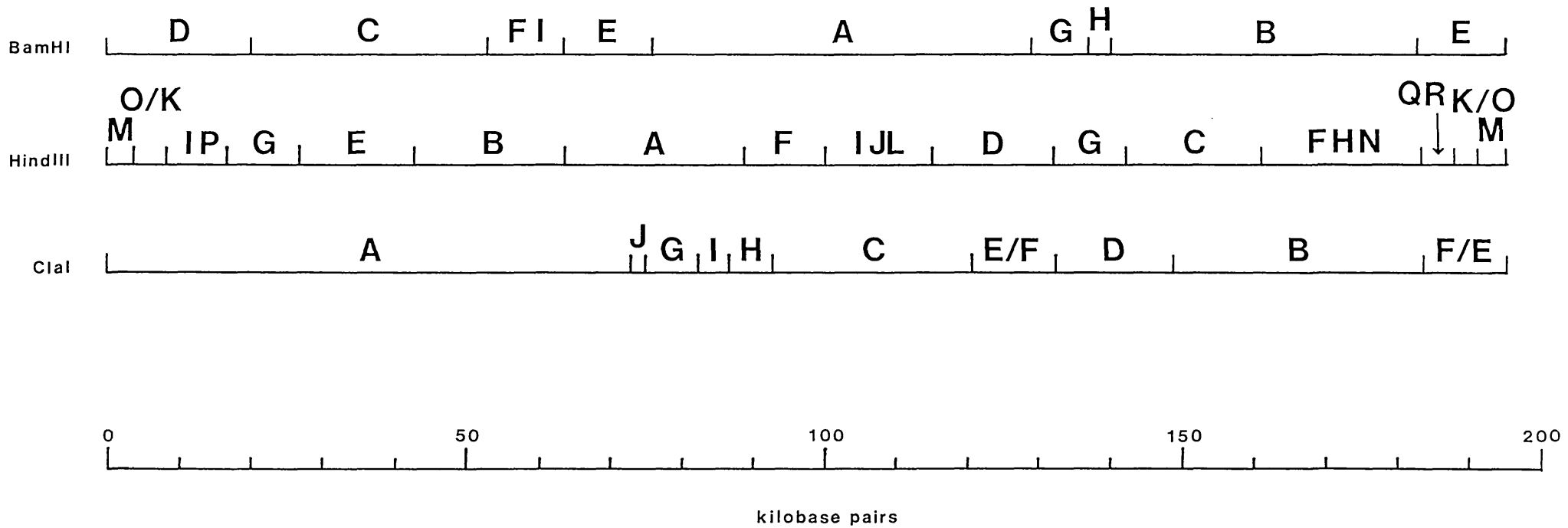
The linkage group was developed rightwards by the hybridisation of Cla/E or F to Bam/A and G. Cla/E or F hybridised to Bam/A, D, E and G: since the terminal component of Cla/E and F would be expected to hybridise to the terminal BamHI fragments (Bam/D and E) and is smaller than either of these, then the other component must link Bam/G to the right of Bam/A. This linkage was confirmed by Hind/D. Bam/G linked Cla/D to the right of Cla/E or F, and Cla/D linked Bam/B and H to the right of Bam/G. Both Bam/B and Hind/C linked Cla/B to the right of Cla/D and the linkage of Bam/G and H was confirmed by Hind/G (one component of the latter having been mapped earlier). Cla/B linked Hind/F, H and N to the right of Hind/C but the order of these latter fragments was not determined. Extension was halted at this point since there

were no further linkages to any of the fragments at the right end of this linkage group. However, Bam/E, Cla/E or F and Hind/K or O, M, Q and R formed a linkage group mapping at the right terminus and thus abutting the linkage group extended from the left terminus. The relative orientaion of Hind/Q and R was not determined. The only fragments not mapped by this analysis were Bam/J and Cla/I, for which no hybridisation data were obtained. The latter almost certainly maps between Cla/G and H since there is a gap at this locus of the ClaI map when fragment sizes are considered. This was subsequently confirmed during an analysis of cross-hybridisation between the MCV I and MCV III genomes (see section 5.8).

The restriction maps for MCV III DNA derived from the above analysis are presented in figure 11.

(Next page)

Figure 11. Restriction maps of MCV III DNA (isolate 8) for the enzymes BamHI, ClaI and HindIII. Fragments are labelled as in figure 10 and table 9. Refer to figure 7 legend for details.


BamHI
D
C
F I
E
A
G
H
B
E
HindIII
O/K
M
I P
G
E
B
A
F
I JL
D
G
C
FHN
QR
K/O
M
ClaI
A
J
G
I
H
C
E/F
D
B
F/E
0
50
100
150
200
kilobase pairs

5.7 Confirmation of MCV genome size by field inversion gel electrophoresis

The sizes of the genomes of MCV subtypes I, II and III were determined above by summation of restriction fragment sizes and compilation of restriction maps to be 190, 200 and 196 kb respectively. Independent confirmation of these figures for MCV I and MCV II was obtained by analysis of intact viral DNA by field inversion gel electrophoresis, a recent technique that extends the range of DNA sizes that can be separated to a limit of several megabases (Carle et al., 1986). The system used was home-made and consisted of a switching device wired to effect inversion of the field applied to a normal electrophoresis tank, with timing of forward and reverse periods controlled by a program running on a BBC microcomputer (Porter et al., 1988). MCV DNA was loaded into the gel and electrophoresed under a periodic field for 16 h, using switching conditions previously determined to provide enhanced separation in the range 50-300 kb (section 3.7). A ladder of bacteriophage lambda DNA oligomers provided size markers. Two different isolates of MCV subtypes I and II were used; see figure 12. A calibration graph of mobility against $\log_{10}$(size in kb) was constructed as for standard electrophoresis, the lambda oligomers giving size markers at 48.5 kb intervals. The sizes of the genomes of MCV I and MCV II determined from this experiment were 185+/-5 kb and 200+/-5 kb respectively. These figures are in close agreement with those obtained above and thus provide independent confirmation of those data.

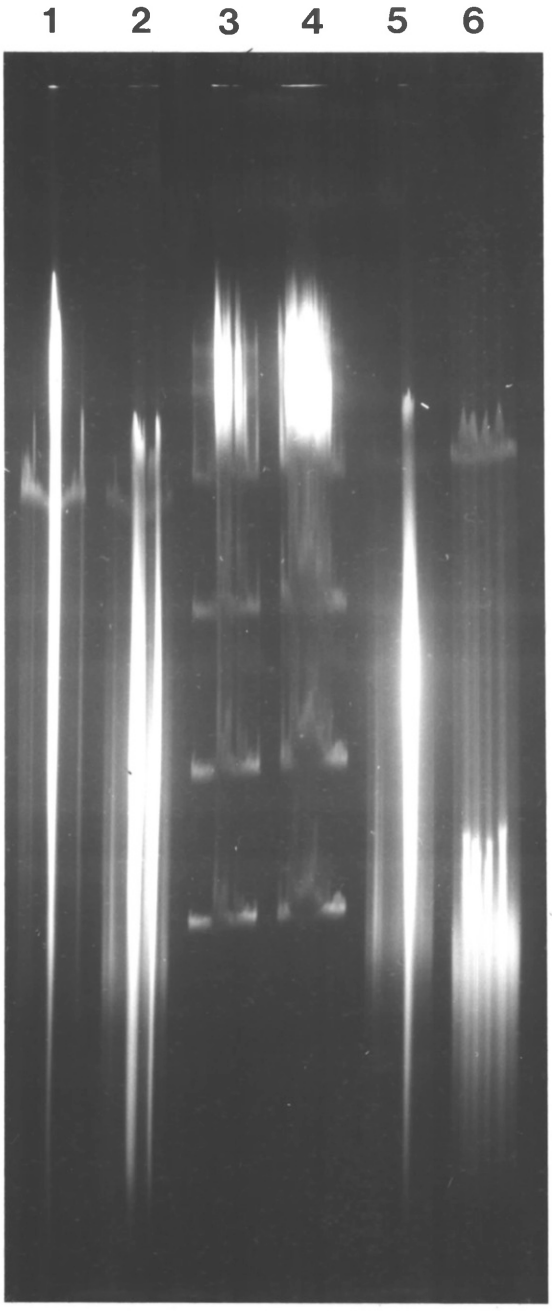


Figure 12. Field inversion gel electrophoresis of intact MCV DNA. Viral DNA was electrophoresed in an inverting field using a switching regime designed to optimise separation in the range 50-300 kb. Tracks 3 and 4 contain 1 ug and 0.5 ug loadings of molecular weight markers: oligomers of bacteriophage lambda DNA at 48.5 kb intervals. Markers are visible up to the sixth oligomer at 291 kb. Two different isolates of MCV I DNA (isolates 14 and 36) are loaded in tracks 1 and 2; similarly, MCV II DNA (isolates 26 and 3) is loaded in tracks 5 and 6. Although the samples are very smeary a band is clearly visible just below (MCV I) and slightly above (MCV II) the fourth lambda oligomer at 194 kb. This experiment provides a direct determination of the size of the MCV genome.

5.8 Comparison of the genome organisation of MCV I, II and III by cross-hybridisation

The relationship between the three MCV subtypes was examined by pairwise cross-hybridisation of their genomes. Initially, probes generated from the complete genome of one subtype were hybridised to Southern blots of BamHI, ClaI and HindIII digests of DNA from a second subtype. Probes were made by digesting MCV DNA with BamHI, heat denaturing and labelling by random oligonucleotide primer extension. These experiments revealed a close relationship between all MCV subtypes since the resulting hybrids were stable when filters were washed stringently in 0.1x SSC, 0.1% SDS at 65 ^{o}C. Such washing conditions select for >95% homology. All fragments of MCV II and MCV III DNAs cross-hybridised to MCV I DNA; all fragments of MCV I DNA cross-hybridised to MCV II and MCV III DNAs, with the exception of MCV I BamHI fragments K and O. These latter fragments map adjacently at approximately 40 kb from the right terminus of the MCV I genome and did not hybridise to MCV II or to MCV III DNA probes. This may be due to the presence within the MCV I genome of sequences having no counterpart in that of the other subtypes (ie. a relative "insertion") or due to sequence divergence between subtypes in this region of the genome to the extent that hybrids are not stable under stringent conditions.

MCV subtype genome organisations were compared in more detail in the same way as for the construction of the restriction maps of individual subtypes, as described above

(section 5.2): isolated DNA fragments from one subtype were used as probes to hybridise to Southern blots of digests of DNA from a second subtype. Control blots of digests of the same subtype and with the same enzyme as those used to generate fragments for labelling were used to confirm probe identities. Cross-hybridisation experiments performed were:-

(1) MCV I BamHI fragments A, B, E-L and O used as probes to Southern blots of MCV II DNA digested with BamHI, ClaI and HindIII.

(2) MCV II HindIII fragments A-L used as probes to Southern blots of MCV I DNA digested with BamHI and ClaI.

(3) MCV I BamHI fragments A-N used as probes to Southern blots of MCV III DNA digested with BamHI, ClaI and HindIII.

(4) MCV III ClaI fragments A-H used as probes to Southern blots of MCV I DNA digested with BamHI, ClaI and HindIII.

(5) MCV III ClaI fragments A-D, G and H used as probes to Southern blots of MCV II DNA digested with BamHI, ClaI and HindIII.

The MCV isolates used for these studies were those used above for restriction mapping, with the exception that the MCV I probes used in (1) were derived from DNA of isolate 41: this was necessary due to the lack of sufficient DNA from isolate 1 to perform these experiments. The restriction fragment

profiles of these two isolates of MCV I were identical for the enzymes BamHI, ClaI and HindIII and so direct comparison with the restriction maps obtained for isolate 1 is valid.

Hybridisation data obtained from these experiments are given in appendix 3D. They indicate that the genomes of all three subtypes of MCV are essentially colinear (ie. geographically equivalent regions of each genome cross-hybridised): the hybridisations observed can be accounted for by aligning all three maps as derived above (see figure 13). The orientations of the restriction maps for each subtype were chosen arbitrarily during data interpretation but they have been presented in the same orientation throughout the forgoing sections to facilitate comparison here. That cross-hybridisation occurred does not mean that the level of homology between isolates is uniformly high since the experiments described here provide only limited resolution. Hence, with the exception already mentioned and discussed in detail below, it can only be stated that there is no evidence that contradicts colinearity.

Alignment of the restriction maps accounts for the hybridisations observed for the regions of the genomes corresponding to positions 18-135 kb and 155-180 kb (relative to the map for MCV I). However the region 135-155 kb between these blocks is divergent and of variable length. This is the region of MCV I that includes Bam/K and O which did not cross-hybridise with other subtypes. Additionally in this region of the genome, MCV III Cla/D did not cross-hybridise to

MCV I Hind/E and neither MCV II Hind/G nor MCV III Cla/D cross-hybridised to MCV I Cla/B, as would be expected on the basis of alignment. Thus the region of MCV I between the ClaI site to the right of Cla/C and the HindIII site to the left of Hind/H, a stretch of approximately 12 kb, is specific to this subtype. Although there is no such contradiction of colinearity between subtypes II and III in this region on the basis of lack of cross-hybridisation, fragment sizes do not agree with colinearity: to allow alignment of the region corresponding to MCV I position 155-180 kb a relative insertion/deletion must be made. In fact, the distance between these two blocks of alignment is different for all three subtypes (see figure 13).

The right terminal regions of the MCV subtypes (from position 180 kb of MCV I) are also colinear, with that of MCV II extending the furthest to the right. The right terminal HindIII fragments of MCV I and MCV III DNAs align with the HindIII fragment adjacent to the terminal one for MCV II: this fragment is known to contain part of the terminal inverted repeat and some of the tandem repetitions. This situation is in agreement with the observed larger terminal repeat of MCV II relative to the other subtypes. The situation at the left terminus would be expected to be similar. However, alignment does not allow for protrusion of the MCV II terminus but places one of the blocks of tandem repetitions of MCV II with a part of the MCV I terminal repeat/unique sequence junction region which does not contain such repetitions (Hind/K). It is therefore possible that

there is a short (2-3 kb) relative deletion in the genome of MCV II in the region near the left terminus (ie. between the end of the terminal repeat and the first BamHI site): such a deletion would not be evident from the hybridisation experiments described here but is necessary in order to align the terminal repeats of both subtypes and to maintain symmetry at both termini. Clones covering the region of both MCV I and MCV II DNAs between 8-18 kb on the MCV I map (between the HindIII and BamHI sites) were obtained as part of another investigation (clones pMBH-I and pMBH-II respectively; see section 6.4). Their restriction maps are identical for 5 kb from the HindIII site but differ over the remainder. However, nucleotide sequence data obtained from the latter half of the clones showed that the two subtypes are highly related in this region of their genomes despite the lack of conserved restriction sites (see section 6.9). Thus, alignment of the MCV I and MCV II genomes continues as far left as the HindIII site to the left of Hind/D, 11 kb from the left terminus of MCV II. If a deletion as postulated above does exist it must occur on the terminal side of this HindIII site, ie. between 7.2 kb (the minimum size of the terminal repeat) and 11 kb from the terminus.

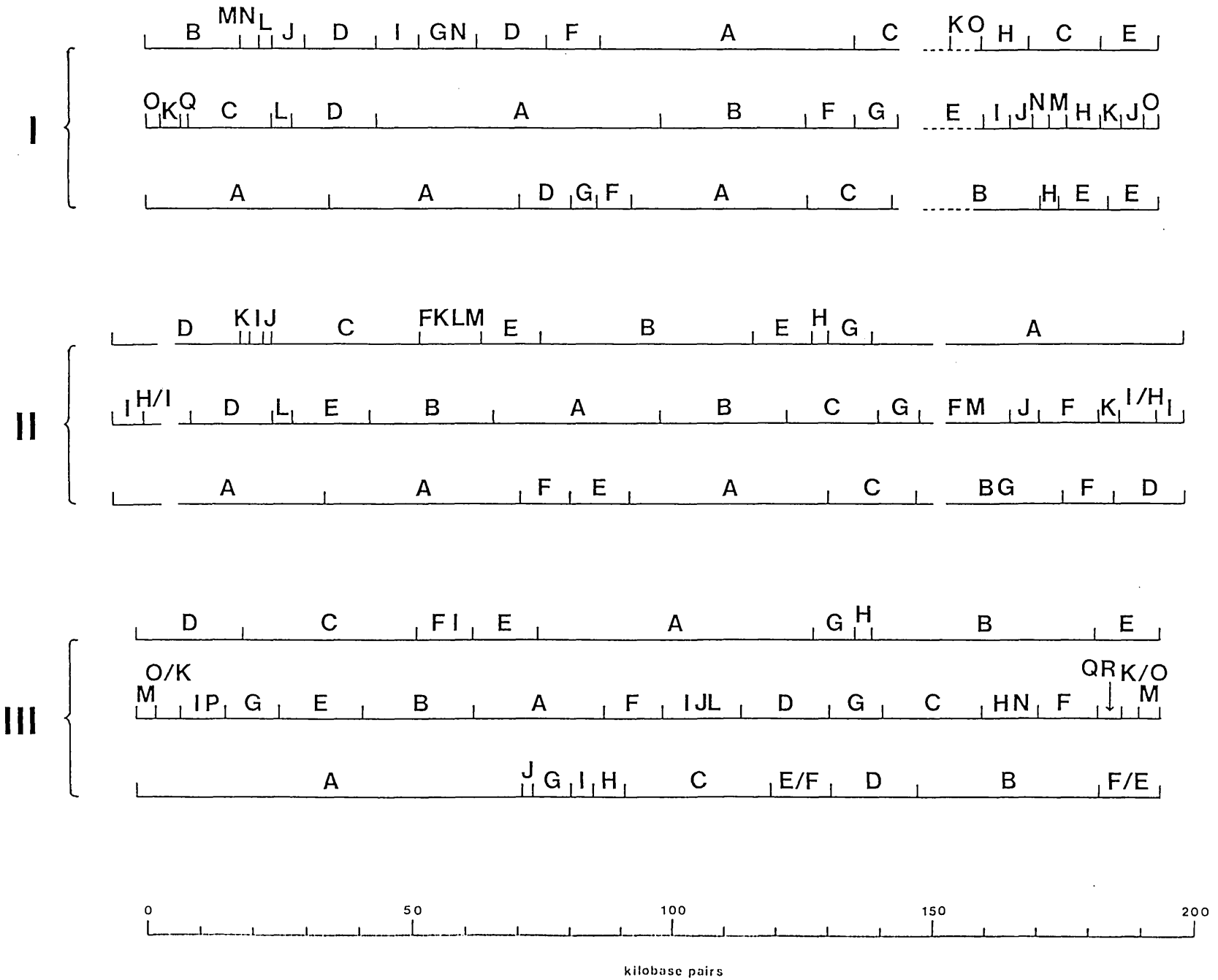
In a number of cases in the individual subtype restriction maps it was not possible to determine the relative orientation of two or more fragments mapping adjacently. The observation of colinearity of the three subtypes made it possible to resolve some of these uncertainties. Thus, the observation that MCV I Bam/L hybridised to MCV II Bam/I and

that MCV I Bam/J hybridised to MCV II Bam/J and C determined the order Bam/K, I, J (left to right) in the MCV II map. Hybridisation of MCV I Bam/G to MCV II Bam/F and M placed these fragments adjacent but it was still not possible to order all of Bam/F, K, L and M in the MCV II map. Hybridisation of MCV I Bam/H to MCV III Hind/H and N partially resolved the order of Hind/F, H and N, and the hybridisation of MCV II Hind/J to MCV I Bam/C and Cla/B placed Hind/J to the left of Hind/F in the MCV II map. These new data are incorporated into figure 13.

Although the genomes of the three MCV subtypes are essentially colinear there is little conservation of restriction sites between the maps as aligned in figure 13 on the basis of cross-hybridisation. This contrasts with the situation for the orthopoxviruses whose genomes show a central region of conserved restriction sites (Mackett and Archard, 1979). If there were any analagous restriction site conservation for the MCV subtypes then they would share some complement of common restriction fragments. However, when digests of DNA of all three subtypes were electrophoresed in adjacent tracks there were no such co-migrating fragments (see figure 4). The degree of conservation of restriction sites between two closely related linear DNA molecules can be used to estimate their level of sequence homology (Schumperli et al., 1977). However, the situation for the MCV subtypes does not fit this model since very few sites are conserved, despite the high level of sequence homology implied by formation of stable hybrids under stringent washing conditions.

(Next page)

Figure 13. Comparison of the restriction maps of the genomes of MCV I, MCV II and MCV III for the enzymes BamHI, ClaI and HindIII. The individual subtype maps are redrawn from figures 7, 9 and 11. Their relative orientations and alignments were determined by cross-hybridisation. All three genomes are colinear and cross-hybridise along their lengths with the exception of the region surrounding BamHI fragments K and O in MCV I. Discontinuities have been drawn in the MCV I and MCV II maps in this region to allow for alignment on either side. The portion of the genome unique to MCV I is marked by the broken line. An additional discontinuity in the MCV II map is drawn near the left terminus to take account of the alignment of the terminal inverted repeat, which is longer in MCV II.


I
B MN L J D I GN D F A C K O H C E
O K Q C L D A B F G E I J N M H K J O
A A D G F A C B H E E
II
D K I J C FKLM E B E H G A
I H/I D L E B A B C G FM J F K I/H I
A A F E A C BG F D
III
D C F I E A G H B E
O/K
M IP G E B A F IJL D G C HN F QR K/O M
A J G I H C E/F D B F/E
0 50 100 150 200
kilobase pairs

5.9 Variations in the restriction profiles of different isolates of MCV DNA

As an example of the variations observed in the restriction profiles of independent isolates of MCV DNA, six isolates of MCV II are compared in figure 14 for the enzymes BamHI, ClaI and HindIII. These serve to illustrate the variations seen for MCV subtypes I and II. Variations were of two types: those associated with variable size of the terminal repeat and those due to the loss or gain of individual restriction sites.

Variation in terminal repeat size is a common feature of the genomes of poxviruses due to the presence of variable numbers of short tandem repetitions (Wittek and Moss, 1980) and has been discussed extensively above (sections 5.1 and 5.5). The terminal HindIII fragment of MCV I DNA (Hind/O) contains these repetitions and varied in size between 2.2 kb and 3.4 kb in the isolates studied. For MCV II the repetitions are distributed between the terminal HindIII fragment (Hind/I) and that adjacent to it (Hind/H or I'; see section 5.5), with both fragments showing variability. MCV II Hind/I varied between 5.2 kb and 5.7 kb, and Hind/H varied between 6.2 kb and 7.0 kb: the total variation in the sum of Hind/I and Hind/H varied between 11.5 and 12.6 kb, a similar range to that observed for MCV I.

The second type of variation is a reflection of genetic drift between isolates in which a base change destroys or

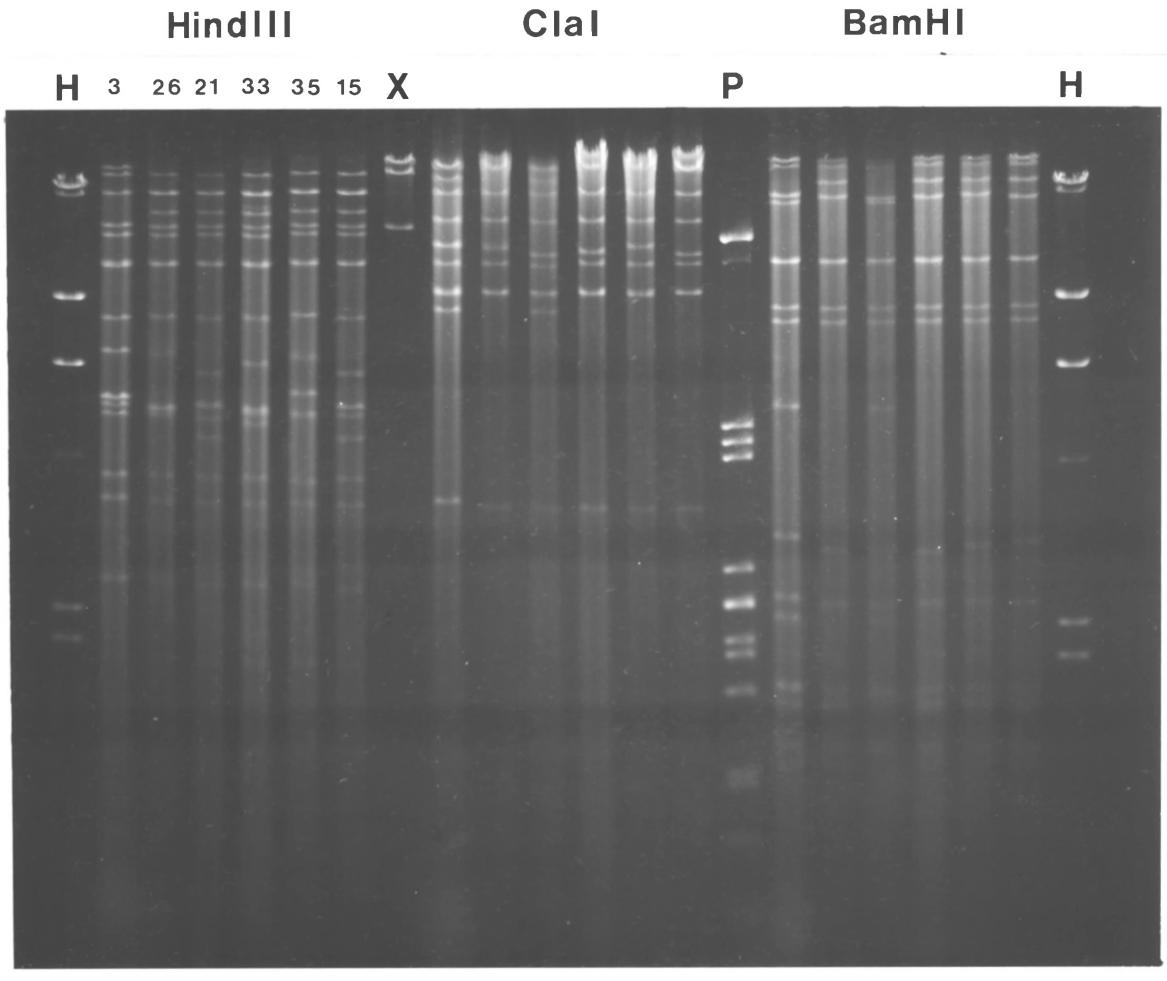


Figure 14. Variations in the restriction profiles of six isolates of MCV II DNA. Digests of isolates 3, 26, 21, 33, 35 and 15 with the enzymes BamHI, ClaI and HindIII are shown. The figure illustrates the types of variation discussed in section 5.9. Isolate 3 lacks a HindIII site relative to the other isolates, as demonstrated by the loss of one component of the 2-molar HindIII fragment B and of fragment C, and the presence of a new fragment above A (fragment identification relates to isolate 26, illustrated here and in figure 8). Isolates 3 and 21 possess a novel BamHI site within BamHI fragment C, as demonstrated by the loss of this fragment and the presence of two new fragments, one just above D and one below G. Size variations associated with the terminal repeat are also evident, particularly in the HindIII digests: there are three variable fragments (corresponding to fragments H, I and I' of isolate 26 as discussed in the text). Variations in the size of ClaI fragment D, parallelling those in the sum of the sizes of the terminal and near-terminal HindIII fragments, can also be seen. The molecular weight markers are digests of bacteriophage lambda DNA with HindIII (H), PstI (P) and XhoI (X).

creates the recognition sequence for a given enzyme. A number of instances of the former have been referred to in the above derivation of the restriction maps. Some examples of the latter were also observed: for instance, an extra BamHI site was present in MCV II Bam/C in three isolates, generating two new fragments in its place (see figure 14).

One variation was particularly frequent and deserves special mention, namely that the BamHI site of MCV I that maps between Bam/D and F (see figure 7) was absent in a large proportion of isolates of this subtype (30 of a total of 92: 33%). This was evident by the absence of restriction fragments Bam/D (one component of a 2-molar band) and Bam/F and the presence of a new fragment whose size was that of the sum of these two fragments. That this variation was so common, whilst all others of its kind were observed only rarely, is indicative that either this region of the MCV I genome is highly variable (ie. is a mutagenic "hotspot") or that there are two distinct "classes" of this MCV subtype, if the presence or absence of the BamHI site reflects any functional difference. In the list of MCV isolates given in appendix 2 those whose DNA lacked this BamHI site are indicated. There is no correlation with other data listed there. In order to examine further the nature of this variation, the Cla/D fragment spanning this BamHI site was cloned from two isolates of MCV I, one of which lacked this site (clones pC/D+ and pC/D-, with and without the BamHI site respectively). The Cla/D fragments were isolated from low melting-point agarose and ligated with T4 DNA ligase to pBR322

digested with ClaI and phosphatased (section 3.13). The ligation reactions were used to transform E. coli strain DH1 to ampicillin resistance (section 3.14). Plasmid DNA was prepared from the transformants by alkaline lysis (section 3.15.1) and the clones were compared by digestion with a number of different enzymes in order to assess whether differences other than that involving the BamHI site existed. The only difference detected between them was in the occurrence of XhoI sites, as deduced from differences in the restriction profile with this enzyme. However, the BamHI site of pC/D+ occurred within a XhoI fragment common to both clones. Thus there does not appear to be any gross difference at this genomic locus between isolates possessing or lacking this highly variable BamHI site; the possibility remains that the extent of the variation is too limited to have been detected by this means. The mechanism behind such variation is also a matter of speculation.

5.10 Summary of chapter

The work described in this chapter has established the structural organisation of the MCV genome, allowing comparison with other characterised poxviruses and also laying the foundation for further studies of the molecular biology of the virus discussed in subsequent chapters.

There are three distinct subtypes of MCV. Their genomes have different restriction maps but are closely related at the nucleotide level, indicated by their stable cross-hybridisation under stringent washing conditions. MCV DNA is 190-200 kb long: the isolates studied here have genome sizes of 190 kb, 200 kb and 196 kb for subtypes I, II and III respectively. The genomic termini are covalently closed, forming a hairpin. There are terminal inverted repeats containing variable numbers of tandem repetitions, which are organised into two blocks in the case of MCV II. The repeat sizes of MCV I, II and III are 3.9-5.4 kb, 7.2-10.6 kb and 3.7-7.0 kb respectively for the isolates studied. In all these respects the physical characteristics of MCV DNA are similar to those of DNA of the orthopoxviruses. However, MCV DNA has a relatively high G+C content (an estimated 58%) compared to vaccinia virus DNA (36%).

DNA of each of the three MCV subtypes has been mapped with the enzymes BamHI, ClaI and HindIII. Comparison of restriction maps reveals no region of conservation, in contrast to the situation for the orthopoxviruses. However,

the genomes are colinear in terms of their content since geographically equivalent regions cross-hybridised, with the exception of a region unique to MCV I. Two discontinuities (ie. relative insertions or deletions) are postulated in order to maintain the observed alignment.

CHAPTER 6

Use of synthetic oligonucleotide probes to detect MCV DNA sequences capable of encoding a conserved domain of EGF-like growth factors

In order to determine whether Molluscum contagiosum virus (MCV) DNA encodes a growth factor belonging to the family of polypeptides related to epidermal growth factor (EGF), synthetic oligodeoxyribonucleotides (oligonucleotides), designed on the basis of conserved domains of this family, were used in hybridisations to Southern blots of MCV DNA. It was previously established that cloned DNA probes of the human EGF cDNA and of the vaccinia virus growth factor (VGF) gene did not cross-hybridise to MCV DNA. This chapter describes the use of a 14-mer and two 17-mer oligonucleotides, the cloning of hybridising loci of the MCV genome and sequence data derived from such clones.

6.1 Screening of MCV DNA with cloned human EGF and VGF probes

Clones of the cDNA for the human EGF gene in the plasmid pUC9 (phEGF) and for the VGF gene in the plasmid pBR328 (pVGF) were available. Plasmid DNA was recovered by alkaline lysis from small-scale cultures of E. coli containing these clones (section 3.15.1) and inserts were cleaved from

the vector with the appropriate restriction enzymes and separated by electrophoresis through low melting-point agarose. The insert fragments were isolated, labelled with [α^{32}P]dCTP by random oligonucleotide primer extension (section 3.10.1) and used as probes to Southern blots on nitrocellulose of digests of MCV DNA. Hybridisation was carried out for 16 h at 65 ^{o}C, after which the filters were washed to a final stringency of 2x SSC, 0.1% SDS. There was no hybridisation of either probe to MCV DNA under these low stringency conditions. Neither was there cross-hybridisation of MCV and vaccinia DNAs under these conditions. Similarly, the inserts of phEGF and pVGF did not cross-hybridise, as expected from the low level of nucleotide sequence homology between the EGF and VGF genes, despite the functionally significant homology of their products (see sections 1.3.3 and 1.3.4).

6.2 Conserved features of the EGF-family of growth factors: design of oligonucleotide probes 1-3

The family of EGF-like growth factors has been discussed in section 1.3. Those members of the family for which sequence data are currently available consist of EGF from four species, transforming growth factor-alpha (TGF) from two species, and three poxviral growth factors (see table 2). At the outset of this project, data for only a single representative of the poxviral growth factors (namely VGF) were available and so the experimental design of this and the following chapter was more restricted than that of chapter 8,

for which all nine sequences were available. The overall homology between EGF, TGF and VGF is low, whilst conservation is absolute in the relatively small regions which presumably dictate function (see sections 1.3.4 and 1.3.5). Nucleotide sequence homology between the genes for these growth factors is not significant overall. However, the short regions of strict amino-acid conservation provide information for the design of oligonucleotide probes able to detect DNA sequences capable of encoding such conserved domains.

The region of greatest homology between all members of the EGF family for which sequence data were available (seven members, consisting of EGF from four species, TGF from two species, and VGF) is that comprising the third disulphide-loop: over a stretch of seven amino-acids (EGF residues 36-42), five are absolutely conserved. The consensus sequence of this domain (domain A) is:-

G Y - G - R C

where the single letter amino-acid code is used (see legend to table 2) and a dash indicates a non-conserved position, one either side of the central amino-acid. There is little chemical similarity between the residues occurring at these non-conserved positions although within the group of four EGF sequences the second of them consists of either glutamate or aspartate, both of which are acidic. However, an acidic residue is clearly not requisite since both TGF sequences and the VGF sequence encode hydrophobic residues at this position

(alanine, valine and isoleucine). There is some preference for hydrophobic residues at the first of the non-conserved positions: isoleucine and valine are common, although serine and threonine occur. The position immediately following domain A is either glutamate or glutamine in every case, and histidine is common at the next position. Amongst the data relating to structure-function studies of EGF that were available at the time, this conserved domain A was implicated as being that responsible for binding to the EGF receptor. More recent data show that such a view is over-simplified although residues within this domain do have a role in receptor binding (see section 1.3.5).

A second region of more restricted homology exists within the first disulphide-loop. This consists of a penta-peptide (EGF residues 11-15) that is absolutely conserved between all four EGF sequences and VGF but is not conserved in either TGF sequence. Thus, the functional significance of this domain was not clear. The sequence of this domain (domain B) is:-

D G Y C L

The position immediately following domain B is either histidine or asparagine.

Oligonucleotides were designed to hybridise to DNA sequences encoding these domains. Domain A was expected to be conserved in a putative MCV-encoded EGF-like growth factor.

Whilst domain B is relatively short and may or may not be similarly conserved, it offered an advantage over domain A for oligonucleotide construction in that the conserved amino-acids are contiguous (ie. they are not interrupted by non-conserved residues).

For the five conserved amino-acids of domain A, there are 6 codons for one of them (arginine, R), 4 codons for two (both glycine, G) and 2 codons for two (tyrosine, Y; cysteine, C). Since the other two non-conserved positions could be encoded by one of 61 possible codons (64 total codons minus 3 stop codons) then the total number of different nucleotide sequences that could encode the septapeptide of domain A is 1,428,864. Each of these possibilities might reasonably be expected to encode a domain with the function sought: that is, there is no a priori reason to exclude any particular combinations of non-conserved amino-acids. A similar analysis of domain B shows that one amino-acid has 6 codons (leucine, L), one has 4 codons (glycine, G) and three have 2 codons (aspartate, D; tyrosine, Y; cysteine, C) giving a total of 192 coding possibilities for this pentapeptide.

Several steps were taken to reduce the number of different oligonucleotides that needed to be considered for domain A. Firstly, degenerate oligonucleotides consisting of pools of related oligonucleotides were used. Secondly, the amino-acids used by the vaccinia virus in VGF were used at the two non-conserved positions. The rationale behind this decision was that the amino-acids used by MCV were thought

likely to be related more closely to those used by another poxvirus than to those used in a mammalian system. In VGF, these amino-acids are threonine (T, 4 codons) and isoleucine (I, 3 codons), for which the total number of coding possibilities for domain A is reduced to 4,608. This was still considered to be too high a level of degeneracy for an oligonucleotide probe, since the oligonucleotide forming the best hybrid would represent only 0.02% of the total radioactive label incorporated into the probe and the likelihood of spurious matches of no relevance would be too great (section 3.9.6). A corollary of this decision is that the oligonucleotide forming the best hybrid to the putative MCV gene may not match the target sequence at every position: ie. the hybrid could contain mismatches and be less stable as a consequence.

A method often used to reduce the degeneracy of synthetic oligonucleotides is based on information regarding the relative usage of possible codons for a given amino-acid in the organism of study (ie. codon preference). However, this was not applicable to MCV since no sequence data were available for this virus. Also, it was noted that codon preference exhibited by vaccinia virus differed significantly from that shown by mammalian (or specifically, human) gene sequences and so there was no reason to believe that human codon preferences would be applicable in reducing degeneracy. That the base compositions of the vaccinia virus and MCV genomes were previously noted to be significantly different (section 5.1) was interpreted to indicate that vaccinia codon

preference data were unlikely to reflect the situation for MCV. The final opportunity of reducing degeneracy was the suggestion that, of the twelve possible mismatched pairs of bases, the formation of a mismatched G to T (the so called "wobble" pairing that occurs in codon-anticodon hybrid formation by mRNA and tRNA) causes the least helix destabilisation (Thein and Wallace, 1986). It was argued that, if the identity of the base at a given position of the target sequence was unknown, only the bases G and T needed to be considered in the oligonucleotide probe at this position, since then the hybrid would have a 50% chance of being correctly matched (if the target base was A or C) and a 50% chance of being a G-T "wobble" mismatch (if the target base was G or T): thus, the formation of strongly destabilising mismatches would be prevented. In this way, the degeneracy could be significantly reduced without a corresponding loss of scope in terms of the sequences to which the oligonucleotide probe would hybridise, and the oligonucleotide forming the best hybrid to the target sequence would represent a much greater proportion of the total radioactivity incorporated into the probe.

Oligonucleotide 2 was designed to be 17 bases long (ie. a 17-mer) and 32-fold degenerate, covering the amino-acids GYTGIR of domain A of VGF (of which T and I are the non-conserved positions) but avoiding the third postition of the highly degenerate arginine codon. Use of potential G-T "wobble" pairing was made at five positions (the maximum possible in this sequence). The same strategy was used to

design oligonucleotide 1, a 17-mer with 4-fold degeneracy covering the whole of domain B and the following amino-acid (commonly histidine, H): DGYCLH. Additionally for this domain, a 14-mer (oligonucleotide 3) was designed without taking any steps to reduce degeneracy: it was 64-fold degenerate and covered the whole of domain B. The oligonucleotide sequences in each case were designed to hybridise to the coding (ie. "sense") strand of DNA encoding these domains. The sequences of oligonucleotides 1-3 are given in table 4.

6.3 Synthesis and use of oligonucleotides 1-3

Oligonucleotides 1-3 were synthesised automatically using phosphoramidite chemistry, deprotected by treatment with acetic acid, and purified by ultraviolet shadowing and isolation from a polyacrylamide gel (section 3.9). They were 5' end-labelled with T4 polynucleotide kinase and [γ^{32}P]ATP and separated from unincorporated nucleotides by exclusion chromatography (section 3.10.2). Southern blots of digests of DNA from all three subtypes of MCV on nitrocellulose were prepared and pre-hybridised, using tRNA as the blocking agent (section 3.11.2). The blots also contained control digests of pVGF. Probes were added and hybridisation was continued overnight.

The theoretical melting temperatures for perfectly matched oligonucleotides were calculated using the Wallace

rule (section 3.11.2). This gave a range of possible values for the T_m of 32-56 ^oC, 36-58 ^oC and 36-46 ^oC for oligonucleotides 1, 2 and 3 respectively. For initial experiments, the temperature for pre-hybridisation and hybridisation was 30 ^oC in each case: ie. 2-6 ^oC below the lower limit for T_m. Following hybridisation, filters were washed in 6x SSC, 0.1% SDS for 2 h at 30 ^oC, sealed in cling-film (this was necessary to prevent the filters from drying out) and autoradiographed using pre-flashed X-ray film. Filters were subsequently re-washed in 6x SSC, 0.1% SDS at a higher temperature for 30 min and the process of autoradiography and re-washing was repeated until the best signal-to-noise ratio was obtained: this was at the temperature above which the signal also was lost. For oligonucleotide 2, these conditions were 36 ^oC for MCV I and MCV II DNAs, and 38 ^oC for MCV III DNA. The difference in the observed Tm with different MCV subtypes was interpreted to indicate that the target sequence showed minor variation. Oligonucleotides 1 and 3 did not give significant hybridisations to MCV DNA and their use was not pursued further.

The results obtained for oligonucleotide 2 are illustrated in figure 15. The strongly hybridising fragments of each MCV subtype were BamHI fragment B, ClaI fragment A and HindIII fragment C for MCV I; BamHI fragment D, ClaI fragment A and HindIII fragment D for MCV II; and BamHI fragment D, and HindIII fragment G for MCV III. These results were regarded as significant since these fragments come from equivalent regions of the viral genomes. Some fainter bands were present

on the autoradiographs and could not be removed by re-washing the filters without loss of those signals already mentioned: most obvious was that corresponding to the poorly resolved MCV I BamHI fragments C and D, both of which are 2-molar, which is assumed to have resulted from non-specific hybridisation to the large amount of viral DNA present at this position on the filters. The absence of hybridisation signals for co-mapping fragments in other digests confirmed that the latter was not significant. No hybridisation of oligonucleotide 2 to pVGF DNA was detected: the best match would involve three G-T mismatches, at positions 3, 6 and 15 from the 3'-end of the oligonucleotide.

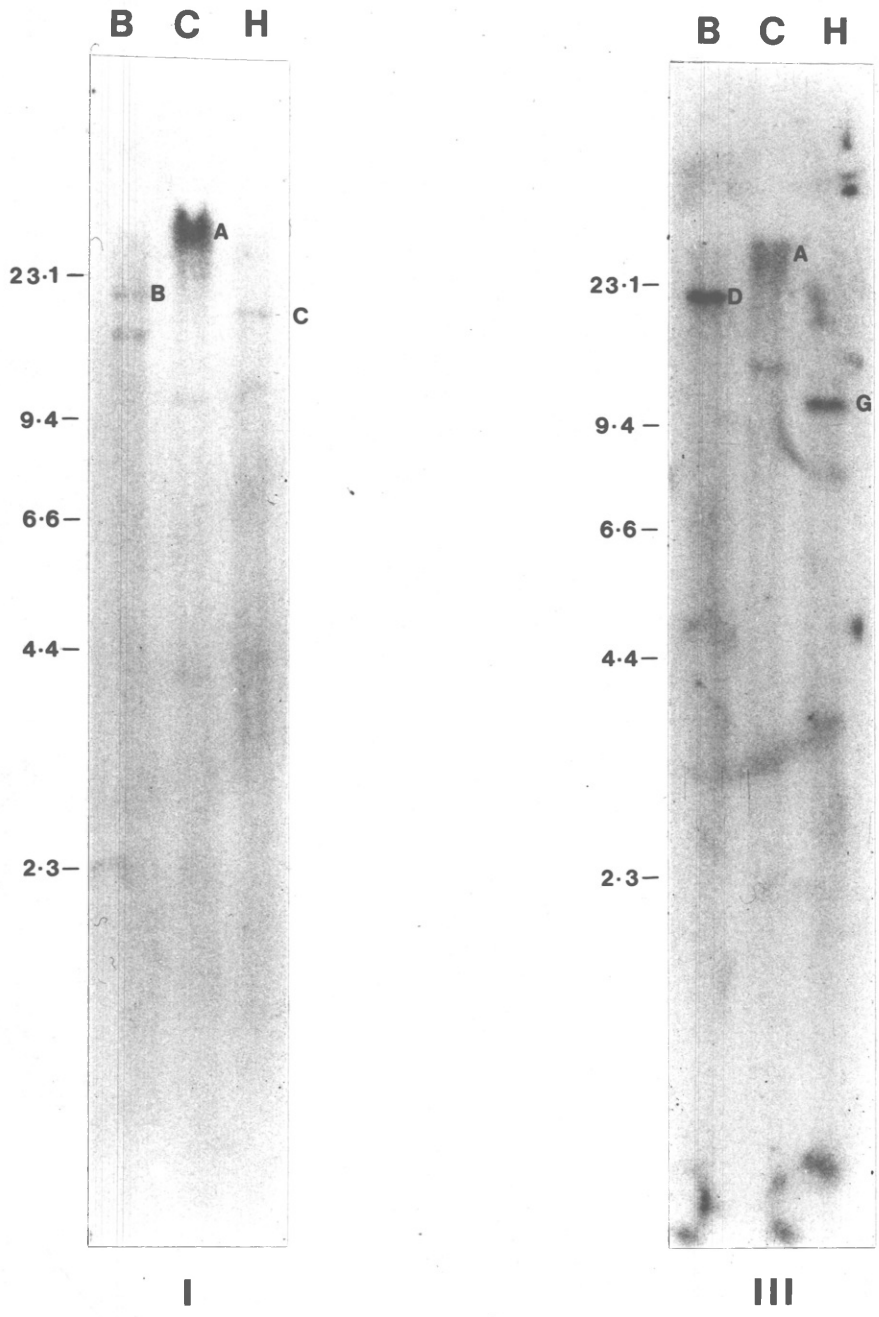


Figure 15. Hybridisation of oligonucleotide 2 to BamHI, ClaI and HindIII digests of MCV I and MCV III DNA. The hybridising fragments are identified by letter. Figures refer to bacteriophage lambda DNA HindIII fragment size markers.

For each of the three MCV subtypes, the fragments from BamHI, ClaI and HindIII digests to which the oligonucleotide probe hybridised map together, delineating the region of the viral genome containing the target sequence. Furthermore, based on the alignment of the viral genomes established earlier (see section 5.8 and figure 13), this hybridising locus is equivalent in all three subtypes (figure 16). The position of the target sequence is conserved and is within the region of the genome corresponding to 14-18 kb from the left terminus of MCV I; that is, at least 8.6 kb from the end of the terminal repeat of 3.9-5.4 kb in this subtype (section 5.1). This compares with the location of the VGF gene within the terminal repeat of the vaccinia virus WR genome (Venkatesan et al., 1982), and of its homologues in wild-type orthopoxviruses within unique sequences close to the junction with the left terminal repeat (Archard et al., 1986).

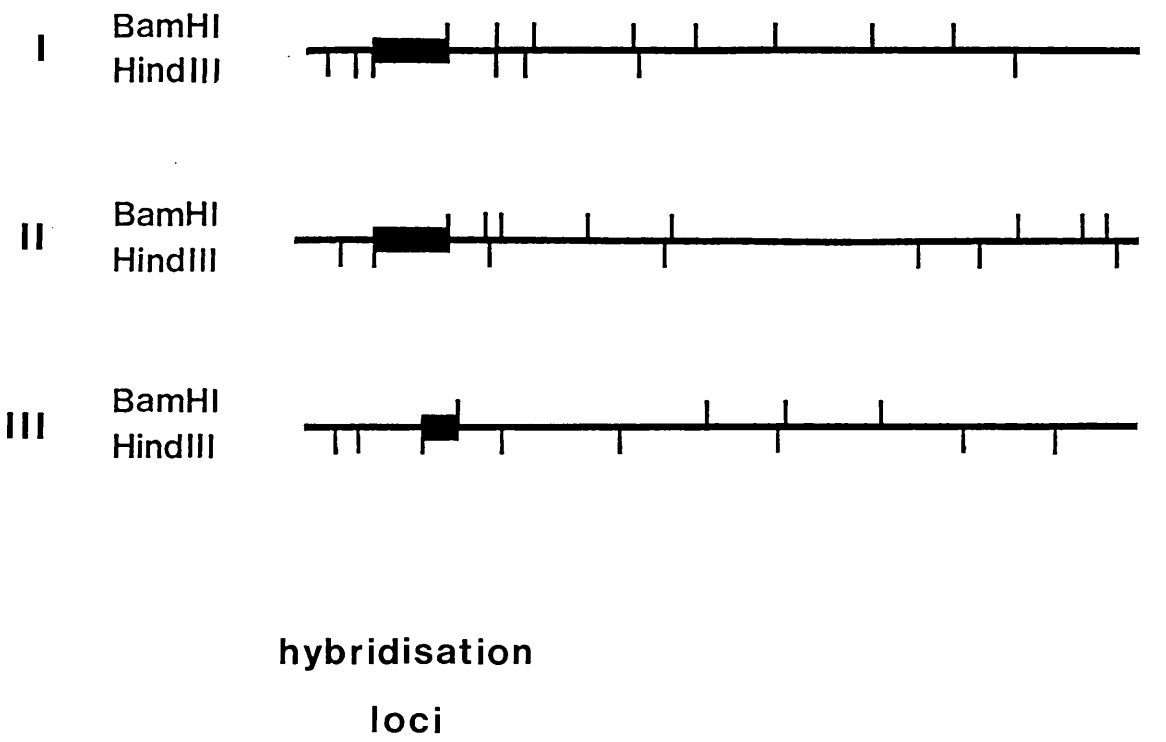


Figure 16. The locus of oligonucleotide 2 hybridisation to MCV subtypes I, II and III. The BamHI and HindIII genomic restriction maps for the left half of the genome are presented and the hybridisation locus is indicated for each subtype. The loci are geographically equivalent in all three subtypes: refer to figure 13.

6.4 Cloning the oligonucleotide-positive locus

There was insufficient DNA of the only isolate of MCV III to allow the region to which the oligonucleotide 2 probe hybridised to be cloned from this subtype. This region was however cloned from an isolate of MCV I DNA (isolate 17) and an isolate of MCV II DNA (isolate 15). In the case of the latter MCV subtypes the oligonucleotide-positive locus was bounded by the BamHI site nearest the left terminus of the genome and a HindIII site approximately 10 kb nearer the terminus. It was not possible to directly clone either of the relevant BamHI or HindIII fragments into a plasmid vector due to the possession of a covalently-closed end by the former and due to the large size of the latter (16.5 and 15.5 kb for MCV I and MCV II respectively). Instead, the overlap of these fragments, a 10 kb BamHI/HindIII fragment, was cloned. The plasmid vector chosen was pBR322.

Approximately 10 ug of MCV DNA (75 fmol) was digested with BamHI and electrophoresed through 0.7% low melting-point agarose. The gel slices containing the BamHI fragment B (MCV I) and D (MCV II) were cut from the gel without ethidium bromide staining or exposure to ultraviolet light, using as a guide a small amount of the digest that was visualised (section 3.13.1). The DNA was recovered by melting the gel slice, extracting several times with phenol and then with chloroform, and precipitating with ethanol. Plasmid pBR322 DNA was prepared by digestion with BamHI and HindIII, and the 4,016 bp vector fragment was separated from the 346 bp

fragment by electrophoresis through low melting-point agarose and recovered as were the MCV fragments. The MCV BamHI fragment was ligated to the BamHI- and HindIII-cut pBR322 vector with T4 DNA ligase for 16 h at 16 ^{o}C in an 8 ul reaction volume. The estimated DNA concentrations were 1 nM insert DNA and 3 nM vector DNA. Following ligation, the reaction volume was increased to 30 ul by the addition of TE and extracted with phenol:chloroform:amyl alcohol (25:24:1) and then ether in order to remove the enzyme. After addition of medium salt restriction enzyme buffer, the reaction products were digested with HindIII in a 40 ul reaction volume which was extracted as before. The desired result of these steps was a linear construct consisting of the 10 kb BamHI/HindIII MCV fragment ligated to pBR322 via the BamHI end. The final stage was to circularise this product and ligate its two HindIII ends. This was achieved by adding ligation buffer and T4 DNA ligase and incubating for 16 h at 16 ^{o}C in a total volume of 50 ul. A relatively large volume was used in order to favour intra- over inter-molecular ligation. Circularisation of remaining vector DNA was not possible due to its possession of non-compatible ends. The cloning scheme is illustrated in figure 17.

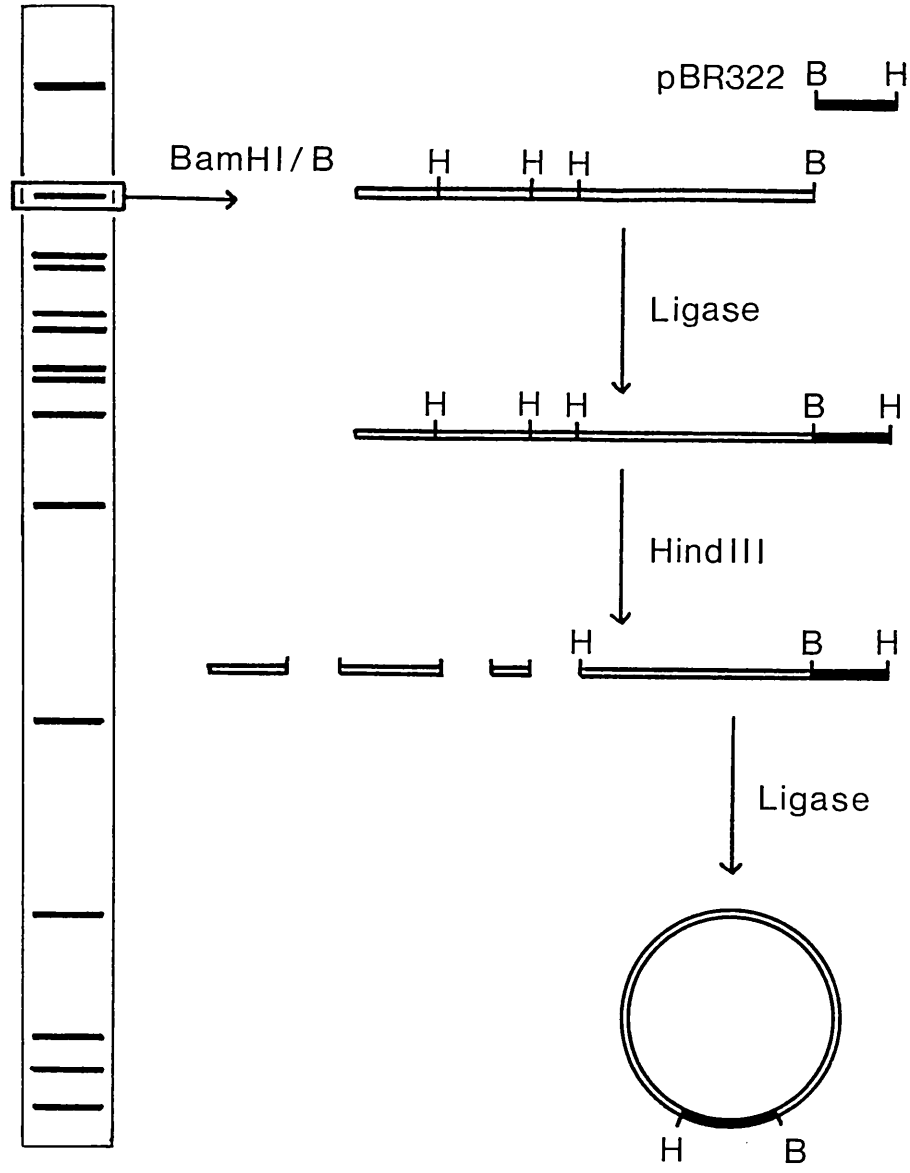


Figure 17. The scheme for cloning the overlap of MCV I BamHI fragment B and HindIII fragment C containing the locus of oligonucleotide hybridisation. The BamHI fragment B was isolated from low melting-point agarose after electrophoresis: it possesses one BamHI end and one covalently closed end, the viral genomic terminus, and there are three internal HindIII sites. Following ligation to the BamHI end of pBR322, the DNA was digested with HindIII and the construct was circularised by self-ligation.

E. coli strain DH1 was transformed with the products of the above procedure and plated on L. agar with ampicillin selection (section 3.14.3). Individual colonies were inoculated into L. broth, grown overnight in the presence of ampicillin and used for small-scale preparation of plasmid DNA by alkaline lysis (section 3.15.1). Plasmid DNA was digested with HindIII and the products were analysed by gel electrophoresis to determine the nature of the resultant clones. One clone from each of MCV I and II that yielded a single fragment of the expected size upon HindIII digestion was also analysed by digestion with BamHI and double-digestion with BamHI and HindIII to confirm that it was the required product. Further confirmation was sought by hybridisation of the clone inserts (labelled in low melting-point agarose with [α^{32}P]dCTP by random oligonucleotide primer extension) to BamHI and HindIII digests, and BamHI/HindIII double-digests of MCV DNA, and additionally, by hybridisation of oligonucleotide 2 to BamHI/HindIII double-digests of the clone DNA. Both experiments confirmed that the clones obtained were of the desired construction. The clones were named pMBH-I and pMBH-II for MCV I and MCV II respectively.

E. coli containing clones pMBH-I and pMBH-II were inoculated into 1 litre of L. broth and grown in the presence of ampicillin. Plasmid DNA was made by alkaline lysis and caesium chloride density centrifugation (section 3.15.2). The restriction sites for the enzymes EcoRI, PstI, SalI and SmaI within this cloned MCV DNA were mapped by analysis of fragments resulting from single- and double-digests using

BamHI and HindIII in addition to these enzymes (figure 18). The insert sizes of pMBH-I and pMBH-II were determined to be 9.7 and 9.5 kb respectively. The SmaI and PstI fragments contained within pMBH-I are consistent in size with those fragments previously mapped to the position of the MCV I genome from which the clone was derived (namely, PstI fragment I and SmaI fragment L; section 5.3).

The clone locus to which the oligonucleotide hybridised was mapped by probing Southern blots of the digests used for mapping the clones. This experiment indicated that the target sequence is within a 3.1 kb region of pMBH-I, bounded by a SmaI and a PstI site, and within a 1.0 kb region of pMBH-II, bounded by a SmaI and a SalI site (see figure 18). The relative orientation of the inserts of both clones is the same, as dictated by the cloning protocol, allowing comparison of their maps. The restriction sites mapped within pMBH-I and pMBH-II are identical from the HindIII site as far as and including the SmaI site almost 5 kb away, from which point they differ. The region of identity corresponds to that nearest the terminus of the MCV genome. The oligonucleotide loci are within the non-identical regions of the clones: however, they are geographically equivalent as judged by the alignment of the two restriction maps.

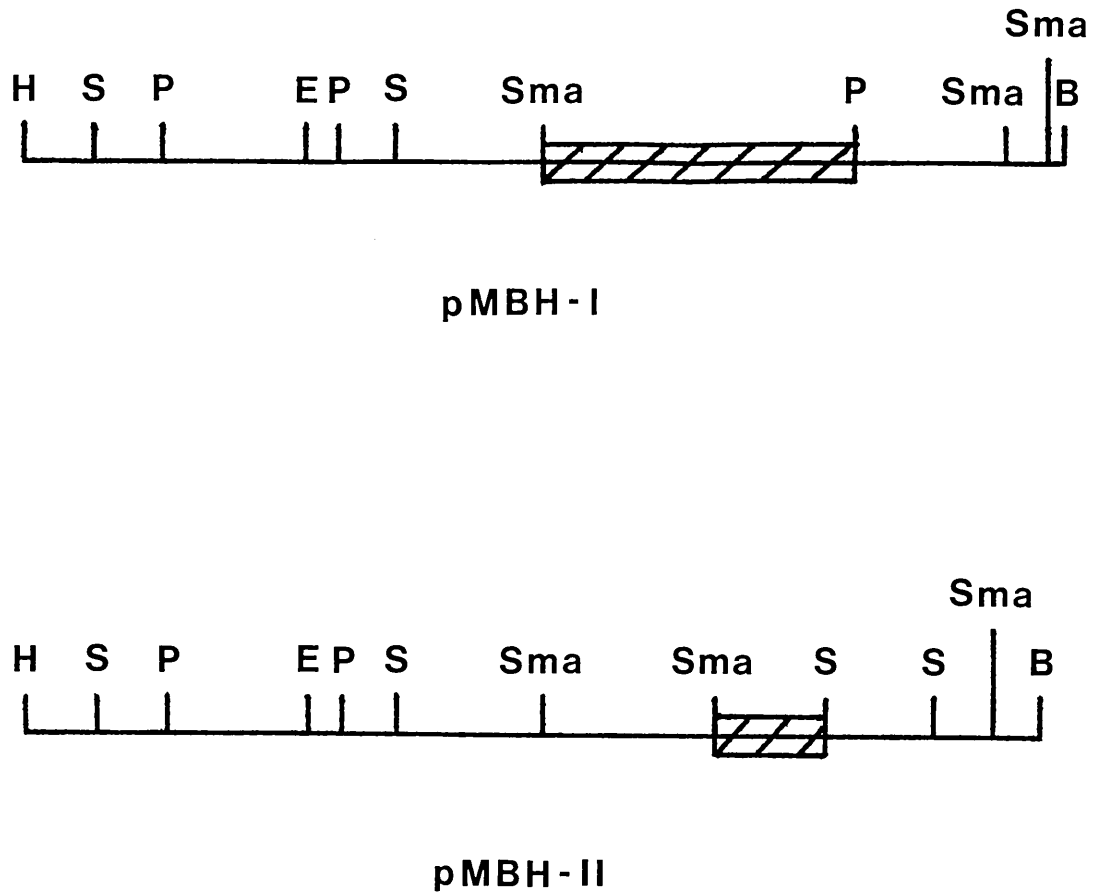


Figure 18. Restriction maps of clones pMBH-I and pMBH-II. The enzymes used are BamHI (B), EcoRI (E), HindIII (H), PstI (P), SalI (S) and SmaI (Sma). The loci of oligonucleotide 2 hybridisation are indicated. The SmaI/PstI fragment of pMBH-I and the SmaI fragment of pMBH-II containing these loci were subcloned into M13 for sequencing.

6.5 Subcloning from pMBH-I and pMBH-II into M13mp8

For sequencing the oligonucleotide-positive locus the Sanger dideoxynucleotide method of DNA sequencing from single-stranded M13 templates was chosen. The method described by Henikoff (1984) using unidirectional deletion with exonuclease III for the production of an ordered series of clones deleted to varying extents was chosen, to allow sequence determination from both DNA strands throughout the region of interest. These choices required that the region to be sequenced was subcloned into M13 in both orientations. Exonuclease III is a 3' to 5' exonuclease and digests double-stranded DNA from a single-stranded 5'-overhang but not from a 3'-overhang. The method of unidirectional deletion requires (i) that there is a restriction site between the insert and the sequencing primer binding site in the vector for an enzyme leaving a 3'-overhang, (ii) that there is a restriction site between the former restriction site and the insert for an enzyme leaving the more usual 5'-overhang, and (iii) that neither of these enzymes cuts within the insert or elsewhere within the vector. For M13mp8, the only choice of enzyme leaving a 3'-overhang is PstI. In the case of pMBH-II, it was decided to subclone the 2.5 kb SmaI fragment into the EcoRI site of M13mp8, using synthetic EcoRI linkers. This would allow the use of PstI and BamHI, both cutting within the multiple cloning site of M13, as the pair of restriction enzymes required for the deletion protocol. The situation was more complicated for pMBH-I due to PstI being one of the enzymes with a site bounding the oligonucleotide-positive locus. In this case it was decided

to subclone the 3.1 kb SmaI/PstI fragment into the EcoRI site of M13mp8 by first blunting the PstI end (and thus losing this site) and then using synthetic EcoRI linkers.

pMBH-I and pMBH-II DNA was digested with SmaI. For the former only, this digest was subsequently digested with PstI: each of dATP, dCTP, dGTP and dTTP (to a final concentration of 0.25 mM) and 14 units of E. coli DNA polymerase I "Klenow" fragment were added and incubation was continued for 80 min at 37 ^{o}C. This last step was used to blunt the 3'-overhang generated by PstI digestion by virtue of the 3'-5' exonuclease activity of the polymerase. Digests were electrophoresed through 0.7% low melting-point agarose and the 3.1 kb SmaI/PstI fragment of pMBH-I and the 2.5 kb SmaI fragment of pMBH-II were isolated (section 3.13.1). Synthetic EcoRI linkers were phophorylated using ATP and T4 polynucleotide kinase. Blunt-ended insert DNA and kinased linkers were ligated with T4 DNA ligase for 16 h at 12 ^{o}C, using estimated concentrations of 7 nM and 19 uM respectively (section 3.13.2). The ligase was inactivated by heating at 65 ^{o}C and the reaction was digested with EcoRI prior to passage through a Sephadex G-50 column to separate the insert DNA, now possessing EcoRI ends, from the unligated linker fragments by exclusion chromatography. The DNA was eluted and 100 ul fractions were collected, aliquots of which were analysed by agarose gel electrophoresis to determine which fractions contained the DNA. This occurred in two fractions, which were pooled for ethanol preciptation. M13mp8 RF DNA was digested with EcoRI and phosphatased with calf intestinal phosphatase

(section 3.13.3). Insert and vector DNAs were combined and ligated with T4 DNA ligase for 16 h at 12 ^{o}C, using estimated concentrations of 5 nM and 1.4 nM respectively (section 3.13.4).

The ligation reactions were incubated at 65 ^{o}C for 10 min and used to transform E. coli strain JM101, which was plated on a lawn of JM101 on H. agar with IPTG/XGal selection (section 3.14.3). Individual plaques were picked, grown in fresh JM101 and M13 DNA was prepared (section 3.16). RF DNA was analysed by EcoRI digestion and agarose gel electrophoresis to determine the size of the insert. Several verified subclones from pMBH-II were similarly analysed with SalI in order to determine the orientation of their inserts, since this enzyme cut asymmetrically within the insert as well as within the multiple cloning site of M13mp8. One clone in each orientation was chosen for further work: these were named mIIa and mIIb. The same analysis for pMBH-I subclones was not based on any mapped restriction sites but on PstI and XhoI double-digestion, the latter enzyme having been determined to cut asymmetrically within the insert although not within M13. The XhoI site was mapped onto the restriction maps previously determined for pMBH-I, allowing clones mIa and mIb to be named in the same relative orientations as mIIa and mIIb.

6.6 Generation of nested deletion clones; DNA sequencing

Large-scale preparations of double-stranded RF DNA of the M13 clones mIIa and mIIb were made by alkaline lysis and caesium chloride density centrifugation (section 3.16.2). For each clone (representing the same insert fragment in both orientations) DNA was digested with BamHI and PstI and uni-directionally deleted using the method described by Henikoff (1984): experimental details are given in section 3.17. This involved digestion of the 3' to 5' DNA strand with exonuclease III (for which the end generated by BamHI is a substrate but that generated by PstI is not) for increasing times, thus leaving increasing lengths of single-stranded insert DNA from the BamHI end (the end nearest the M13 sequencing primer binding site), which were removed by digestion with nuclease S1. The rate of digestion by exonuclease III was approx-imately 250 bases/min. The end of the insert DNA and the vector DNA (the latter possessing a PstI end) were blunted with E. coli DNA polymerase I "Klenow" fragment and the deleted products were circularised by self-ligation with T4 DNA ligase. The reactions were used to transform E. coli strain JM101 (section 3.14.3). Several resultant deletion clones were picked from each time point and single-stranded DNA was made for use as a sequencing template (section 3.16.1). Clones were identified according to their derivation from mIIa or mIIb and to the time of digestion with exonuclease III.

Clones were sequenced by the Sanger dideoxynucleotide

method (section 3.18), using either E. coli DNA polymerase I "Klenow" fragment or modified T7 DNA polymerase (Sequenase^{TM}). The radioactive label used was [α^{35}S]dATPαS. Reactions were electrophoresed for 2-16 h through 6% polyacrylamide gels containing 7 M urea, which were fixed and dried for autoradiography. Sequences were read from autoradiographs with a sonic digitiser and recorded on computer disc (section 3.19.1). Since all sequencing reactions were primed with the "universal" 17-mer primer, the sequence obtained from a given clone was derived from within the insert of mIIa/b at a position determined by the extent of its deletion by exonuclease III. Thus, use of deletion clones from a series of time points allowed sequence data to be obtained throughout the 2.5 kb insert. Gel readings were screened to remove M13 vector sequence from clones that had deleted to near the end of the insert, or to within the vector, and were compiled into a contiguous sequence (section 3.19.2). A total of 66 gel readings was used to obtain the complete sequence of the 2.5 kb MCV II SmaI fragment subcloned from pMBH-II. Sequence data were obtained covering both strands of the insert throughout.

The sequence of the 2.5 kb SmaI fragment of pMBH-II is 2,571 bases long and is given in appendix 4A. It has a base composition of 58% G+C, consistent with the earlier observation that MCV DNA is more GC-rich than that of the orthopoxviruses (section 5.1). However, the local base composition varies widely: there are long tracts of either AT-rich or GC-rich sequence. Several of the latter regions were particularly difficult to sequence through (resulting in band

"compressions" on gels) and required the use of dGTP analogues (dITP and 7-deaza-dGTP) in the sequencing reactions. One region required the inclusion of formamide in addition to urea in the gel to increase the denaturing conditions during electrophoresis (section 3.18.4): this region is 88% GC-rich over a stretch of 60 bases (positions 351-410). The sequence lacks restriction sites for the enzymes BamHI, ClaI, EcoRI, HindIII, PstI and SmaI, as expected on the basis of the restriction map of pMBH-II (see figure 18).

6.7 Identifying the oligonucleotide target sequence

As was described above (section 6.4), oligonucleotide 2 hybridised to a 1 kb region of pMBH-II bounded by a SmaI site (position 4 of the derived sequence) and a SalI site (position 1,010). In order to map the position of the target sequence more finely, the oligonucleotide was used to probe a Southern blot of the RF DNA of M13 deletion clones spanning this region: these clones were digested with EcoRI and HindIII to allow separation of the inserts from the vector. Hybridisation was observed to an mIIa deletion clone containing only the first 263 bases (as determined from the position of sequence data derived from it within the completed contiguous sequence) but not to an mIIb deletion clone with the first 181 bases deleted. The oligonucleotide was also used to probe single-stranded DNA of clones mIIa and mIIb, to determine to which DNA strand it bound. Hybridisation occurred only to the strand present in single-stranded DNA of clone mIIa, whose

sequence is that derived from reactions with mIIb templates. Thus, a complementary match should be present in the sense of the sequence as derived from the mIIb templates (the "b" sense, that presented in appendix 4A) within the first 181 bases.

In an attempt to functionally define the target sequence, sequencing reactions were performed in which oligonucleotide 2 was used as the primer. 50 ng oligonucleotide was used and the reactions were otherwise unchanged (section 3.18). However, priming occurred within the M13 sequence at multiple sites and the sequence derived was not readable.

A match to the oligonucleotide was sought in the first 200 bases of sequence, using the computer programs FIND and GAP (section 3.19.3). No good match was found without the insertion of a gap in the alignment. The best direct match (ie. without gaps) was at 12 of 17 positions, in which the longest contiguous match was 5 bases: it is unlikely that such a hybrid would form. The best match with the insertion of gaps allowed had a single gap of one base and matched at 14 of the remaining 16 positions, with the two mismatches occurring at the oligonucleotide ends:-

```
          103                 118
           .                   .
5'....AGCTACACCG AATACT....3'
      :::::::::: :::::
3'-CCGATGTGGC TTATGC-5'
             C
```

The best matching oligonucleotide of the degenerate

mixture is shown. The two unpaired terminal bases would not be expected to destabilise the hybrid. Thus, the only destabilising feature is the single looped-out, unpaired oligonucleotide base. Given that this is between two reasonable lengths of matched sequence this destabilisation would be minimised. The Wallace rule (section 3.11.2) predicts a T_m for this hybrid of 42 ^{o}C, with no account taken of the unpaired base, which accords reasonably with the experimentally determined value of 36 ^{o}C for MCV II DNA (section 6.3).

6.8 Analysis of sequence data

From the above it was clear that the region of the MCV II genome from which sequence data had been derived did not contain a significant match to the oligonucleotide probe used to identify this region as a candidate for encoding an EGF-like polypeptide: the oligonucleotide probe did not select a contiguous match. The nucleotide sequence was translated in all six possible reading frames and scanned (i) for a match to the conserved domains of the EGF-like family, and (ii) for a pattern of cysteine residues corresponding to that present in this family: neither was found. The sequence was, nevertheless, analysed further to see if it contained other information of interest (section 3.19.3).

The sequence was analysed for open reading frames (orfs; ie. regions between start codons and stop codons in the

same frame). There were two complete orfs of significant length and the 3' end of a third potential orf (figure 19): position 1,693-1,115 on the "a" sense strand (orf A; coding for 193 amino-acids), position 525-1,103 on the "b" sense strand (orf B; also 193 amino-acids) and position 2,571-2,158 on the "a" sense strand (orf C; the C-terminal 138 amino-acids of a potentially larger gene product). There are just five bases between the stop codons of orfs A and B on the opposite DNA strands: such clustering of genes is not uncommon for poxviruses (eg. Pluciennicczak et al., 1985).

Three gene-searching algorithms were applied to the sequence. These were the 'Fickett testcode', based on the fact that coding places a period three constraint on base occurrence; and the McLachlan, Staden and Boswell algorithms, based on the fact that only coding regions give rise to codon bias above the random expectation from the base composition and that only coding regions give rise to bias in codon usage for each amino-acid. Orf A scored highly with all algorithms and so can be considered a potential MCV gene; orf B scored less well but significantly; orf C scored poorly. There were no other orfs of significant length or that scored significantly with these algorithms.

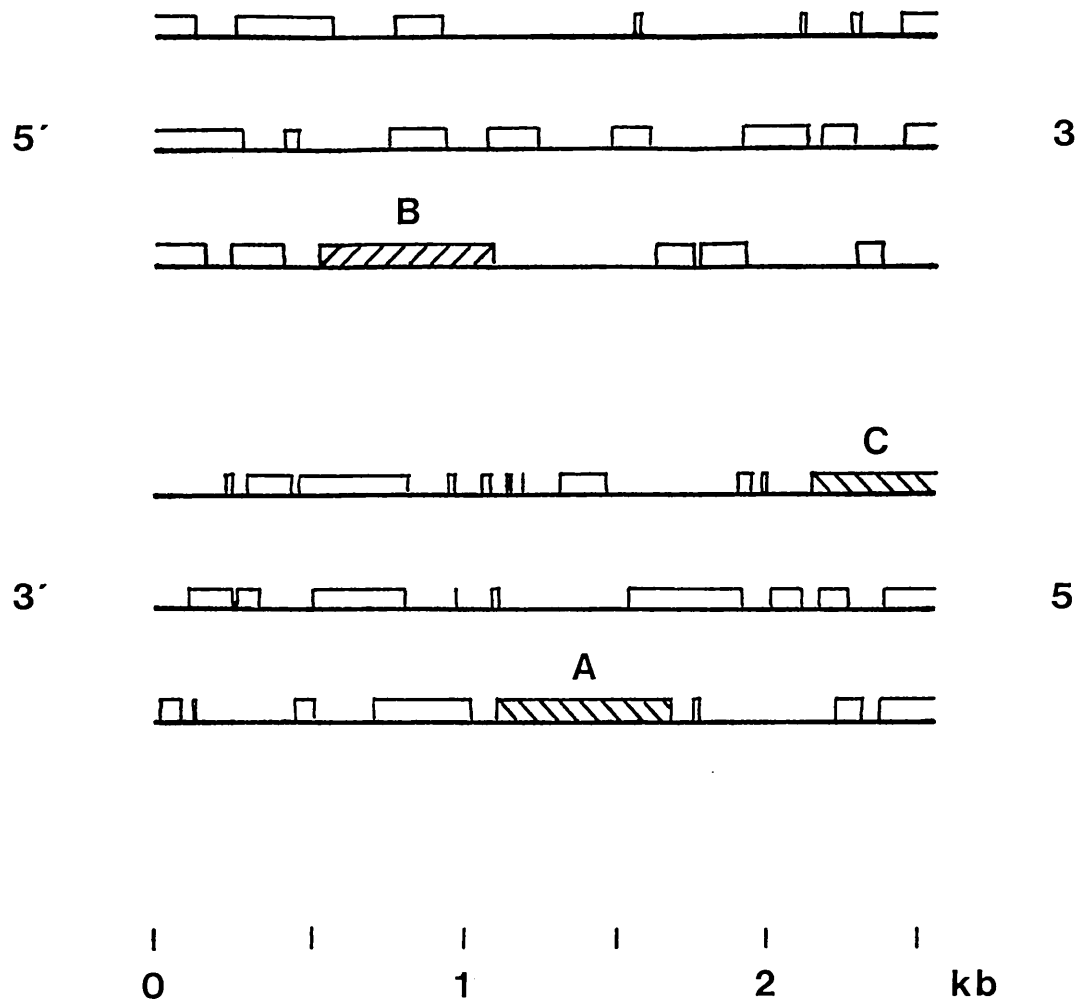


Figure 19. Open reading frames within the 2,571 bases of nucleotide sequence obtained from clone mII. Boxes indicate regions between start and stop codons for each of the six reading frames. The three open reading frames referred to in the text are indicated: orf B on the top strand (bases 525-1,103), and orf A (bases 1,693-1,115) and the 3'-end of orf C (bases 2,571-2,158) on the bottom strand.

The sequence upstream and downstream of these orfs was analysed for the presence of features similar to control sequences known to be of importance in the regulation of other poxviral genes (see section 1.2.4). Downstream of orf A are two matches to TTTTTXT, the putative vaccinia virus early gene transcription termination signal (Rohrmann et al., 1986), one overlapping with TGTAT, a sequence occurring five times close to the 3' end of the VGF mRNA (Yuen and Moss, 1986): TTTTTTTGTAT and TTTTTTTTGT, beginning 549 and 589 bases 3' to the stop codon respectively. Downstream of orf C is one match to this sequence: TTTTTTCT, beginning 95 bases 3' to the stop codon. Orf B has no matches upstream or downstream, although the sequences TTTTTTCT and TTTTTTT occur internally. There are no matches to any other sequence motifs known to regulate poxviral genes. The context of the start codon of orf B matches well with the consensus sequence for the eukaryotic ribosome binding site (Kozak, 1984; Staden, 1984). The sequence upstream of vaccinia virus genes is AT-rich (Pluciennicczak et al., 1985): no such feature is present upstream of orfs A and B. The 100 bases 5' to the start codon from each of orfs A and B were compared: there was no sequence similarity shown between these two putative promoter regions.

The translates of orfs A, B and C (figure 20) were used in a homology search of the NBRF protein sequence database: there were no significant matches. Additionally, the three translates were compared with each other: there is no homology between them. The nucleotide sequence of clone mII (2,571 bases) and the limited sequence derived from clone mIa (364

Orf A

```
  1  MTPLPPCGEI WDLWNQTFPL MTPQNCTSGP DVTVSGYFTY QNGTCHQGKY
 51  IGEVMSSSGS PPNVLGLKGL CQHGNDSERF PWFKFFYYGG GDGSDGDDGQ
101  EDGGYYYDGG GDDGGDDGGG EKCTLSVYNQ TAEDLTDYSP VYWPANTTMP
151  PELQVRLNQT SDSNVLNVTI PPGLFDEAAF VNEQTSGFPL SRI
```

Orf B

```
  1  MVFRLDFDAY TKKRKELWHV AAARSCAAAS QRTAADTPRA EYTHLCRELP
 51  LRTRQCRGHR TCHATAACRR HRCVTPARSH ARELSSRSNC GWLHLCPSLS
101  QRRFSCCFFL LCFFPVMEQK VWGKDGPSLP SKRMTSPEKG GRGAMYMSCS
151  HTALKYWTFV GRHPAPATML VCESGRLRIT CASVLSVCLY HEA
```

Orf C (3'-end)

```
  1  PGLRPALQRH GGRAARGPSR ASARRAACAA RAALVRSAHG ICRRRHARAR
 51  VGATSFFSAK LRIKRPLPKL RAACASCDVL ARWGRTRRVP FLRYSGAGGI
101  GVRHGVACSR ARRDARSLVK VRAQHSLLLA LLMMLNNL
```

Figure 20. Amino-acid sequences of the putative translates of orfs A, B and C from the nucleotide sequence of clone mII. Refer to figure 19. The single letter amino-acid code is used (see legend to table 2).

bases; see below) were used in a homology search of entries in the Genbank nucleic acid sequence database containing poxviral sequence data. No homology with previously sequenced poxviral DNA was detected.

6.9 Sequence comparison with MCV I

Some sequence data (364 bases) were obtained from the MCV I clone mIa, allowing comparison with that from MCV II (figure 21). The MCV I sequence matches the MCV II sequence at the position expected from alignment of the restriction maps of pMBH-I and pMBH-II (see figure 18): they are 86.8% homologous. This level of homology applies to the region of the clones for which their restriction maps are not identical and so for which homology is expected to be lowest. The PstI site of MCV I and the SalI site of MCV II each differ in the sequence of the other subtype by a single base within the recognition sequence of 6 bases: this was the mechanism assumed to account for the different genomic restriction maps of the MCV subtypes despite their high level of homology as determined by hybridisation studies (section 5.8). An additional difference observed between the two sequences was the relative deletion of 1 and 4 bases in that from MCV II and MCV I respectively, at two different positions within the region of comparison.

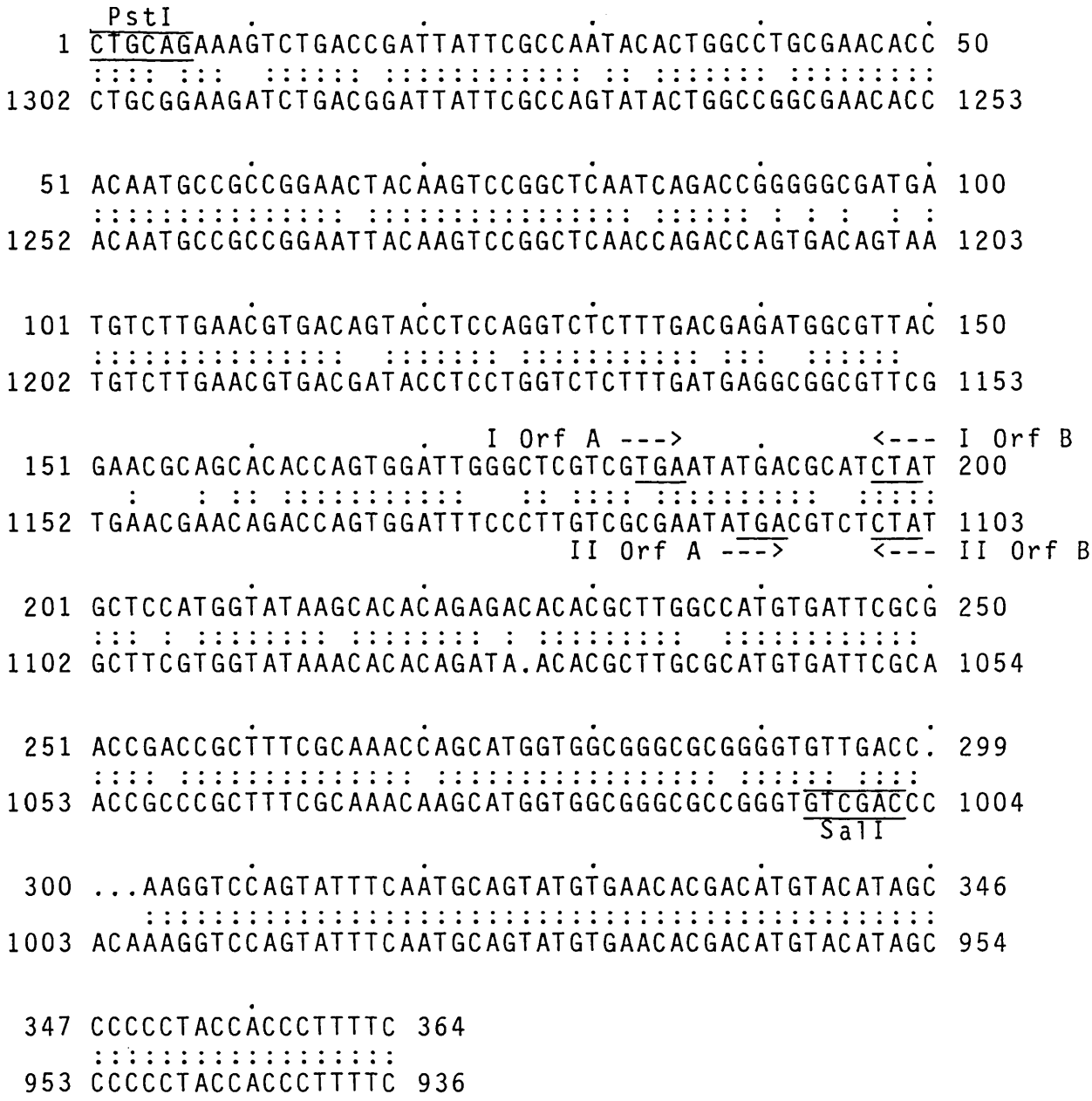


Figure 21. Comparison of the 364 bases of sequence data obtained from clone mIa (from MCV I) with that of mII (MCV II). The upper sequence is that from mIa. The alignment is 368 bases long and has two gaps, one of a single base and one of four bases. The overall homology is 86.8%. The PstI site of the MCV I sequence (position 1-6) and the SalI site of the MCV II sequence (position 1,006-1,011) are marked: for each there is a single base change in the sequence from the other MCV subtype. Also indicated are the stop codons of orf A and orf B of the MCV II sequence and those of the equivalent orfs from MCV I, for which only the 3' sequence is available.

The region of the MCV II sequence that was compared with that from MCV I includes the 3'-ends of both orf A and orf B. The stop codons of these orfs are indicated in figure 21. The MCV I sequence equivalent to these orfs is also open in the same reading frame as MCV II: orf A finishes two codons earlier and orf B fininshes at the same position relative to MCV II. The C-terminal 60 amino-acids of the translate of MCV I orf A are 86.7% similar (71.7% identical) to those of MCV II. The two short nucleotide deletions mentioned above occur within orf B. The 4 base deletion in MCV I deletes a codon and advances the reading-frame by one base relative to MCV II: the reading-frame is restored 74 bases later by the single base deletion in MCV II. The reading frame in MCV I stays open despite switching frames relative to MCV II, but the translate differs from that of MCV II for 24 amino-acids. The C-terminal 55 amino-acids of the translate of MCV I orf B are 72.7% similar (60.7% identical) to those of MCV II. The C-termini of the translates of orfs A and B from MCV I and MCV II are compared in figure 22.

Orf A: MCV I/II

```
  1 AESLTDYSPIHWPANTTMPPELQVRLNQTGGDDVLNVTVPPGLFDEMALR 50
    :: ::::::: :::::::::::::::::::: :::::::::::::: ::
132 AEDLTDYSPVYWPANTTMPPELQVRLNQTSDSNVLNVTIPPGLFDEAAFV 181

 51 TQHTSGLGSS   60
    ::::::  :
182 NEQTSGFPLSRI 193
```

Orf B: MCV I/II

```
  1 EKGGRGAMYMSCSHTALKYWTLVNTPRPPPCWFAKAVGRESHGQACV.SV 49
    ::::::::::::::::::::::::    :::  :    ::     ::: ::
138 EKGGRGAMYMSCSHTALKYWTFVGRHPAPATMLVCESGRLRITCASVLSV 187

 50 CLYHGA 55
    ::::::
188 CLYHEA 193
```

Figure 22. Comparison of the C-terminal amino-acid sequences of the translates of MCV II orf A and orf B with their MCV I equivalents. The single letter amino-acid code is used (see legend to table 2). The upper sequence is that derived from the MCV I nucleotide sequence. The putative translate of MCV I orf A terminates 2 amino-acids earlier than that from MCV II: the level of similarity is 86.7% (71.7% identity). The putative translate of MCV I orf B is frame-shifted relative to that from MCV II over a region of 24 amino-acids, due to the short relative deletions in the nucleotide sequences: the level of similarity is 72.7% (60.7% identity).

6.10 Summary of chapter

Cloned human EGF and VGF gene probes do not cross-hybridise with MCV DNA. A degenerate oligonucleotide (a 17-mer), designed to identify DNA sequences encoding the conserved third disulphide-loop domain of members of the EGF-family of growth factors, has been used to probe MCV DNA. Hybridisations delineated a genomic locus, geographically equivalent in all three MCV subtypes, close to the left terminus. This locus from MCV II DNA was cloned into pBR322 and subcloned into M13, from which 2,571 bases of sequence were obtained by dideoxynucleotide sequencing of a series of deletion clones generated by exonuclease III. The oligonucleotide target sequence was identified but is not contiguous (a single base is looped out) and does not represent a sequence of the type sought.

The sequence contains three open reading frames which are potential MCV genes. The only previously described poxviral control sequence present is an early gene transcription termination signal downstream of two of these. The 3'-ends of two of these open reading frames, encoded on opposite strands and finishing with just five bases between them, were also sequenced from MCV I. Nucleotide sequence homology is high (87%), as is that of the putative translates, despite a frame-shifted region of one relative to the other viral subtype. Homology was not detected between the sequence data obtained and known poxviral sequences.

CHAPTER 7

Use of synthetic oligonucleotide probes, part 2: the MCV homologue of the vaccinia virus major envelope antigen

Despite being unsuccessful with the oligonucleotide probes described in the previous chapter, new probes were constructed to detect sequences capable of encoding the highly conserved domain A of EGF-like growth factors. The design rationale was changed in this second attempt.

7.1 Design of oligonucleotides 5 and 6

Two new degenerate oligonucleotides were designed. They were 20-mers covering the region GYTGIRC of domain A of VGF (of which T and I are the non-conserved positions; see section 6.2). This domain includes one amino-acid with 6 codons (arginine, R), three with 4 codons (two glycines, G; threonine, T), one with 3 codons (isoleucine, I) and two with 2 codons (tyrosine, Y; cysteine, C). The sequence of VGF was chosen for the reasons discussed in section 6.2. However, the means of reducing degeneracy by allowing G-T "wobble" mismatches was not used. Rather, the degeneracy was reduced from the 4,608 possibilities for the above septapeptide by opting for a single codon for arginine and by synthesising two related oligonucleotides, each with half of the total degeneracy.

The third position of the cysteine codon was not covered by the oligonucleotides and codon usage was applied for arginine: in vaccinia virus genes, one of the six codons predominates (namely AGA, used for 50% of arginine codons) and this was chosen in the design of oligonucleotide probes 5 and 6 in the hope that a similar usage might be true of MCV genes. It was not convenient to use just three bases at the third position of the isoleucine codon as well as using all four bases at other positions of the oligonucleotide: thus, the degeneracy was 512 for practical purposes. Two oligonucleotides were made, each having one of the possible bases at the doubly-degenerate third position of the tyrosine codon, and thus each having a degeneracy of 256. As well as being a means of reducing the degeneracy of each, this decision meant that one of the pair of oligonucleotides should match the putative MCV gene better than the other. Hence, some encouragement would be given if they did behave differentially. This design of oligonucleotide also allowed the use of a positive control in hybridisation experiments, which was lacking for the earlier designs: one component of the degenerate mixture of oligonucleotide 5 is an exact match to the sequence in the VGF gene. The sequences of oligonucleotides 5 and 6 are given in table 4A.

7.2 Synthesis and use of oligonucleotides 5 and 6

Oligonucleotides 5 and 6 were synthesised automatically using phosphoramidite chemistry, deprotected automatically and purified by polyacrylamide gel electrophoresis (section 3.9). They were 5' end-labelled with T4 polynucleotide kinase and [γ^{32}P]ATP and separated from unincorporated nucleotides by exclusion chromatography (section 3.10.2). Southern blots of digests of MCV I and MCV II DNA on nitrocellulose were prepared and pre-hybridised, using tRNA as the blocking agent (section 3.11.2). The blots also contained digests of pVGF as a positive control. Probes were added and hybridisation was continued overnight.

The theoretical melting temperatures for perfectly matched oligonucleotides were calculated using the Wallace rule (section 3.11.2). This gave a range of possible values for the T_m of 54-62 ^{o}C and 56-64 ^{o}C for oligonucleotides 5 and 6 respectively. The T_m for a hybrid of oligonucleotide 5 and pVGF was calculated to be 56 ^{o}C: based on the rule of 5 ^{o}C destabilisation per mismatch, that for oligonucleotide 6 and pVGF would be approximately 49 ^{o}C (due to the loss of an A-T match and gain of a mismatch). Initially, hybridisation was at 40 ^{o}C and washing was for 30 min at 40 ^{o}C and 40 min at 45 ^{o}C. This yielded a signal from both probes to pVGF but no hybridisation of either to MCV DNA. The signal due to oligonucleotide 6 was approximately 10-fold weaker than that for oligonucleotide 5, reflecting destabilisation by a single G-T mismatch at the sixth position from the 3' end of the

oligonucleotide. This experiment thus provided evidence of the significance of the destabilising effect of mismatches. Particularly relevant was the nature of this mismatch: namely, a G-T "wobble" mismatch that has been regarded as being poorly destabilising (Thein and Wallace, 1986) and that was featured prominantly in the design of oligonucleotides 1-3 in the previous chapter.

The experiment was repeated with a hybridisation temperature of 38 ^{o}C and washing for 50 min at this temperature. Signals with both probes to pVGF were comparable but all other controls were as above. Only oligonucleotide 6 showed hybridisation to MCV DNA, to MCV II BamHI fragment C, ClaI fragment A and HindIII fragment M (figure 23). There were also faint signals to ClaI fragments E and F, and to HindIII fragment I. There was no hybridisation to MCV I DNA. As discussed in section 5.9, the terminal and near-terminal HindIII fragments of MCV II DNA vary in size between different isolates of this subtype. For the isolate used in this experiment (number 86), HindIII fragment M corresponds to HindIII fragment L in the isolate from which restriction maps were derived in chapter 5. (This is because the 3-molar fragment I of the latter isolate was "split" into its component 2-molar terminal fragment I and the near-terminal fragment I', labelled J in the present isolate; see section 5.5).

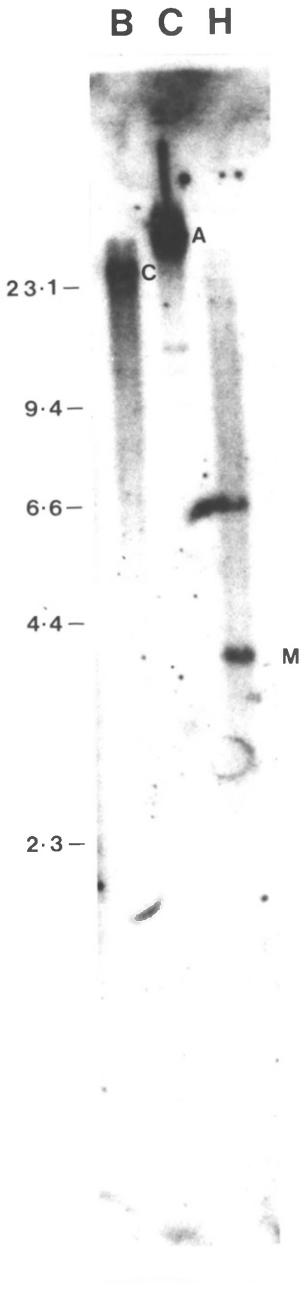


Figure 23. Hybridisation of oligonucleotide 6 to BamHI, ClaI and HindIII digests of MCV II DNA. The hybridising fragments are identified by letter. Figures refer to bacteriophage lambda DNA HindIII fragment size markers.

The hybridising fragments of MCV II DNA map together, delineating the region of the genome containing the target sequence as being 27-31 kb from the left terminus (see figure 9). That oligonucleotide probes 5 and 6 behaved differentially was taken as indicative of a valid hybridisation (see section 7.1). That the hybrid formed between oligonucleotide 6 and MCV II DNA was weaker than that formed between oligonucleotide 6 and pVGF (which has one mismatch) suggested the presence of mismatches in the former. Although oligonucleotide 6 did not hybridise to MCV I DNA, MCV II HindIII fragment L does so - to MCV I HindIII fragment L at the geographically equivalent position (section 5.9). The differential hybridisation presumably reflects sequence differences between the two subtypes.

7.3 Cloning the oligonucleotide-positive locus

The HindIII fragment M of this isolate of MCV II was of a suitable size for cloning (3.4 kb). Approximately 8 ug (60 fmol) of DNA was digested with HindIII and electrophoresed through 0.7% low melting-point agarose. The gel slice containing the HindIII fragment M was removed and the DNA was recovered (section 3.13.1). pBR322 DNA was digested with HindIII and phosphatased with calf intestinal phosphatase (section 3.13.3). Insert and vector DNAs were combined and ligated with T4 DNA ligase for 16 h at 17 ^{o}C, using estimated concentrations of 3 nM and 1 nM respectively (section 3.13.4).

The ligation reactions were used to transform E. coli strain DH1, plating on L. agar with ampicillin selection (section 3.14.3). Transformants were grown overnight in L. broth containing ampicillin and used for the preparation of plasmid DNA by alkaline lysis (section 3.15.1). Plasmid DNA was digested with HindIII to verify the clone insert size. One clone was chosen and its identity was confirmed by hybridisation to MCV II DNA. Oligonucleotide 6 also hybridised to this clone, which was named pH/M-IIB.

Restriction sites for the enzymes BamHI, EcoRI, PstI, SalI, SmaI and XhoI were mapped on plasmid DNA of this clone by analysis of single- and double-digests. A Southern blot of these digests was probed with the oligonucleotide and the locus to which it hybridised was determined to be within a 0.45 kb region bounded by a SmaI site and a XhoI site (figure 24).

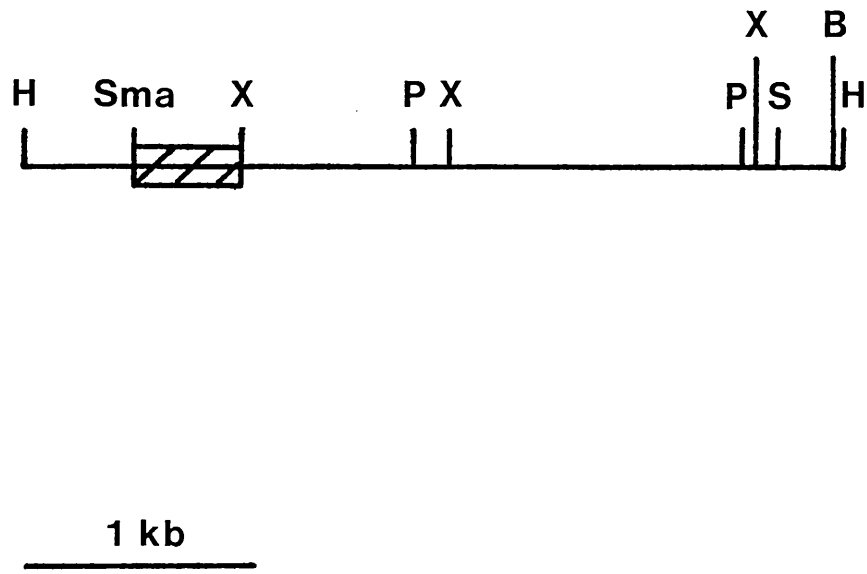


Figure 24. Restriction map of clone pH/M-IIB. The enzymes used are BamHI (B), HindIII (H), PstI (P), SalI (S), SmaI (Sma) and XhoI (X). The locus of oligonucleotide 6 hybridisation is indicated. The SmaI/XhoI fragment containing this locus was subcloned into M13 for sequencing.

7.4 Subcloning from pH/M-IIB into M13mp18

It was decided to subclone the SmaI/XhoI fragment of pH/M-IIB into M13mp18, from which the complete sequence could be derived by sequencing each strand. Thus, both orientations were required. pH/M-IIB DNA was digested with SmaI and XhoI and recovered after extraction with phenol:chloroform:amyl alcohol (25:24:1). The DNA was incubated in 50 mM Tris.Cl pH 7.2, 10 mM $MgSO_4$, 0.1 mM DTT, 50 ug/ml BSA, 80 uM each of dATP, dCTP, dGTP, dTTP, and 6 units *E. coli* DNA polymerase I "Klenow" fragment for 30 min at room temperature, in order to fill in the 5'-overhang left by XhoI digestion. The digest was electrophoresed through 1.5% low melting-point agarose and the required fragment was excised and recovered (section 3.13.1). M13mp18 RF DNA was digested with SmaI and phosphatased with calf intestinal phosphatase (section 3.13.3). Insert and vector DNAs were combined and blunt-end ligated with T4 DNA ligase for 16 h at 19 ^{o}C, using estimated concentrations of 6.7 nM insert and 6.7 nM vector (section 3.13.4).

The ligation reactions were incubated for 10 min at 65 ^{o}C and used to transform *E. coli* strain JM101, which was plated on a lawn of JM101 on H. agar with IPTG/XGal selection (section 3.14.3). Individual plaques were picked, grown in fresh JM101 and RF DNA was prepared to determine the insert size. Due to the inefficiency of blunt-end ligation, only one clone was obtained, named mSX. The opposite orientation was obtained by "clone turnaround", as follows. mSX RF DNA was

digested with BamHI and EcoRI, which cut within the M13 multiple cloning site flanking the insert, and the insert fragment was recovered from low melting-point agarose as above. M13mp19 RF DNA was digested with the same two enzymes and phosphatased. Insert and vector DNAs were combined and ligated as above, using estimated concentrations of 15 nM and 8 nM respectively. The ligation reactions were used to transform JM101, and RF DNA from transformants was analysed by digestion with SmaI (for which the site was re-created at one end of the insert only) and by double-digestion with SmaI and BamHI. The double-digest revealed the clone orientation with respect to the sequencing primer site. A clone in the opposite orientation to mSX was selected and named mSXrev.

7.5 DNA sequence analysis and the oligonucleotide target sequence

Single-stranded DNA of clones mSX and mSXrev was prepared and sequenced, using Sequenase™. The reactions were electrophoresed through 6% polyacrylamide gels containing 7 M urea for 2.5 and 16 h (section 3.18). Sequence data were read using a sonic digitiser and compiled into a contiguous sequence (section 3.19.1). The sequence derived is 423 bases long and is given in appendix 4B. The complete sequence was derived from each strand. It has a base composition of 67% G+C, similar to that observed earlier (58%; see section 6.6).

The oligonucleotide target sequence was searched for

using the computer programs FIND and GAP (section 3.19.3). The best match consists of an uninterrupted run of 13 matched bases, with the oligonucleotide ends poorly matched:-

```
          149                 168
           .                   .
5'....GCTTACACGGGCATTCCTTT....3'
       : ::::::::::::::    :
   3'-CCAATGTGCCCGTAATCTAC-5'
```

The best matching oligonucleotide of the degenerate mixture is shown. It is unlikely that the 3'-terminal base of the oligonucleotide would be bound, due to destabilisation by the adjacent mismatch. Similarly, the five bases at the 5' end would not bind. However, terminal mismatches would be unlikely to destabilise the contiguous match of 13 bases, for which the T_m calculated from the Wallace rule (section 3.11.2) is 38 ^{o}C. This is the same as the experimentally determined value. Oligonucleotide 5, which does not hybridise to this DNA, differs from number 6 at the sixth position from the 3' end, interrupting the contiguous match. Despite the oligonucleotide match, there was no match to the conserved features of the EGF-family at this or any position within the sequence data obtained.

7.6 MCV homologue of the vaccinia virus p37 gene

The sequence was used in a homology search of entries within the Genbank nucleic acid sequence database containing poxviral sequence data. A significant match was found to the

sequence of the vaccinia virus gene encoding the major envelope antigen of 37 kD (p37; Hirt et al., 1986), indicating that part of the homologue of this gene had been cloned and sequenced. The p37 gene is expressed late in vaccinia virus infection and its product comprises 5-7% of the total protein mass of the virus. It maps to approximately 38 kb from the left terminus of the vaccinia virus genome (190 kb total length), similar to the locus from which clone mSX was derived, 30-31 kb from the left terminus of the MCV II genome. These figures respectively represent distances of 28 kb and 24 kb from the junction of the terminal repeat and unique sequences at the left end of the genomes. Also, the direction of transcription of the MCV homologue, as deduced from the orientation of the sequence, is towards the terminus, as is that of the vaccinia gene.

Nucleotides 423-177 of the mSX sequence correspond to the region of the gene encoding the C-terminus of p37. There are no gaps in the alignment between the MCV and vaccinia virus nucleotide sequence data in this region and there is a 42% level of homology. This region of the mSX sequence was translated (82 amino-acids) and compared with the sequence of p37 (figure 25). The MCV translate extends 13 amino-acids beyond the C-terminus of p37. Over the region of alignment (69 amino-acids) there is 62.9% similarity (49.3% identity). Over a stretch of 56 of these amino-acids (positions 8-64) there is 75% similarity, including 57% identity, and the nucleotide homology within this region is 53%.

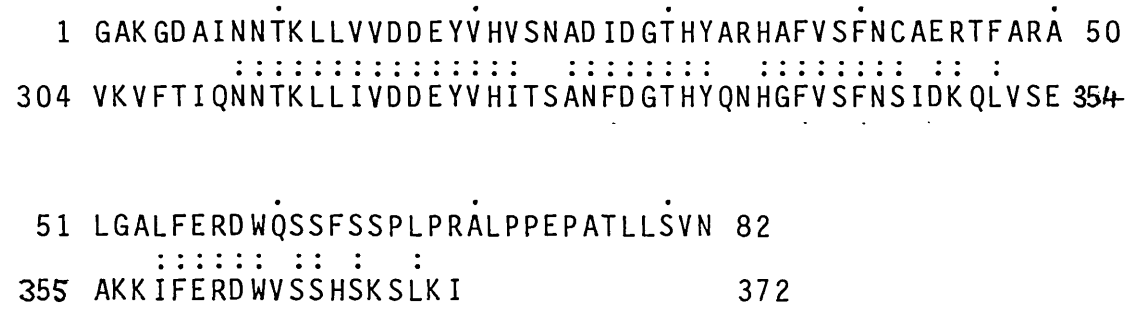


Figure 25. Comparison of the amino-acid sequences of the C-termini of the vaccinia virus 37 kD major envelope antigen (p37) and its MCV II homologue. The lower sequence is that of p37 and is taken from Hirt et al. (1986). The terminal 69 and 82 amino-acids are shown, the MCV homologue being 13 longer than the vaccinia protein. The single letter amino-acid code is used (see legend to table 2). The level of homology is 62.9% similarity (49.3% identity).

The frequency with which codons were used was analysed from this limited data of 82 amino-acids in the hope that information on MCV codon usage might be of value in the construction of further oligonucleotides. Codon usage was found to differ from that displayed by vaccinia virus. Significantly, of five occurrences of arginine within the C-terminus of the MCV homologue of p37, the codon AGA was not used: this codon is used 50% of the time by vaccinia genes and was that chosen for the design of oligonucleotide probes 5 and 6. There is a strong bias towards G or C at the third position of codons (87%), reflecting the relatively high G+C base content of the MCV genome (67% in the sequence of clone mSX). In contrast, A and T are favoured in the third position of vaccinia virus codons (70%).

The sequence immediately following the stop codon of the p37 gene is A-rich: TAAAAAAAAGAAAATA etc., where the underlined bases are the stop codon (Hirt et al., 1986). The MCV gene finishes similarly: TGAAAAAAAAAGGAAT etc., implying a functional role for this motif, perhaps as a termination of transcription signal. No other poxviral genes for which data are available possess this feature. 173 bases of nucleotide sequence 3' to the stop codon of the MCV gene was obtained: there were no matches to other previously described poxviral regulatory sequences within this region (see section 1.2.4).

7.7 Summary of chapter

A pair of oligonucleotides (20-mers) was designed to identify DNA sequences encoding the third disulphide-loop domain of members of the EGF-family of growth factors. One of these identified a locus of MCV II DNA which was cloned into pBR322 and subcloned into M13, from which 423 bases of sequence were obtained by dideoxynucleotide sequencing. The oligonucleotide target sequence was identified but does not represent a sequence of the type sought: the hybrid is matched at only 13 contiguous bases.

The sequence contains the 3'-end of an open reading frame whose translate is 63-75% homologous to the vaccinia virus major envelope antigen (p37). The MCV II homologue is encoded at a similar genomic position to that of the p37 gene in the vaccinia genome and the two genes are in the same orientation (towards the left terminus).

CHAPTER 8

Use of synthetic oligonucleotide probes, part 3

Although unsuccessful in identifying the gene of interest, the investigation of the previous chapter yielded useful data regarding the organisation of coding sequence for the MCV homologue of the vaccinia virus p37 gene. This information was exploited in the design of new oligonucleotides for the detection of a putative MCV EGF-like gene.

8.1 Design of oligonucleotides 8-13

Six new oligonucleotides were designed, as pairs, covering both domain A and domain B of the EGF-like family. They were 24-mers (domain A) and 22-mers (domain B). Pairs of oligonucleotides, differing at one position only, were again used in order to reduce the degeneracy of each and to gain insight into the their behaviour in hybridisation experiments. In all three cases the two oligonucleotides differed in the base covering the doubly-degenerate third position of a cysteine codon: differential hybridisation would provide evidence of this codon being involved in hybrid formation.

Two pieces of information relevant to the design of new oligonucleotide probes were extracted from the partial sequence of the MCV homologue of the vaccinia virus p37 gene

(see section 7.6). The first of these was the observation that the bases G or C are highly preferred at the third position of codons, occurring in 87% of the 82 codons sequenced, and is a reflection of the relatively high G+C content of the MCV genome. Some indication of MCV codon preferences was also given, these differing from those of vaccinia virus. However, the data are limited (82 codons only) and were not relied upon too heavily for this and the following reason. It was considered possible that the codon preferences exhibited by a late gene which is abundantly expressed may differ from those of an early gene with a lower level of expression (drawing the analogy from the vaccinia virus p37 and VGF genes respectively). This possibility was tested by comparing codon preferences in the p37 and VGF genes: there are some significant differences but the third codon positions are consistently 70% A or T. Thus, the codon preference data from the MCV p37-homologue partial sequence were not considered a wholly reliable guide regarding those of a putative EGF-like gene. However, the design of the new oligonucleotides did fully incorporate the base preference data at the third position of each codon covered. The only amino-acids for which any further choices were made were arginine and leucine (the codons AGA/G and TTA/G did not occur at all in the 5 and 8 occurrences of these amino-acids respectively): codon preferences for these are similar in the p37 and VGF genes.

As noted in section 6.2, the amino-acid following the conserved domain A is either glutamine (Q) or glutamate (E) in every sequenced member of the family of EGF-like growth

factors. Oligonucleotides 8-11 were designed against this extended domain:-

G Y - G - R C Q/E

where dashes indicate the non-conserved positions. Previous oligonucleotide designs to this domain used the amino-acids present in VGF at these positions (T and I respectively). This was again the case for the oligonucleotide pair 8 and 9. Although an assumption of this sort is necessary in order to reduce oligonucleotide degeneracy to a reasonable level, it was considered that this choice alone was too restrictive. The first of the non-conserved positions is valine (V) in four cases and isoleucine (I) in another two cases of the nine sequences available (the sequences of VGF-homologues from Shope rabbit fibroma virus and myxoma virus were now available in addition to those discussed in section 6.2; see section 1.3). The codons expected to be preferred (ie. with G or C in the third position) for these two amino-acids differ at one position only (underlined): ATC for I, GTG/C for V. Similarly, the second of the non-conserved positions is aspartate (D) or glutamate (E) in four cases: GAC for D, GAG for E. The oligonucleotide pair 10 and 11 was designed against V and E/D at the non-conserved positions, this being a composite of the slight preferences shown at these two positions. The only known instance when this combination is known to occur is in guinea-pig EGF, which uses V and E respectively (see table 2). It was reasoned that, between them, oligonucleotides 8-11 covered many of the coding possibilities

of this domain and that the second design overcame many of the restrictions of the first.

Oligonucleotides 8-11 were 24-mers and were 64-fold degenerate. They were synthesised to allow hybridisation to the coding (ie. "sense") strand since it was hoped to use them to prime sequencing from selected clones, thus giving upstream sequence data. The sequences of oligonucleotides 8-11 are given in table 4A. Because of their design, positive controls were not possible: the best match was that of oligonucleotide 11 to the human EGF gene, matching at 22 of 24 positions.

Domain B, shared by all EGF sequences and that of VGF, is followed by the amino-acids histidine (H) or asparagine (N) in every sequenced member of the family of EGF-like growth factors, and this position is followed by glycine (G) in six cases (see table 2). The domain is preceded by histidine or tyrosine (Y) in eight cases. Oligonucleotides 12 and 13 were designed against this extended domain:-

H/Y D G Y C L H/N G

As discussed above, their design was based on the preference for G or C at the third position of codons in MCV genes. The codons thus expected to be preferred for amino-acids at the variable positions differ by one base only (underlined): CAC for H, TAC for Y; CAC for H, AAC for N. The first base of the H/Y codon and the third of the final G codon were not covered by the oligonucleotides.

Oligonucleotides 12 and 13 were 22-mers and were 8-fold degenerate. They were synthesised in the coding sense since it was hoped to use them to prime sequencing from selected clones, thus giving downstream sequence data covering the conserved domain A. The sequences of oligonucleotides 12 and 13 are given in table 4B. As with oligonucleotides 8-11, positive controls were not available for these probes, the best match being that of oligonucleotide 12 to the human EGF gene, matching at 19 of 22 positions.

8.2 Synthesis and use of oligonucleotides 8-13

Oligonucleotides 8-13 were synthesised automatically using phosphoramidite chemistry, and purified and deprotected using NENSORB-PREP columns (section 3.9). They were 5'-end labelled with T4 polynucleotide kinase and [γ^{32}P]ATP, and separated from unincorporated nucleotides by exclusion chromatography (section 3.10.2). Southern blots of digests of MCV I and MCV II DNA on nitrocellulose were prepared and pre-hybridised, using tRNA as the blocking agent (section 3.11.2). Probes were added and hybridisation was continued overnight.

The theoretical melting temperatures for perfectly matched oligonucleotides were calculated using the Wallace rule (section 3.11.2). This gave values for the T_m of 82 and 80 ^{o}C for probes 8 and 9, 84 and 82 ^{o}C for probes 10 and 11, and 72-74 and 70-72 ^{o}C for probes 12 and 13. However, oligo-

nucleotides of this length are approaching the size limit for which this rule is valid and so these figures are probably over-estimates. Initially, hybridisations were at 45 ^{o}C and washing was for 20 min at this temperature. Filters were re-washed at higher temperatures as appropriate.

Oligonucleotide 8 hybridised to many MCV fragments when washed at 50 ^{o}C, although its partner (9) did not: however, upon further washing, there was no particular hybridisation that was more significant than any others. No further use was made of these oligonucleotides (which were designed against the VGF sequence, as were all previous oligonucleotides).

At 45 ^{o}C, oligonucleotide 10 hybridised to many MCV fragments, whilst its partner (11) hybridised more selectively to a few MCV I fragments. The signal from oligonucleotide 11 was stable and strong after washing for 30 min at 58 ^{o}C (figure 26A), although that from its partner was washed off, providing evidence that the stable hybrid involved base pairing to the oligonucleotide sequence covering the cysteine codon. Oligonucleotide 11 hybridised to MCV I BamHI fragment B and HindIII fragment Q, which map together near the left genomic terminus (see figure 7). There was no significant hybridisation to MCV II DNA, presumably due to sequence divergence.

Oligonucleotide 12 hybridised to several MCV fragments when washed at 45 ^{o}C, whilst its partner (13) did not give any signals. The signal was still present at 54 ^{o}C, although

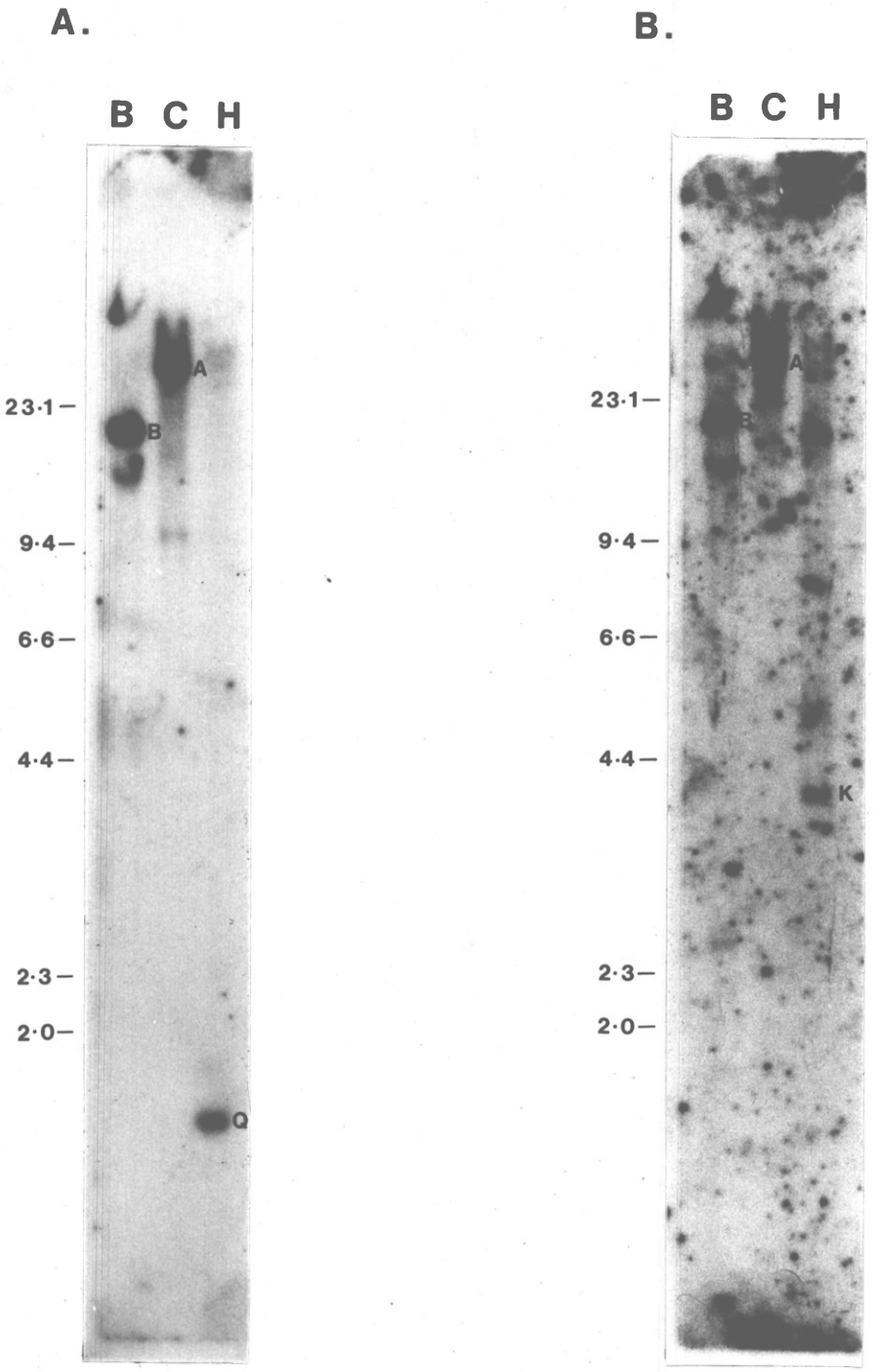


Figure 26. Hybridisation of oligonucleotides 11 (A) and 12 (B) to BamHI and HindIII digests of MCV I DNA. The hybridising fragments are identified by letter. Figures refer to bacteriophage lambda DNA HindIII fragment size markers.

much weaker than that obtained for oligonucleotide 11 (figure 26B). Oligonucleotide 12 hybridised to MCV I BamHI fragment B and HindIII fragment K, which map together near the left genomic terminus, this locus being adjacent to that for oligonucleotide 11 (see figure 7). Again, there was no significant hybridisation to MCV II DNA.

The hybridisation of an oligonucleotide for each of domains A and B to HindIII fragments mapping adjacently in the near-terminal region of MCV I suggested the possibility of hybridisation to similar domains in an MCV EGF-homologue, between which there is a HindIII site. The positions of the target sequences are comparable to that of the VGF gene in the vaccinia virus WR genome (Venkatesan et al., 1982) and, more particlarly, to the position of the VGF gene homologues in wild-type orthopoxvirus genomes, for which the gene occurs within the unique sequences close to the left terminal repeat (Archard et al., 1986).

8.3 Cloning the oligonucleotide-positive loci

It was decided to clone directly the HindIII fragments K (4.0 kb) and Q (1.3 kb) from MCV I DNA. HindIII fragment K is 2-molar, the required component being that mapping near the left terminus (K_L). 8 ug of MCV I DNA (60 fmol) was digested with HindIII and the desired fragments were isolated after electrophoresis through low melting-point agarose (section 3.13.1). pBR322 DNA was digested with HindIII and phosphatas-

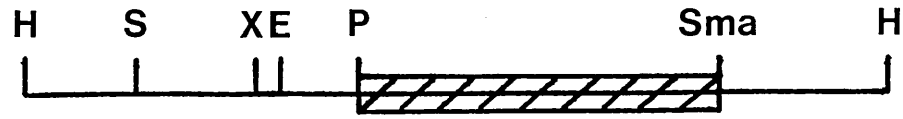
ed with calf intestinal phosphatase (section 3.13.3). Insert and vector DNAs were combined and ligated with T4 DNA ligase for 16 h at 18 ^{o}C, using estimated concentrations of 5 nM and 1.6 nM respectively (section 3.13.4).

The ligation reactions were used to transform E. coli strain DH1, plating on L. agar with ampicillin selection (section 3.14.3). Plasmid DNA was prepared from transformants (section 3.15.1), digested with HindIII and screened by hybridisation with the relevant oligonucleotide probe. One clone selected with each was chosen and its identity was confirmed by hybridisation to MCV I DNA. The clones were named pH/K_L-I and pH/Q-I for HindIII fragments K_L and Q respectively.

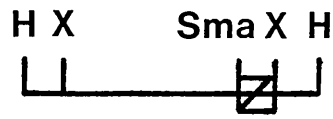
Restriction sites for the enzymes EcoRI, PstI, SalI, SmaI and XhoI were mapped, using single- and double-digestion of plasmid DNA with these enzymes and BamHI (figure 27). Southern blots of these digests were probed with the relevant oligonucleotide (11 for pH/Q-I, 12 for pH/K_L-I) to map the target sequences more precisely. Oligonucleotide 11 hybridised to a region of approximately 150 bp bounded by a SmaI site and a XhoI site. Oligonucleotide 12 hybridised to a 1.6 kb region bounded by a PstI site and a SmaI site. The orientation of clone pH/K_L-I relative to the genomic maps was determined from its internal PstI site, as follows. The left terminal PstI fragment was known to be either fragment C or G (see section 5.3). The former (7.1 kb) is larger than the sum of the sizes of HindIII fragments K and O whilst the latter

(5.4 kb) is 2.5 kb greater than that of HindIII fragment O: since there is a PstI site 2.5 kb from one end of pH/K_L-I, the left terminal PstI fragment must be G and the clone could thus be orientated. This orientation of the clone map was also in accord with the sizes of the terminal SmaI and XhoI fragments. This analysis revealed that the oligonucleotide 12 target sequence in HindIII fragment K_L is within unique sequence (since there was no hybridisation to HindIII fragment J at the right terminus), close to the junction with the terminal repeat.

From the above analysis it was determined that the target sequences for the two oligonucleotides (11 and 12) are at least 1 kb apart in the MCV genome. This suggested that at least one of them was not part of the gene being sought, since domain A and domain B of the EGF-family of growth factors are only 19-23 amino-acids apart. The hybridisation signals with oligonucleotide 11 to HindIII fragment Q were much stronger and cleaner than those with oligonucleotide 12 and so clone pH/Q-I was selected for nucleotide sequencing.



pH/K_L-I



pH/Q-I

1 kb

Figure 27. Restriction maps of clones pH/Q-I and pH/K_L-I. The enzymes used are EcoRI (E), HindIII (H), PstI (P), SalI (S), SmaI (Sma) and XhoI (X). The loci of hybridisation with oligonucleotides 11 and 12 are indicated for pH/Q and pH/K_L respectively. The SmaI/HindIII fragment of pH/Q-I containing this locus was subcloned into M13 for sequencing.

8.4 Subcloning from pH/Q-I into M13mp18 and 19

The 350 bp SmaI/HindIII fragment of pH/Q-I was subcloned into M13mp18 and M13mp19 to obtain both orientations. pH/Q-I DNA was digested with SmaI and HindIII, and the required fragment was isolated following electrophoresis through 1% low melting-point agarose (section 3.13.1). M13 RF DNA was similarly digested and recovered. Insert and vector DNAs were combined and ligated with T4 DNA ligase for 16 h at 19 ^{o}C, using estimated concentrations of 3.5 nM and 2 nM respectively (section 3.13.4).

The ligation reactions were used to transform E. coli strain JM101, which was plated on a lawn of JM101 on H. agar with IPTG/XGal selection (section 3.14.3). Plaques were grown and insert sizes and orientations were determined from RF DNA. Subclones from pH/Q-I were named mSH18 and mSH19 for M13mp18 and mp19 respectively.

8.5 DNA sequence analysis and the oligonucleotide target sequence

Single-stranded DNA of clones mSH18 and mSH19 was prepared and sequenced, using Sequenase^{TM} (section 3.18). The sequence derived from mSH18/19 is 392 bases long and is given in appendix 4C: the complete sequence was derived from each strand. It has a base composition of 56% G+C, similar to values previously determined (58% and 67%; sections 6.6 and

7.5). The oligonucleotide target sequence was sought, using the computer programs FIND and GAP (section 3.19.3). The best direct match (without gap insertion) matches at 16 of 24 positions but can not be that responsible for hybridisation since it spans the XhoI site within pH/Q-I, whilst the target sequence is to one side of this site (as demonstrated by hybridisation to XhoI-digested pH/Q-I DNA: cleavage within the oligonucleotide target would abolish hybridisation).

Additional sequencing reactions for each of clones mSH18 and mSH19 were performed in which oligonucleotide 11 was the primer. 50 ng oligonucleotide was used in place of the normal sequencing primer, with reaction conditions otherwise unchanged. Both clones gave identical sequence, indicating that priming had occurred within the vector: as expected, the sequence matched that of M13. The position at which priming was initiated (M13 nucleotide position 2609) was determined from the derived sequence, and thus the match of the oligonucleotide to the vector was found:-

```
                C
5'....GGCT AAGTCGGTGA CGGTGATAA....3'
     :::: : ::::: :: ::::: :
3'-CCGA TGCAGCCSCT GCCACASTC-5'
                  S
```

The best matching oligonucleotide of the degenerate mixture is shown: S represents G or C. The 3'-end of the oligonucleotide is bound, as expected to be necessary for priming to occur. In order to draw a reasonable match for this experimentally determined hybrid, two looped-out bases are necessary, one in each strand. This observation provides

evidence that it is possible for oligonucleotides to form hybrids containing looped-out unpaired bases, as well as with mismatched pairs, albeit under the low stringency conditions of the sequencing reaction compared to the hybridisation washing protocols. A hybrid containing a looped-out base was postulated for oligonucleotide 2 in section 6.7. The best match of oligonucleotide 11 to the mSH sequence, with the insertion of such gaps allowed, has a single gap of one base and matches at 15 positions, with the ends unpaired:-

```
          264                      287
           .                        .
5'....AAAATCGT GGCGACCGCTGTTTC....3'
         ::: ::::::::::::
   3'-CCSATGCA CCGCTGGCGACASTC-5'
             S
```

As above, the best matching oligonucleotide is shown. The 3'-end is not bound, accounting for the lack of sequence priming within the insert. Another possibility for the match is as shown but with the 5'-terminal six bases (3'-ACACTC-5') matching the target DNA exactly after a gap of 8 bases from the previous oligonucleotide base: the destabilising effect of such a large gap is unknown but would be expected to be significant. The T_m of the hybrid shown calculated from the Wallace rule (section 3.11.2) is 52 ^{o}C, with no account taken of the destabilising effect of the looped-out base: this compares with the experimental value of 53 ^{o}C. Oligonucleotide 10 hybridised less stably and differs from number 11 at the last of the bound positions indicated above (the A-T base pair), which would be expected to depress the T_m by 2 ^{o}C. The lack of a contiguous match indicates that the oligo-

nucleotide did not select a sequence of the type sought. There was no match to the conserved features of the EGF-family within the sequence data obtained.

The sequence of clone mSH was used in a homology search of entries within the Genbank nucleic acid sequence database containing poxviral sequence data. No significant matches were found. The sequence does not contain an open reading frame of significant length and so was not subjected to further analysis.

8.6 Summary of chapter

Three pairs of oligonucleotides (24- and 22-mers) were designed to identify DNA sequences encoding the conserved domains of the EGF-family of growth factors. Use was made of the information obtained from the partial sequence of the MCV homologue of the vaccinia virus p37 gene obtained in the previous chapter. One of these oligonucleotides identified a locus of MCV I DNA which was cloned into pBR322 and subcloned into M13, from which 392 bases of sequence was obtained by dideoxynucleotide sequencing. The oligonucleotide target sequence was identified but is not contiguous and so does not represent a sequence of the type sought.

CHAPTER 9

Summary and discussion of results

9.1 The molecular epidemiology of subtypes of the Molluscum contagiosum virus

Prior to and shortly after the commencement of this project were two reports of the analysis of restriction endonuclease cleavage patterns (as opposed to physical maps) of Molluscum contagiosum virus (MCV) DNA, using virus purified from clinical lesions. The first was by Parr et al. (1977) who presented three cleavage patterns of MCV DNA and referred to viral subtypes I, II and III. In a study group of 10 isolates from 8 patients there were six of subtype I (four patients), one of subtype II and three of subtype III. Subtypes II and III showed similar DNA cleavage patterns and the single instance of II is probably a variant of III. Darai et al. (1986) similarly studied MCV DNA from 14 independent patients and observed two viral subtypes. Their terminology differs from that of Parr et al.: subtypes I and II respectively correspond to the earlier subtypes II/III and I. The later study reported a relative incidence of patients with MCV I:II of 13:1, which compares with 4:4 for the earlier study.

Chapter 4 decribes the analysis of 124 MCV DNA isolates from 105 independent patients. Three viral subtypes were

found, based on DNA cleavage patterns obtained following digestion with the restriction enzymes BamHI, ClaI and HindIII: the independence of these subtypes was confirmed following the determination of physical maps of viral DNA, yet their close relationship was shown by DNA cross-hybridisation (see below). Two of these subtypes are similar to those reported by Darai et al. (1986) and the terminology used is consistent with theirs. The third subtype (III) differs from the others on the basis of its cleavage pattern and restriction maps and has not been reported previously. Total lesion DNA without prior purification of virus was used in the majority of cases for identifying the MCV subtype: the relative incidence of MCV I:II:III was 92:31:1 isolates (80:25:1 patients). Thus, the novel subtype III is also rare. There were no phenotypic differences between lesions induced by the different subtypes.

During the course of this project the initial study by Darai et al. (1986) was expanded to 40 patients, for which the MCV I:II ratio was 39:1 (Reisner et al., 1986). Clearly these relative incidences of MCV subtypes differ from those reported by Parr et al. (1977: ratio of 4:4; see above) and from those for the study group used in this project. One possible reason for this is that there may be a different geographical distribution of each subtype. The group of patients studied by Parr et al. were from the United States and those studied by Reisner et al. included 33 from West Germany and 6 from Hong Kong: those studied in this project were from South-East England and had a ratio of 80:25 for MCV

I:II. In respect of a possible biassed geographical distribution of MCV subtypes, it may be relevant that the single instance of MCV III observed in this project and not otherwise reported was from a patient (a 16-year old girl) in whom lesions developed shortly after returning from a trip to Moscow.

Postlethwaite et al. (1967) have previously noted the rarity of Molluscum contagiosum lesions on the soles and palms and also that lesions occurred 3- to 4-fold more often in males than in females for patients aged 7-15 years in their study group. Similar observations were made here: there were no occurrences of lesions on soles and palms, and the ratio of boys to girls was 11:5. All other body sites were represented, with the genitalia being a common site of infection in adults. The relative incidences of MCV subtypes I and II in different patient groups were subjected to statistical analyses, the only significant finding being that MCV II did not occur in patients less than 15 years of age ($P=0.035$). The reason for this restriction is unknown but is not due to sexual transmission of this subtype in a manner analagous to herpes simplex virus II (Chaney et al., 1983). There was no difference in the relative incidence of MCV subtypes in lesions at genital sites (ie. with presumed sexual transmission), contrary to the early suggestion of restriction of MCV II to this site (Darai et al., 1986).

Shortly after the completion of the work described in this thesis, the studies of Darai et al. (1986) and Resiner et

al. (1986) were extended by the inclusion of 108 patients from Scotland (Scholz et al., 1988 and 1989), giving a cumulative ratio of 142:5 for patients with MCV I:II. However, two of the four newly reported isolates of MCV II were described as "variants" of their prototype MCV II due to differences in the restriction endonuclease cleavage profiles (Scholz et al., 1989). Close examination of the BamHI digestion products of their so-called MCV II variant reveals it to be the same as the subtype III reported in this project - ie. they have unknowingly reported two new instances of MCV III as well as two new MCV II isolates. This mistake is easily appreciated due to the superficial similarity of BamHI digests of these two subtypes (see figure 4). Thus, the 108 Scottish patients gave a ratio of 104:2:2 cases of MCV I:II:III. The new data confirm the lack of association of MCV II with genital infection. There was one occurrence of each of subtypes II and III in patients less than 15 years of age, contrary to the lack of occurrence in this age group reported in this project. Scholz et al. (1988) additionally demonstrated that virus isolates obtained from two or more members of families with Molluscum contagiosum yielded identical restriction patterns of MCV DNA.

The cumulative experience of the occurrence of the three MCV subtypes reported to date is given in table 10.

	Country	MCV subtype I	II	III
(a)	Maryland, U.S.A.	4	4	0
(b)	West Germany	32	1	0
	Hong Kong	6	0	0
	Scotland	104	2	2
(c)	South-East England	80	25	1
	Totals	226	32	3

Table 10. Cumulative data relating to the occurrence of MCV subtypes. Figures represent the numbers of patients with subtypes I, II and III, grouped geographically, as discussed in the text. Based on data from (a) Parr et al. (1977), (b) Scholz et al. (1989) and (c) that given in chapter 4.

9.2 Structure and organisation of the Molluscum contagiosum virus genome

A viral isolate of each MCV subtype was purified from clinical lesions of Molluscum contagiosum from a single patient and the genomic DNA was studied by restriction digestion. The genome sizes of these isolates are 190, 200 and 196 kb for subtypes I, II and III respectively. The presence of a genomic terminal cross-link was demonstrated for MCV I and II by rapid renaturation of terminal DNA fragments. These observations are in agreement with those of Parr et al. (1977) who reported a genome size of 179 kb for an isolate corresponding to subtype II, and demonstrated terminal cross-links by electron microscopy of DNA spreads. The presence of terminal inverted repeats was demonstrated by mapping the genomic terminal restriction sites and by cross-hybridisation of terminal DNA fragments. The repeat sizes of the chosen MCV I, II and III DNA isolates are 3.9-5.4 kb, 7.2-10.6 kb and 3.7-7.0 kb respectively. These terminal repeats are variable in length for different isolates of the same subtype, indicative of the presence of variable numbers of short tandem repetitions. In each of these structural respects, MCV DNA is typical of poxviral DNA (see section 1.2.3). MCV DNA does not cross- hybridise with vaccinia virus DNA.

The G+C content of sequence data derived from cloned DNA from several regions of the MCV genome averages 58% and is consistent with estimates based on denaturation studies (Parr et al., 1977) and the high frequency of sites for the

restriction enzyme SmaI. This high G+C content is reflected in the high frequency of these bases at the third positions of codons for the MCV homologue of the vaccinia virus major envelope antigen (p37) gene and so gives rise to a different pattern of codon preference in these two viruses.

The restriction sites for the enzymes BamHI, ClaI and HindIII in the genomes of each of MCV I, II and III were mapped by hybridisation of isolated restriction fragments to genomic digests. The maps for each subtype are different and there are no regions of conservation as for orthopoxvirus DNA (Mackett and Archard, 1979). Neither is there any evidence from the mapping data that one subtype arose by recombination of the other two, or that any subtype is a mutant of the copy-round type described by Archard et al. (1984). DNA of each subtype cross-hybridises with that of the other two subtypes, revealing a close sequence relationship between them. Pairwise cross-hybridisations using isolated DNA fragments demonstrated the colinearity of the subtype genomes, with two exceptions: a region of MCV I does not have a counterpart (ie. it did not cross-hybridise) in the other subtypes, and two short insertions/deletions are required to maintain map alignment consistent with hybridisation studies, one of these occurring in the MCV I-unique region (see figure 13). It would be of interest to know whether these variable regions encode distinguishable features of the subtypes although, as noted above, there are no known phenotypic differences in the lesions they induce. These physical mapping data indicate that the three MCV subtypes are more

divergent than are the different species of the orthopoxvirus genus when genome structure is analysed and so could be considered as different species of a Molluscum contagiosum genus of poxviruses.

The high degree of DNA sequence homology between MCV subtypes, >90% as determined by cross-hybridisation under stringent washing conditions, yet their low level of restriction site conservation implies that the orthopoxviruses, for which there is extensive restriction site conservation, have even less sequence diversity. The degree of sequence homology between MCV I and II DNA is not uniform throughout the genome, as indicated by the comparison of restriction maps of an equivalent clone from each, for which a portion is identical whilst the remainder shows no conservation (see figure 18): sequence data derived from equivalent regions showing no restriction site conservation are 87% homologous. Orthopoxvirus homologies are also not uniform but are greater in the central region of their genomes.

When restriction digests of DNA from independent isolates of the same MCV subtype were compared, variations of two types were observed. One related to variations in the size of the terminal repeat: these affected the sizes of terminal DNA fragments and also of the near-terminal HindIII fragments for MCV II, due to the larger repeat which contains a HindIII site for this subtype. Variations were always identical at each end of the genome, as was demonstrated for

cowpox and monkeypox virus mutants (Mackett and Archard, 1979; Dumbell and Archard, 1980), reflecting the mechanism of viral DNA replication (see section 1.2.4). The other type of variation related to the occasional loss of individual restriction sites or the gain of new ones (ie. restriction fragment length polymorphisms), as is expected on the basis of some sequence heterogeneity between isolates. One such variation was unusual in that it occurred at high frequency: loss of the BamHI site at 76 kb from the left terminus occurred in 33% of MCV I DNA isolates. The genomic region spanning this variable BamHI site was cloned from an isolate with and an isolate without such a site: restriction analysis did not reveal other significant differences. The presence or absence of this restriction site did not correlate with other factors of lesion presentation and so the significance of this variation, if any, remains unknown. The same variation has also been noted by Scholz *et al.* (1989).

During the course of this project, sequence data were obtained that were homologous to part of the gene for the vaccinia virus 37 kD polypeptide (p37), the major antigen in the envelope of extracellular vaccinia virus (Hirt *et al.*, 1986). The envelope containing this antigen, which is highly abundant, is acquired by the virus from the Golgi system (Hiller and Weber, 1985; see section 1.2.2). The sequence data encode the C-terminal 82 amino-acids of the MCV II p37-homologue and represent the first *bona fide* MCV gene to have been identified, although other open reading frames for potential genes have also been described (see below). Both

the gene locus (30 and 38 kb from the left termini of MCV and vaccinia virus DNA respectively) and its direction of transcription (towards the left terminus) are the same for vaccinia virus and for MCV II. This implies a similar genomic organisaion for the two viruses despite their unrelatedness: this being the case, then the orientation of the MCV genomic restriction maps presented in chapter 5 is the same as that established for the orthopoxviruses (Mackett and Archard, 1979).

The degree of sequence homology between the vaccinia virus p37 gene and the portion of its MCV homologue for which data were obtained is 42%, with 63% similarity (49% identity) at the amino-acid level. The relatively low level of nucleotide homology reflects the differing base composition of vaccinia virus and MCV DNAs, as discussed above. The MCV gene encodes a protein that extends 13 amino-acids beyond the C-terminus of the vaccinia virus product. It is of interest to observe such a high degree of homology between the major envelope antigen of these two viruses, particularly as they have been reported not to cross-react immunologically (see section 1.2.6). However, such studies have been based on sera raised against vaccinia virus released by experimental lysis of infected cells: such "intracellular" virions do not possess the outer membrane containing p37 that is present around "extracellular" virions.

The complete MCV I p37 gene homologue has recently been cloned and sequenced using methodology described in chapters 3

and 6. The protein product is 389 amino-acids long, compared to 372 for vaccinia virus. The sequence data show that the high level of amino-acid sequence homology is not confined to the C-terminus but exists throughout the protein (Neil Blake, personal communication). There are some small regions of dissimilarity, including relative insertions, but the overall homology is 62% amino-acid similarity (40% identity). The gene is 66% GC-rich, with 82% of the codons having G or C in their third position (as observed for the MCV II homologue) and the 100 base pairs upstream are 72% GC-rich. The presence of open reading frames flanking the p37-homologue implies tight clustering of genes in this region of the genome.

During the sequencing studies described in chapter 6, a number of open reading frames which may represent MCV genes were identified. However, their putative products are not homologous to any previously sequenced proteins. Neither did their upstream or downstream sequences show homology with each other or with previously characterised poxviral regulatory motifs, with the exception of the vaccinia virus early gene transcription termination signal (TTTTTXT, Shuman et al., 1988). Their identity as MCV genes is still in question. Two such open reading frames occur on opposite DNA strands and only five bases exist between their stop codons, reminiscent of the tight packing of vaccinia virus genes (Pluciennicak et al., 1985). Sequences immediately 5' and 3' to the MCV p37 gene-homologue are similar to those flanking the p37 gene itself, suggestive of conserved control mechanisms. The start codon (underlined) occurs in the sequence TAAAATG for MCV and

TAAATG for vaccinia virus, the latter being highly conserved for vaccinia virus late genes (see section 1.2.4); the stop codon (underlined) occurs in the sequence TGAAAAAAAAAGGAAT for MCV and TAAAAAAAAGAAAATA for vaccinia virus (see section 7.6). The latter motif does not occur at the 3'-ends of other sequenced poxvirus genes and its similarity for MCV is all the more significant given that the MCV genome is GC-rich (whilst that of vaccinia virus is AT-rich). This sequence motif may be involved in signalling termination of transcription of this late gene.

9.3 Molluscum contagiosum virus-induced tumourigenesis: use of synthetic oligonucleotide probes to identify members of the epidermal growth factor gene family

A mechanism of tumourigenesis by MCV was proposed in section 1.4, based on the following information:-

(1) Although MCV replicates in differentiating keratinocytes of the higher layers of the epidermis, it is enhanced proliferation of basal cells, which do not harbour virus, that generates the tumour. This implicates secretion of a factor from infected cells that provides the stimulus to proliferate.

(2) Vaccinia virus, the prototypic orthopoxvirus, directs secretion of a growth-stimulating factor from infected cells when grown *in vitro*. This factor (VGF) is derived from an early viral gene product and is homologous to epidermal growth factor (EGF) in terms of its structure and ability to bind and activate the receptor for this growth factor.

(3) Within the epidermis, it is the cells of the basal layer that possess the greatest numbers of epidermal growth factor receptors. It is these cells that provide the proliferative potential of the epidermis.

The proposed scheme involves infection of keratinocytes by mechanical transmission of the virus followed by its replication in these cells. A growth factor analogous to that of vaccinia virus is encoded by MCV and is synthesised in the

infected cells, from which it is released (either by secretion or by cleavage from a membrane-bound precursor). This growth factor has little or no effect on the infected cells, which have lost proliferative potential due to differentiation, but enhances the proliferation of basal cells by augmenting the normal physiological signal to divide (see figure 3).

Crucial to this scheme is the possession by MCV of the coding potential for such a growth factor. It was not possible to directly assay such a factor in the same way as for vaccinia virus since there is no in vitro culture system for MCV. Thus, it was decided to attempt to clone the gene for the putative viral growth factor, as the first step to dissection of the proposed mechanism of tumourigenesis.

It was not possible to select a clone by using gene probes for VGF or EGF, since neither cross-hybridised to MCV DNA: neither did these probes cross-hybridise to each other. The reason for this is evident when the amino-acid and nucleotide sequences of members of the EGF-family, which includes transforming growth factor-alpha (TGF) in addition to EGF and VGF, are compared (see table 2). The overall degree of amino-acid homology is low (30-40%) and that at the nucleotide level is insignificant due to differences in codon choice. In fact, there is insignificant nucleotide sequence homology between the VGF gene and its homologue in another poxvirus, Shope rabbit fibroma virus, indicating the sequence diversity of genes in this family. However, there are short amino-acid sequence motifs that are highly conserved: notably

the third disulphide-loop (domain A), conserved in all members of the EGF-family, and a region within the first disulphide-loop (domain B), with homology confined to EGF and VGF but not TGF. As discussed in section 1.3.5, structure-function analyses have implicated the highly conserved residues of domain A in receptor binding. Any new member of the EGF growth factor family is expected to possess an equivalent domain. This suggested a cloning protocol for the putative MCV gene, with selection based on hybridisation to synthetic oligonucleotide probes for DNA sequences capable of encoding this domain.

An alternative to cloning the gene on the basis of its structural features would be to express randomly cloned DNA fragments and assay for growth stimulating or receptor binding activity. Such an approach has the advantage of not presupposing the nature of the growth factor but is technically complex for the following reasons. Expression would have to be driven from a heterologous promoter since those of poxviral genes do not conform to those of host nuclear genes but are recognised by a virus-specific RNA polymerase (see section 1.2.4). A system with an inducible eukaryotic promoter with assay for growth stimulating activity by transfection of 3T3 fibroblasts and selection by focus formation was considered. Immunological assay of expression products would not necessarily be successful since there is little cross-reactivity between EGF, TGF and VGF: although VGF does cross-react with antisera to mouse EGF it does not do so with antisera to human EGF or rat TGF (Stroobant et al., 1985). An

expression approach would require cloning, on a random basis, such that the coding region of the putative gene was correctly positioned with respect to a promoter. Given these complexities, the structural approach on the basis of oligonucleotide probes was thought to be more likely to yield an answer to the specific question of whether MCV encoded an EGF-like growth factor.

Oligonucleotides (typically 15-20 nucleotides) have been widely used in recent years to select genes for which partial amino-acid sequence data are available (Wallace and Miyada, 1987). The normal course of events is to select a region of amino-acid sequence for which codon degeneracies are minimal and synthesise oligonucleotides, often making use of codon preferences shown by the organism concerned. Longer oligonucleotides (30-100 nucleotides) have a greater likelihood of being specific to the gene of interest than do short oligonucleotides and also allow more stringent washing of hybridisation filters and minimise the effects of mismatches (Wood, 1987). The situation for this project is a step removed from this protocol in that the amino-acid sequence data used in oligonucleotide design do not relate to the gene of interest but rather to a family of related genes. This introduces a degree of uncertainty into the approach. Although the conservation of domain A of the EGF-family is impressive, it is difficult to work with since it is restricted both in length and in degree: strict conservation applies only to five positions of a septapeptide sequence whilst an eighth residue shows some degree of conservation.

Domain B shows strict conservation of a pentapeptide, with some conservation of flanking residues, but is absent from TGF species. Given this information, and the functional implication of the former domain (see above), most attention was given to oligonucleotides designed against domain A. Unlike the optimal situation discussed above, it was not possible to incorporate codon preference data into oligonucleotide design since none were available for MCV genes and those for vaccinia virus are not applicable since the base compositions of the two viral genomes are significantly different.

Eleven different oligonucleotide probes were used during the course of this project. Their design principles have been extensively discussed in chapters 6, 7 and 8 (see also table 4). Briefly, these were (i) the use of G-T "wobble" pairing to reduce the degeneracy of 17-mer probes (chapter 6), (ii) the use of highly degenerate (256-fold) 20-mer probes with the only other step being to assume codon usage for arginine (chapter 7), and (iii) the use of base positional preference data within codons of an identified MCV gene (the p37-homologue) in the construction of 22- and 24-mer probes.

A hybridisation worthy of follow-up was obtained in each of the three rounds of probe design. The first of these occurred near the left terminus and an equivalent locus was present in all three MCV subtypes. However, sequence data for this locus did not contain an open reading frame with homology

to the EGF-family. The second attempt yielded a hybridisation signal to MCV II DNA, but not to that of MCV I. The locus was cloned and sequenced but again did not yield the gene of interest: instead, it contained part of the MCV homologue of the gene for the vaccinia virus major envelope antigen (p37). The third attempt was the most promising in terms of the clarity and strength of the hybridisation signal, giving a hybridisation to MCV I DNA, but not to that of MCV II. The latter oligonucleotide design incorporated information relating to base position and codon preferences derived from the p37-homologue sequence. The locus was close to the junction of unique and repeat sequences at the left genomic terminus, consistent with the position of the VGF gene in wild-type orthopoxviruses (Archard et al., 1986) and the SFGF and MGF genes in Shope rabbit fibroma and myxoma viruses (Chang et al., 1987; Upton et al., 1987) but in contrast to the location of the VGF gene within the terminal repeat of vaccinia virus strain WR (Venkatesan et al., 1982). However, sequence data again denied the presence of an EGF-like gene.

In all three cases a match to the oligonucleotide probes within the derived sequence data has been proposed. Such matches are consistent with the experimental data and indicate that the oligonucleotide probes did not behave in the desired fashion: they formed stable hybrids with sequences not conforming to the target against which they were designed. A resynthesis of the oligonucleotide yielding the first hybridisation signal behaved identically to the initial synthesis, verifying correct probe construction. On two

occasions, hybrids involved "looped-out" bases in one strand - ie. the matched bases were not contiguous and so the hybrids were meaningless in terms of selecting for a given amino-acid sequence since the reading frame was not maintained - whilst in the third, the arrangement of matches and mismatches minimised the effect of the latter.

Although there are insufficient data available in the literature to estimate the destabilising effect of such "looping-out" in this type of experiment there is no satisfactory alternative explanation of the oligonucleotide behaviour: the data presented in chapters 6 and 8 imply that the destabilising effect is similar to that of a mismatched base pair. Moreover, formation of such a structure between one oligonucleotide and M13 DNA was demonstrated to occur since the former was capable of priming sequencing reactions with the latter as a template: the resulting sequence data pinpointed the position of priming at which formation of a "looped-out" hybrid was indicated. Such structural features are thought to be common in RNA with extensive secondary structure (eg. ribosomal RNA and tRNA). Solution nuclear magnetic resonance has been used to study the structures of oligonucleotide duplexes with extra looped-out bases in one strand: the unpaired bases are incorporated into an otherwise intact helix by stacking internally (Woodson and Crothers, 1988; Joshua-Tor *et al.*,1988; van den Hoogen *et al.*, 1988), suggesting that their destabilising effect is not great.

There are two possible explanations for the failure to

locate a homologue of the VGF gene in this way. Firstly, there is the possibility that such a gene does not exist. However, this seems unlikely since homologues have been sequenced from two species of the leporipoxviruses (SFV and myxoma virus) and cumulative experience of studies of poxviral molecular genetics suggests conservation of gene functions. Additionally, recent studies of VGF show that it is responsible for a proliferative response, normally masked by cytolysis, that is associated with vaccinia virus infection and that occurs through a similar mechanism to that proposed for MCV (Buller et al., 1988b; see section 1.4): it would be remarkable if MCV induced proliferation by a different mechanism although in principle a growth factor not related to EGF could be involved. A likely candidate for the latter would be the recently described keratinocyte growth factor which is related to the fibroblast growth factors (Rubin et al., 1989). The second, more probable explanation is that the oligonucleotides do not closely represent the sequence of the MCV homologue of VGF. This is entirely possible since the EGF-family domain A, conserved though it is, is interrupted by two non-conserved amino-acids and oligonucleotide design must assume something of the nature of the residues at these positions. In this respect, the oligonucleotide sequences chosen in this project represent only a subset of those that could encode such a conserved domain. It is unfortunate that, rather than leading to no hybridisation, some of these probes were able to bind fortuitously to other sequences and yield significant hybridisation signals.

9.4 Use of oligonucleotides to identify new members of gene families

With hindsight, the constraints of the above experimental system demanded too much of oligonucleotide probes. However, it was a reasonable approach to have taken. Indeed, such an approach has been used successfully to clone the gene for the VGF-homologue from SFV (SFGF; Chang *et al.*, 1987) which was subsequently used to obtain the homologue from the related myxoma virus (MGF; Upton *et al.*, 1987). It is notable that the domain A of these sequences differs by the substitution of asparagine for the first glycine residue, which is otherwise conserved in all members of the EGF-family for which sequence data were previously available (see table 2). Virus-free lysates of cultured cells infected with SFV and myxoma virus have also been directly assayed. A factor with EGF-receptor binding activity and immunological cross-reactivity with EGF has been described, although its relationship to SFGF/MGF has not been established (Strayer and Leibowitz, 1987). The finding of an EGF-like function in members of a second poxvirus genus (the leporipoxviruses) strengthens the idea that this is a universal poxviral function and that the failure to identify the MCV homologue was not due to its non-existence but due to the limitations of the approach, complicated by unfortunate "spurious" mismatches to the oligonucleotide probes used.

Chang *et al.* (1987) also used an oligonucleotide to the

VGF sequence of conserved domain A of the EGF-family. It was a 128-fold degenerate 21-mer that did not cover the glycine residue that turned out not to be conserved in SFV but did cover the semi-conserved residue following this domain (ie. it covered the amino-acid sequence YTGIRCQ). Their oligonucleotide design incorporated vaccinia virus codon preferences for the last three residues (R, arginine; C, cysteine; Q, glutamine), a reasonable assumption given the similar base compositions of vaccinia virus and SFV DNAs. The hybrid formed was subsequently found to be mismatched at three positions, one of which was central and the other two of which were adjacent at the fourth and fifth positions from the 3'-end of the oligonucleotide, all due to the use of amino-acids that differed from those in VGF at the two non-conserved positions of domain A. These authors appear to have been fortunate in the choice of their oligonucleotide sequence in terms of the region selected, in that the codons used in the target gene corresponded to those preferred by vaccinia virus (which is not true for MCV), and in that hybrids to non-contiguous sequences of the type experienced in this project did not form more stably than those to the "correct" incompletely-matched target.

Similar situations in which oligonucleotide probes have been successfully employed are in the cloning of the vaccinia virus ribonucleotide reductase large subunit gene (Tengelsen *et al.*, 1988) and for identifying new members of the *ras* gene family (Touchot *et al.*, 1987). In the former case, a fully degenerate (128-fold) 18-mer was designed against a hexa-

peptide of which five positions were conserved in the ribonucleotide reductase large subunits of E. coli, herpes simplex virus, Epstein-Barr virus and mouse: the isolated vaccinia virus gene matched the oligonucleotide probe at all 18 positions. In the second example, a hexapeptide is strictly conserved in all members of the ras gene superfamily and a seventh amino-acid shows partial conservation. The authors used three approaches in the construction of 20-mer oligonucleotide probes: the construction of a 512-fold degenerate probe encompassing all possible sequences encoding the conserved amino-acid sequence, the use of inosine at highly degenerate positions to reduce the total degeneracy to 8, and the use of weak G-T "wobble" base-pairing (with 8-fold degeneracy) in the same fashion as that used in the design of oligonucleotides in chapter 6. They were in the fortunate position of having a bank of clone ras-related genes to use as positive controls. The first approach gave hybridisation to all controls but additionally yielded relatively high non-specific signals: the second did not hybridise to all controls. The third approach was ultimately the one successfully used: two oligonucleotides, each incorporating G-T "wobble" base-pairing at six positions, were designed to hybridise to either the coding or the non-coding DNA strand and were thus partially complementary. Although each hybridised to only a subset of control sequences, in combination they hybridised to all. This combination probe was used to identify 14 cDNA clones, of which 10 were shown to belong to the ras-family (with at least 16 of 20 positions matching) whilst the remainder were not ras-related (and had at most 15

matches, although a high proportion of the variable positions involved the more stable G-C base pair). A similar combination probe applied to the oligonucleotides designed in chapter 6 may have been more successful in identifying an MCV gene belonging to the EGF-family, although the major disadvantage in this system is that the conserved amino-acids are not contiguous but are interrupted by non-conserved residues.

9.5 Scope for further work

It is still possible that the oligonucleotide approach could be successful for identifying an MCV-encoded growth factor. However, all rational choices of oligonucleotide probes have already been made and so alternative approaches should be considered. One possibility, discussed above (section 9.3), is based on random cloning in a eukaryotic expression vector and selection for focus formation following transfection and expression in a suitable recipient cell type (eg. 3T3 fibroblast).

A second alternative now exists, with the advent of VGF-minus (VGF^-) mutants of vaccinia virus (Buller *et al.*, 1988a) and a means of selecting wild-type virus by growth in A431 cells with the generation of cell foci instead of plaques (Buller *et al.*, 1988b; see section 1.4). MCV DNA could be randomly cloned into the VGF^- mutant - eg. by recombination into the non-essential thymidine kinase locus, applying methodology developed for generating recombinant vaccinia

virus (reviewed in Mackett and Smith, 1986) - and clones restoring the wild-type phenotype selected. This approach would depend on the recognition of the MCV promoter for the VGF-homologue by the vaccinia virus RNA polymerase (analagous to non-genetic reactivation; see section 1.2.2).

Once the gene responsible for MCV-induced tumourigenesis is cloned and characterised a number of further studies would be possible. Its sequence could be compared to those of the existing members of the EGF-family and would contribute to greater understanding of the essential structural features of polypeptides that bind to and activate the EGF receptor. The sequence data would also allow a comparison of potential upstream and downstream regulatory sequences with those of the vaccinia virus and SFV growth factor genes to gain further understanding of the molecular biology of MCV and its relationship to other poxviruses.

A functional approach to the study of an MCV-encoded growth factor would be to clone the coding sequences in an inducible mammalian expression vector which would be transfected into recipient cells (eg. 3T3 fibroblasts) in culture: the effects of expression in a cell type that possesses receptors for EGF could thus be investigated, allowing an experimental test of the proposed model of tumourigenesis. Also, the effect of the medium conditioned by transfected and induced cells on the proliferation of uninfected cells could be studied. A closer experimental approximation to the proposed model of MCV-induced tumour-

igenesis would be the expression of the growth factor in cells that lack the ability to respond themselves (ie. cells lacking EGF receptors and so mimicking the MCV-infected non-proliferative keratinocytes) and their co-culture with EGF-responsive cells (mimicking the uninfected proliferative basal cells). Such systems could be further examined by the generation of anti-sera to synthetic peptides based on the sequence of the MCV translate and their use to investigate expression in the experimentally transformed cells. These sera could also be used for immunocytochemical staining of thin sections of natural tumours.

REFERENCES

Anand, R. (1986) Trends Genet. 2, 278-283

Andrews, C., Pereira, H. G. & Wildy, P. (1978) In "Viruses of vertebrates", 4th edition, pp. 356-389. London: Balliere Tindall

Archard, L. C., Kinchington, D. & Rice, A. P. (1986) Poster presentation at the sixth Poxvirus/Iridovirus workshop, Cold Spring Harbour

Archard, L. C. & Mackett, M. (1979) J. Gen. Virol. 45, 51-63

Archard, L. C., Mackett, M., Barnes, D. E. & Dumbell, K. R. (1984) J. Gen. Virol. 65, 875-886

Atkinson, T. & Smith, M. (1984) In "Oligonucleotide synthesis: a practical approach", pp. 35-81. Edited by Gait, M. J. Oxford: IRL Press

Bagnall, B. G. & Wilson, G. R. (1974) Aust. J. Dermatol. 15, 115-120

Baroudy, B. M., Venkatesan, S. & Moss, B. (1982) Cell 28, 315-324

Barrandon, Y. & Green, H. (1987) Cell 50, 1131-1137

Bateman, T. (1814) In "A practical synopsis of cutaneous diseases", 3rd edition. London: Longman, Hurst, Rees, Orme and Brown

Baxby, D. (1984) In "Topley and Wilson's 'Principles of bacteriology, virology and immunity'" vol. 4, 7th edition, pp. 163-182. Edited by Brown, F. & Wilson, G. London: Edward Arnold

Becker, T. M., Blount, J. H., Douglas, J. & Judson, F. N. (1986) Sex. Transm. Dis. 13, 88-92

Berkowitz, E. M. & Pogo, B. G.-T. (1985) Virology 142, 437-440

Bertholet, C., Van Meir, E., Heggeler-Bordier, B. & Wittek, R. (1987) Cell 50, 153-162

Binns, M. M., Tomley, F. M., Campbell, J. & Boursnell, M. E. G. (1988) J. Gen. Virol. 69, 1275-1283

Bird, A. P. (1986) Nature 321, 209-213

Bishop, M. J., Ginsburg, M., Rawlings, C. J. & Wakeford, R. (1987) In "Nucleic acid and protein sequence analysis: a practical approach", pp. 83-146. Edited by Bishop, M. J. &

Rawlings, C. J. Oxford: IRL Press

Blomquist, M. C., Hunt, L. T. & Barker, W. C. (1984) Proc. Natl. Acad. Sci. USA 81, 7363-7367

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S. (1977) Gene 2, 95-113

Boyle, D. B., Coupar, B. E. H., Gibbs, A. J., Seigman, L. J. & Both, G. W. (1987) Virology 156, 355-365

Brachmann, R., Lindquist, P. B., Nagashima, M., Kohr, W., Lipari, T., Napier, M. & Derynck, R. (1989) Cell 56, 691-700

Brown, J.P., Twardzik, D. R., Marquardt, H. & Todaro, G. J. (1985) Nature 313, 491-492

Brown, S. T., Nalley, J. F. & Kraus, S. J. (1981) Sex. Transm. Dis. 8, 227-234

Broyles, S. S. & Moss, B. (1986) Proc. Natl. Acad. Sci. USA 83, 3141-3145

Buller, R. M., Chakrabarti, S., Cooper, J. A., Twardzik, D. R. & Moss, B. (1988a) J. Virol. 62, 866-874

Buller, R. M., Chakrabarti, S., Moss, B. & Fredrickson, T. (1988b) Virology 164, 182-192

Burgess, A. W., Lloyd, C. J., Smith, S., Stanley, E., Walker, F., Fabri, L., Simpson, R. & Nice, E. (1988) Biochemistry 27, 4977-4985

Cabirac, G. F., Strayer, D. S., Sell, S. & Leibowitz, J. L. (1985) Virology 143, 663-670

Campbell, I. D., Cooke, R. M., Baron, M., Harvey, T. S. & Tappin, M. J. (1989) Progress in Growth Factor Research 1, 13-22

Carle, G. F., Frank, M. & Olson, M. V. (1986) Science 232, 65-68

Carpenter, G. (1987) Ann. Rev. Biochem. 56, 881-914

Carpenter, G. & Cohen, S. (1979) Ann. Rev. Biochem. 48, 193-216

Chaney, S. M. J., Warren, K. G., Kettyls, J., Zbitnue, A. & Subak-Sharpe, J. H. (1983) J. Gen. Virol. 64, 357-371

Chang, A. & Metz, D. H. (1976) J. Gen. Virol. 32, 275-282

Chang, W., Upton, C., Hu, S., Purchio, A. F. & McFadden, G. (1987) Mol. Cell. Biol. 7, 535-540

Chen, W. S., Lazar, C. S., Poenie, M., Tsien, R. Y., Gill, G.

N. & Rosenfeld, M. G. (1987) Nature 328, 820-823

Cobbold, R. J. C. & MacDonald, A. (1970) The Practitioner 204, 416-419

Coffey, R. J., Derynck, R., Wilcox, J. N., Bringman, T. S., Goustin, A. S., Moses, H. L. & Pittelkow, M. R. (1987) Nature 328, 817-820

Cohen, S. (1962) J. Biol. Chem. 237, 1555-1562

Cohen, S. (1987) In Vitro Cell. Dev. Biol. 23, 239-246

Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H. & Sheard, B. (1987) Nature 327, 339-341

Cooley, A. J., Reinhard, M. J., Gross, T. L., Fadok, V. A. & Levy, M. (1987) J. Comp. Pathol. 97, 29-34

Cotton, D. W. K., Cooper, C. Barrett, D. F. & Leppard, B. J. (1987) Brit. J. Dermatol. 116, 871-876

Dahlback, B., Lundwall, A. & Stenflo, J. (1986) Proc. Natl. Acad. Sci. USA 83, 4199-4203

Darai, G., Reisner, H., Scholz, J., Schnitzler, P. & Lorbacher de Ruiz, H. (1986) J. Med. Virol. 18, 29-39

Dawson, R. M. C., Elliott, D. C., Elliott, W. H. & Jones, K. M. (1986) "Data for biochemical research", 3rd edition. Oxford: Clarendon Press

DeLange, A. M., Reddy, M., Scraba, D., Upton, C. & McFadden, G. (1986) J. Virol. 59, 249-259

De Larco, J. E. & Todaro, G. J. (1978) Proc. Natl. Acad. Sci. USA 75, 4001-4005

Derynck, R. (1986) J. Cell. Biochem. 32, 293-304

Derynck, R. (1988) Cell 54, 593-595

Derynck, R. Goeddel, D. V., Ullrich, A., Gutterman, J. U., Williams, R. D., Bringman, A. S. & Berger, W. H. (1987) Cancer Res. 47, 707-712

Devereux, J., Haeberli, P. & Smithies, O. (1984) Nucleic Acids Res. 12, 387-395

Douglas, J. D., Tanner, K. N., Prine, J. R., Van Riper, D. C. & Derwelis, S. K. (1967) J. Am. Vet. Med. Assoc. 151, 901-904

Dumbell, K. R. & Archard, L. C. (1980) Nature 286, 29-32

Earl, P. L., Jones, E. V. & Moss, B. (1986) Proc. Natl. Acad. Sci. USA 83, 3659-3663

Engler, D. A., Matsunami, R. K., Campion, S. R., Stevens, A. & Niyogi, S. K. (1988a) Poster presentation at the UCLA symposium "Growth factors and their receptors: genetic control and rational application", Keystone; J. Cell. Biochem. suppl. 12A, 81

Engler, D. A., Matsunami, R. K., Campion, S. R., Stringer, C. D., Stevens, A. & Niyogi, S. (1988b) J. Biol. Chem. 263, 12384-12390

Eppstein, C. A., Marsh, Y. V., Schreiber, A. B., Newman, S. R., Todaro, G. J. & Nestor, J. J. (1985) Nature 318, 663-665

Epstein W. L., Conant, M. A. & Krasnobrod, H. (1966) J. Invest. Dermatol. 46, 91-103

Epstein, W. L. & Fukayama, K. (1973) J. Invest. Dermatol. 60, 73-79

Esposito, J. J. & Knight, J. C. (1985) Virology 143, 230-251

Feinberg, A. P. & Vogelstein, B. (1984) Anal. Biochem. 137, 266-267

Fenner, F. (1985) In "Virology", pp. 661-684. Edited by Fields, B. N. New York: Raven Press

Francis, R. D. & Bradford, H. B. (1976) J. Virol. 19, 382-388

Gassman, U., Wyler, R. & Wittek, R. (1985) Arch. Virol. 83, 17-31

Gershon, P. D. & Black, D. N. (1988) Virology 164, 341-349

Geshelin, P. & Berns, K.I. (1974) J. Mol. Biol. 88, 785-796

Gillard, S., Sphener, D., Drillien, R. & Kirn, A. (1986) Proc. Natl. Acad. Sci. USA 83, 5573-5577

Gissman, L., Boshart, M., Durst, M., Ikenberg, H., Wagner, D. & zur Hausen, H. (1983) J. Invest. Dermatol. 83, 26s-28s

Gray, A., Dull, T.J. & Ullrich, A. (1983) Nature 303, 722-725

Greenwald, I. (1985) Cell 43, 583-590

Gregory, H. (1975) Nature 257, 325-327

Gullick, W. J., Marsden, J. J., Whittle, N., Ward, B., Bobrow, L. & Waterfield, M. D. (1986) Cancer Res. 46, 285-292

Hanahan, D. (1983) J. Mol. Biol. 166, 557-580

Hanahan, D. (1985) In "DNA cloning: a practical approach" vol. 1, pp. 109-135. Edited by Glover, D. M. Oxford: IRL Press

Harkins, R. N. & Brandis, J. W. (1988) Poster presentation at the UCLA symposium "Growth factors and their receptors:

genetic control and rational application", Keystone; J. Cell. Biochem. suppl. 12A, 81

Heath, W. F. & Merrifield, R. B. (1986) Proc. Natl. Acad. Sci. USA 83, 6367-6371

Henderson, W. (1841) Edinburgh Med. Surg. J. 56, 213-218

Henikoff, S. (1984) Gene 28, 351-359

Hiller, G. & Weber, K. (1985) J. Virol. 55, 651-659

Hirt, P., Hiller, G. & Wittek, R. (1986) J. Virol. 58, 757-764

Hojrup, P., Jensen, M. S. & Peterson, T. E. (1985) FEBS Letters 184, 333-338

Hunter, T. (1984) Nature 311, 414-416

Hursh, D. A., Andrews, M. E. & Raff, R. A. (1987) Science 237, 1487-1490

Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Apella, E. & Sussman, J. L. (1988) Nature 334, 82-84

Juliusberg, M. (1905) Deutsch Med. Wschr. 31, 1598-1599

Katzman, M., Carey, J. T., Elmets, C. A., Jacobs, G. H. & Lederman, M. M. (1987) Brit. J. Dermatol. 116, 131-138

Kilpatrick, D. R. & Rouhandeh, H. (1985) Virology 143, 399-406

King, C. S., Cooper, J. A., Moss, B. & Twardzik, D. R. (1986) Mol. Cell. Biol. 6, 332-336

Komoriya, A., Hortsch, M., Meyers, C., Smith, M., Kanety, H. & Schlessinger, J. (1984) Proc. Natl. Acad. Sci. USA 81, 1351-1355

Kozak, M. (1984) Nucleic Acids Res. 12, 857-872

Krishnamurthy, J. (1974) Arch. Dermatol. 109, 578

Laskey, R. A. (1984) "Radioisotope detection by fluorography and intensifying screens", review 23 (Amersham International)

Lax, I. Kris, R., Sasson, I., Ullrich, A., Hayman, M. J., Beug, H. & Schlessinger, J. (1985) EMBO. J. 4, 3179-3182

Lazar, E., Watanabe, S., Dalton, S. & Sporn, M. B. (1988) Mol. Cell. Biol. 8, 1247-1252

Lin, Y-Z., Caparaso, G., Chang, P-Y., Ke, X-H. & Tam, J. P. (1988) Biochemistry 27, 5640-5645

Mackett, M. & Archard, L. C. (1979) J. Gen. Virol. 45, 683-701

Mackett, M. & Smith, G. L. (1986) J. Gen. Virol. 67, 2067-2082

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) "Molecular cloning: a laboratory manual". New York: Cold Spring Harbour Laboratory

Mars, M. & Beaud, G. (1987) J. Mol. Biol. 198, 619-631

Marsh, Y. V. & Eppstein, D. A. (1987) J. Cell. Biochem. 34, 239-245

McAuslan, B. R. (1963) Virology 21, 383-389

McFadden, G., Pace, W. E., Purres, J. & Dales, S. (1979) Virology 94, 297-313

McMullen, B. A. & Fujikawa, K. (1985) J. Biol. Chem. 260, 5328-5341

Messing, J. (1983) Meth. Enzymology 101, 20-78

Mitchell, J. C. (1953) Brit. J. Exp. Path. 34, 44-49

Montagna, W. & Parakkal, P. F. (1974) "Structure and function of skin", 3rd edition. New York: Academic Press

Montelione, G. T., Wuthrich, K., Nice, E. C., Burgess, A. W. & Scheraga, H. A. (1986) Proc. Natl. Acad. Sci. USA 83, 8594-8598

Montelione, G. T., Sporn, M. B. & Wagner, G. (1988) Poster presentation at the UCLA symposium "Growth factors and their receptors: genetic control and rational application", Keystone; J. Cell. Biochem. suppl. 12A, 136

Moss, B. (1985) In "Virology", pp. 661-684. Edited by Fields, B. N. New York: Raven Press

Nanney, L. B., Magid, M. Stoscheck. C. M. & King, L. E. (1984) J. Invest. Derm. 83, 385-393

Nestor, J. J., Newman, S. R., Delustro, B., Todaro, G. J. & Schreiber, A. B. (1985) Biochem. Biophys. Res. Comm. 129, 226-232

Niles, E. G., Condit, R. C., Caro, P., Davidson, K., Matusick, L. & Seto, J. (1986) Virology 153, 96-112

Oda, H., Ohayama, Y., Sameshima, T. & Hirakawa, K. (1982) J. Med. Virol. 9, 19-25

Overfield, T. M. & Brody, J. A. (1966) J. Paediatrics 69, 640-642

Ozanne, B., Richards, C. S., Hendler, F., Burns, D. & Gusterson, B. (1986) J. Path. 149, 9-14

Parr, R. P., Burnett, J. W. & Garon, C. F. (1977) Virology 81, 247-256

Paterson, R. (1841) Edinburgh Med. Surg. J. 56, 279-288

Pennington, T. H. (1974) J. Gen. Virol. 25, 433-444

Pfeffer, S. & Ullrich, A. (1985) Nature 313, 184

Pirie, G. D., Bishop, P. M., Burke, D. C. & Postlethwaite, R. (1971) J. Gen. Virol. 13, 311-320

Pluciennicczak, A., Schroeder, E., Zettlmeissl, G. & Streeck, R. E. (1985) Nucleic Acid Res. 13, 985-998

Porter, C. D., Porter, A. G. F. & Archard, L. C. (1988) CABIOS 4, 271-278

Postlethwaite, R. (1970) Arch. Environ. Health 21, 432-452

Postlethwaite, R., Watt, J. A., Hawley, T. G., Simpson, I. & Adam, H. (1967) J. Hygeine 65, 281-291

Purchio, A. F., Twardzik, D. R., Bruce, A. G., Wizental, L., Madisen, L., Ranchalis, J. E., Hu, S-L. & Todaro, G. (1987) Gene 60, 175-182

Rall, L. V., Scott, J. & Bell, G. I. (1985) Nature 313, 228-231

Reisner, A. H. (1985) Nature 313, 801-803

Reisner, H., Scholz, J. & Darai, G. (1986) Poster presentation at the sixth Poxvirus/Iridovirus workshop, Cold Spring Harbour

Robinson, A. J., Barns, G., Fraser, K., Carpenter, E. & Mercer, A. A. (1987) Virology 157, 13-23

Rohrmann, G., Yuen, L. & Moss, B. (1986) Cell 46, 1029-1035

Rosenthal, A., Lindquist, P. B., Bringman, T. S., Goeddel, D.V. & Derynck, R. (1986) Cell 46, 301-309

Rubin, J.S., Osada, H., Finch, P.W., Taylor, W.G., Rudikoff, S. & Aaronson, A.A. (1989) Proc. Natl. Acad. Sci. USA 86, 802-806

Sam, C. K. & Dumbell, K. R. (1981) Ann. Virol. (Inst. Pasteur) 132 E, 135-150

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467

Sasaki, M., Kato, S., Kohno, K., Martin, G. & Yamada, Y. (1987) Proc. Natl. Acad. Sci. USA 84, 935-939

Savage, C. R., Inagami, T., & Cohen, S. (1972) J. Biol. Chem. 247, 7612-7621

Savage, C. R., Hash, J. H. & Cohen, S. (1973) J. Biol. Chem.

248, 7669-7672

Schaudies, R.P., Gorman, R. M., Savage, C. R. & Poretz, R. D. (1987) Biochem. Biophys. Res. Comm. 143, 710-715

Schlessinger, J., Schreiber, A. B., Levi, A., Lax, I., Libermann, T, & Yarden, Y. (1983) CRC Crit. Rev. Biochem. 14, 93-111

Scholz, J., Rosen-Wolff, A., Bugert, J., Reisner, H., White, M. I., Darai, G. & Postlethwaite, R. (1988) J. Infect. Dis. 158, 898-900

Scholz, J., Rosen-Wolff, A., Bugert, J., Reisner, H., White, M. I., Darai, G. & Postlethwaite, R. (1989) J. Med. Virol. 27, 87-90

Schumperli, D., Lagadec, R. & Muller, H. K. (1977) J. Gen. Virol. 38, 161-166

Schwer, B. & Stunnenberg, H. G. (1988) EMBO J. 7, 1183-1190

Schwer, B., Visca, P., Vos, J. C. & Stunnenberg, H. G. (1987) Cell 50, 163-169

Shirodaria, P. V. & Matthews, R. S. (1977) Brit. J. Dermatol. 96, 29-34

Shope, R. E. (1932) J. Exp. Med. 56, 803-822

Shuman, S., Broyles, S. S. & Moss, B. (1987) J. Biol. Chem. 262, 12372-12380

Shuman, S. & Moss, B. (1988) J. Biol. Chem. 263, 6220-6225

Simpson, R. J., Smith, J. A., Moritz. R. L., O'Hare, M. J., Rudland, P. S., Morrison, J. R., Lloyd, C. J., Grego, B., Burgess, A. Q. & Nice, E. C. (1985) Eur. J. Biochem. 153, 629-637

Southern, E. M. (1975) J. Mol. Biol. 98, 503-517

Sporn, M. B. & Todaro, G. J. (1980) N. Eng. J. Med. 303, 878-880

Sproat, B. S. & Gait, M. J. (1984) In "Oligonucleotide synthesis: a practical approach", pp. 83-116. Edited by Gait, M. J. Oxford: IRL Press

Sproul, E. E., Metzgar, R. S. & Grace, J. T. (1963) Cancer Res. 23, 671-675

Staden, R. (1982) Nucleic Acids Res. 10, 4731-4751

Staden, R. (1984) Nucleic Acids Res. 12, 521-538

Staden, R. (1987) In "Nucleic acid and protein sequence analysis: a practical approach", pp. 173-217. Edited by

Bishop, M. J. & Rawlings, C. J. Oxford: IRL Press

Stoscheck, C. M. & King, L. E. (1986) J. Cell. Biochem. 31, 135-152

Strayer, D. S. & Leibowitz, J. L. (1987) Am. J. Path. 128, 203-209

Stroobant, P., Rice, A. P., Gullick, W. J., Cheng, D. J., Kerr I. M. & Waterfield, M. D. (1985) Cell 42, 383-393

Stuart, R. J., Muller, H. K. & Francis, J. D. (1971) Med. J. Aust. 2, 751-754

Sudhof, T. C., Russell, D. W., Goldstein, J. L., Brown, M. S., Sanchez-Pescador, R. & Bell, G. I. (1985) Science 228, 893-895

Sutton, J. S. & Burnett, J. W. (1969) J. Ultrastruct. Res. 26, 177-196

Tabor, S. & Richardson, C. C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771

Teixido, J., Gilmore, R., Lee, D. C. & Massague, J. (1987) Nature 326, 883-885

Tengelsen, L. A., Slabaugh, M. B., Bibler, J. K. & Hruby, D. E. (1988) Virology 164, 121-131

Thein, S. L. & Wallace, R. B. (1986) In "Genetic analysis of human diseases: a practical approach", pp. 33-50. Edited by Davies, K. E. Oxford: IRL Press

Tomley, F., Binns, M., Campbell, J. & Boursnell, M. (1988) J. Gen. Virol. 69, 1025-1040

Touchot, N., Chardin, P. & Tavitian, A. (1987) Proc. Natl. Acad. Sci. USA 84, 8210-8214

Twardzik, D. R., Brown, J. P., Ranchalis, J. E. & Todaro, G. J. (1985) Proc. Natl. Acad. Sci. USA 82, 5300-5304

Upton, C., DeLange, A. M. & McFadden, G. (1987) Virology 160, 20-30

Upton, C., Macen, J. L., Maranchuk, R. A., DeLange, A. M. & McFadden, G. (1988) Virology 166, 229-239

Upton, C., Macen, J. L.& McFadden, G. (1987) J. Virol. 61, 1271-1275

Upton, C. & McFadden, G. (1986) J. Virol. 60, 920-927

van den Hoogen, Y. Th., van Beuzekom, A. A., van den Elst, H., der Marel, G. A., van Boom, J. H. & Altona, C. (1988) Nucleic Acids Res. 16, 2971-2986

Van Meir, E. & Wittek, R. (1988) Poster presentation at the

seventh Poxvirus/Iridovirus workshop, Heidelberg

Venkatesan, S., Gershowitz, A. & Moss, B. (1982) J. Virol. 44, 637-646

Wallace, R. B. & Miyada, C. G. (1987) In "Guide to molecular cloning techniques", Meth. Enzymology 152; 432-442. Edited by Berger, S. L. & Kimmel, A. R. London: Academic Press

Weinrich, S. L. & Hruby, D. E. (1986) Nucleic Acid Res. 14, 3003-3016

Wharton, K. A., Kohansen, K. M., Xu, T. & Artavanis-Tsakonas, S. (1985) Cell 43, 567-581

Winkler, J. (1986) J. Biol. Chem. 261, 13838-13843

Wittek, R., Barbosa, E., Cooper, J. A., Garon, C. F., Chan, H. & Moss, B. (1980a) Nature 285, 21-25

Wittek, R., Cooper, J. A., Barbosa, E. & Moss, B. (1980b) Cell 21, 487-493

Wittek, R. & Moss, B. (1980) Cell 21, 277-284

Wittek, R., Muller, H. K., Menna, A. & Wyler, R. (1978) FEBS Letters 90, 414-446

Wong, S. T., Winchell, L. F., McCune, B. K., Earp, H. S., Teixido, J., Massague, J., Herman, B & Lee, D. C. (1989) Cell 56, 495-506

Wood, W. I. (1987) In "Guide to molecular cloning techniques", Meth. Enzymology 152, 443-447. Edited by Berger, S. L. & Kimmel, A. R. London: Academic Press

Woodruff, A. M. & Goodpasture, E. W. (1931) Am. J. Path. 7, 209-222

Woodson, S. A. & Crothers, D. M. (1988) Biochemistry 27, 3130-3141

Wright, C. F. & Moss, B. (1987) Proc. Natl. Acad. Sci. USA 84, 8883-8887

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) Gene 33, 103-119

Yarden, Y., Schreiber, A. B. & Schlessinger, J. (1982) J. Cell Biol. 92, 687-693

Yarden, Y. & Schlessinger, J. (1987) Biochemistry 26, 1434-1442

Yuen, L. & Moss, B. (1986) J. Virol. 60, 320-323

Yuen, L. & Moss, B. (1987) Proc. Natl. Acad. Sci. USA 84, 6417-6421

PUBLICATIONS ARISING FROM THIS PROJECT

Porter, C. D. & Archard, L. A. (1987) "Characterisation and physical mapping of Molluscum contagiosum virus DNA and location of a sequence capable of encoding a conserved domain of epidermal growth factor" Journal of General Virology 68, 673-682

Porter, C. D., Blake, N. W. & Archard, L. C. (1988) "Structure and activity of epidermal growth factor-like peptides: induction of basal cell proliferation by a poxvirus gene product?", "Skin" colloquium of the 625th Biochemical Society Meeting, London. Biochemical Society Transactions 16, 671-674

Porter, C. D., Blake, N. W., Archard, L. C., Muhlemann, M. F., Rosedale, N., & Cream, J. J. (1989) "Molluscum contagiosum virus types in genital and non-genital lesions" British Journal of Dermatology 120, 37-41

Porter, C. D., Muhlemann, M. F., Cream, J. J. & Archard, L. C. (1987) "Molluscum contagiosum: characterisation of viral DNA and clinical features" Epidemiology and Infection 99, 563-567

Porter, C. D., Porter, A. G. F. & Archard, L. C. (1988) "BBC microcomputer controlled field inversion gel electrophoresis" Computer Applications in the Biosciences 4, 271-273

APPENDIX 1

Materials and reagents

Chemical reagents

With the exception of the following, all chemical reagents and solvents were obtained from British Drug Houses. Sodium chloride, tris(hydroxymethyl)aminomethane (Trizma base) and tris(hydroxymethyl)aminomethane hydrochloride (Trizma. HCl), ethidium bromide, β-mercaptoethanol, Sigmacote, spermidine, calcium chloride, N,N,N',N'-tetramethylethylene diamine (TEMED), polyvinylpyrrolidone, thiamine hydrochloride, polyethylene glycol and manganese chloride were obtained from Sigma. Zinc sulphate was obtained from May and Baker. Ethanol was obtained from James Burrough Ltd. Dithiothreitol (DTT) was obtained from Bethesda Research Laboratories. Dimethylsulphoxide (DMSO) was obtained from Fluka. Hexamino cobalt (III) chloride was obtained from Aldrich. Caesium chloride, dextran T10, Ficoll 70 and bovine serum albumin (BSA) were obtained from Pharmacia. Xylene cyanol was obtained from National Diagnostics. Reagents for the automated synthesis of oligodeoxyribonucleotides on a Pharmacia Gene Assembler were obtained from Pharmacia or from Cruachem.

Radioisotopes

Aqueous solutions of [α^{35}S]dATPαS (500 Ci/mmol, 10 uCi/ul), [α^{32}P]dCTP (3,000 Ci/mmol, 10 uCi/ul) and [γ^{32}P]ATP (3,000 Ci/mmol, 10 uCi/ul) were obtained from New England Nuclear, Du Pont.

Resins, chromatography and filtration media, agarose and blotting membranes

Amberlite ion exchange resin was obtained from Bio-Rad. Sephadex G-25 and G-50, and pre-packed G-25 columns (NAP-10) for exclusion chromatography were obtained from Pharmacia. NENSORB-PREP columns for affinity purification of full length oligonucleotides were obtained from Du Pont-NEN. 3MM paper, GF/C filters and 0.22 or 0.45 um filters for solution sterilisation were obtained from Whatman. Agarose and low melting-point agarose for gel electrophoresis were obtained from Bethesda Research Laboratories. Nitrocellulose membranes (0.45 um, BA85) for Southern blotting were obtained from Schleicher and Schuell. Nylon membranes (Hybond-N) for Southern blotting were obtained from Amersham International.

Nucleic acids and DNA sequencing kits

Salmon sperm DNA and *E. coli* tRNA were obtained from Sigma. Hexanucleotides ($p(dN)_6$) for random primer extension, bacteriophage lambda DNA, plasmid pBR322 DNA, M13mp18 and mp19 RF vector DNA, synthetic EcoRI linkers, ATP and deoxyribonucleotides were obtained from Pharmacia. M13mp8 and mp9 RF vector DNA was obtained from Amersham International. An unfrozen stock of bacteriophage lambda DNA for oligomerisation was obtained from New England Biolabs. 7-deaza-dGTP was obtained from Boehringer Mannheim. M13 sequencing kits were obtained from Amersham International and from the United States Biochemical Corporation (Sequenase™). The latter kit included dITP.

Recombinant DNA

A clone of the human EGF cDNA in the plasmid pUC9 (phEGF) was obtained from J. Scott, MRC Clinical Research Centre, London. A clone of the VGF gene in pBR328 (pVGF) was obtained from A. Rice, Imperial Cancer Research Laboratories, London.

Photography and autoradiography

Agarose gels stained with ethidium bromide and visualised by transillumination were photographed using FP4 film (5" x 4") obtained from Ilford Photographics. X-ray film for autoradiography was either X-Omat XAR-5 (Kodak), Rx-100 (Fuji) or Hyperfilm-MP (Amersham International). FP4 was developed with Autophen or Ilfotec-HC, X-ray film with Phenisol, and both were fixed with Hypam, all obtained from Ilford Photographics.

Microbiological culture media

Agar and bacto tryptone were obtained from Difco. Bacto yeast extract was obtained from Difco or from Beta Lab. Ampicillin and chloramphenicol antibiotics were obtained from Sigma. Isopropyl β-D-thiogalactoside (IPTG) was obtained from Pharmacia. 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (XGal) was obtained from Bethesda Research Laboratories or Northumbria Biologicals Ltd.

Enzymes

Lysozyme was obtained from Sigma. Proteinase K and calf intestinal phosphatase were obtained from Boehringer Mannheim. Ribonuclease A and S1 nuclease were obtained from Pharmacia. Exonuclease III and T4 DNA ligase were obtained

from Bethesda Research Laboratories. T4 polynucleotide kinase was obtained from Northumbria Biologicals Ltd. or from Pharmacia. E. coli DNA polymerase I "Klenow" fragment was obtained from Pharmacia or Bethesda Research Laboratories. Restriction endonucleases were obtained from a variety of suppliers, including Northumbria Biologicals Ltd., Pharmacia, Boehringer Mannheim and Bethesda Research Laboratories.

APPENDIX 2

Patient details of MCV isolates, grouped according to subtype

DNA isolated from lesions of *Molluscum contagiosum* was digested with BamHI and the viral subtype determined on the basis of the restriction pattern produced. Isolates 1, 26 and 8 are the standards to which all others have been compared and represent MCV subtypes I, II and III respectively. Of 124 independant viral isolates from 105 patients there were 92 instances of MCV I, 31 of MCV II and only one of MCV III. Details of patients and isolates are listed below, grouped according to the viral subtype found. Each isolate is numbered (column 1) and the patient's sex and age are given (columns 2 and 3) as is the site at which the lesion occurred (column 4). Lesions were obtained from patients presenting to hospital departments or to Genito-Urinary clinics (section 3.3): those presenting to the latter (all female) are indicated by a * following the record of lesion site. Fifteen patients presented on more than one occasion or had lesions removed from different sites on the same occasion: lesions so derived from the same patient are cross-referenced in column 5. The final column ('Other features') contains information relating to other relevant medical conditions of the patient, peculiarities of the clinical presentation and notes relating to features of the viral DNA (see sections 4.2 and 5.9).

Abbreviations: SLE, systemic lupus erythematosis. At., atopic. Giant, giant Molluscum contagiosum. Notes: 1, viral DNA (MCV I) lacks the BamHI site between fragments D and F as mapped for isolate 1.

No.	Sex	Age	Lesion site	Ref.	Other features
A.	Isolates of MCV I				
1	M	22	Antecubit fossa		At., ichthyosis
2	F	62	Face		
4	M	13	Trunk		Atopic dermatitis
5	M	12	Axilla/shoulder		
6	F	24	Trunk/thigh		
7	F	21	Trunk		
9	M	11	Shoulder	10	1
10	M	11	Axilla	9	1
11	M	12	Groin/abdomen		
12	F	19	Genitalia *	13,91	1
14	F	8	Thigh		
16	M	8	Antecubit		
17	M	10	Axilla		
18	M	27	Arm	19	
19	M	27	Chest	18	
20	F	30	Face/trunk/shoulder/leg		
22	M	3	Forehead		1
23	F	26	Thigh		1
24	F	19	Genitalia *		
25	F	24	Genitalia *		
27	F	3	Forearm		Atopic dermatitis
28	F	50	Forehead		1
29	F	49	Neck		
30	F	28	Genitalia		
31	F	25	Outer canthus		
32	M	9	Axilla		Atopic dermatitis
34	M	22	Genitalia		
36	F	19	Genitalia *		
37	F	23	Genitalia *	38	
38	F	23	Genitalia *	37	
39	F	19	Genitalia *		
40	M	31	Groin		1
41	F	53	Forehead	45,66	
43	F	19	Genitalia *		1
45	F	53	Forehead	41,66	
48	M	24	Axilla	49,50	SLE, on steroids
49	M	24	Axilla	48,50	as above
50	M	24	Axilla	48,49	as above, Giant
52	F	23	Genitalia *		1
53	F	18	Genitalia *	62	
54	F	22	Genitalia *		1

No.	Sex	Age	Lesion site	Ref.	Other features
55	F	26	Genitalia *		
56	F	19	Genitalia *		
58	M	13	Elbow		Giant (solitary)
59	M	38	Thigh		1
60	F	23	Genitalia *		
62	F	18	Genitalia *	53	
63	F	23	Thigh (left)	64	Atopic dermatitis, 1
64	F	23	Thigh (right)	63	as above
65	F	16	Thigh		
66	F	53	Face/neck	41,45	
67	F	22	Genitalia		
68	F	24	Genitalia *		1
69	F	40	Genitalia *		1
72	F	22	Genitalia *		1
73	F	20	Neck		1
74	F	20	Genitalia *		
75	M	18	Chest		Atopic dermatitis
76	F	8	Antecubit fossa		
79	F	24	Genitalia *	119	
80	M	21	Genitalia		Intrafollicular, 1
81	F	28	Genitalia *		
82	F	25	Genitalia *		1
83	M	14	Neck	115	
84	M	17	Popliteal fossa		Atopic dermatitis
85	F	2	Eyebrows		1
87	M	24	Beard area		Intrafollicular
88	M	11	Neck		
90	M	11	Genitalia		1
91	F	21	Genitalia *	12,13	1
92	F	21	Genitalia *		1
94	F	19	Genitalia *		
95	F	24	Thigh		1
96	F	19	Genitalia *		1
99	M	19	Genitalia		
100	F	22	Genitalia *		1
102	F	18	Face		
103	F	50	Thigh		1
104	F	51	Genitalia *		
106	M	31	Genitalia		
107	F	23	Genitalia *		1
112	M	27	Flank		
113	F	35	Genitalia *		
114	F	6	Waist		
115	M	15	Neck	83	
116	F	23	Neck/chin		
117	F	26	Genitalia *		1
118	F	41	Face		Erythema, SLE ?
119	F	25	Genitalia *	79	
121	F	22	Genitalia *		1
123	F	19	Genitalia *		
124	F	22	Genitalia		

No.	Sex	Age	Lesion site	Ref.	Other features
B.	Isolates of MCV II				
3	F	53	Forehead		At., pancytopenia
13	F	20	Genitalia *	12,91	
15	F	19	Face		
21	F	24	Axilla		
26	F	23	Back/chest/neck		
33	F	24	Chin		
35	F	37	Genitalia *		
42	M	63	Genitalia	57	
44	F	29	Genitalia *		
46	F	22	Genitalia *		
47	F	23	Genitalia *		
51	F	23	Genitalia *		
57	M	63	Back	42	Giant (solitary)
61	F	20	Genitalia *		
70	F	15	Cheek		
71	F	25	Genitalia *		
77	F	34	Genitalia *		
78	F	23	Genitalia *	101	
86	F	28	Thigh	122	
89	M	24	Chest/axilla		
93	F	25	Genitalia *		
97	M	70	Head	98	
98	M	70	Head	97	
101	F	23	Genitalia *	78	
105	F	35	Genitalia *		
108	F	25	Genitalia *	109,110	
109	F	25	Genitalia *	108,110	
110	F	25	Genitalia *	108,109	
111	F	28	Genitalia *		
120	F	21	Genitalia *		
122	F	29	Thigh	86	
C.	Isolates of MCV III				
8	F	16	Back		Large lesions

APPENDIX 3

Hybridisation and cross-hybridisation data for restriction mapping the genomes of MCV subtypes I, II and III

The following tabulated data were obtained by hybridisation of isolated restriction fragments of MCV DNA to Southern blots of digests of the same DNA. Isolates 1, 26 and 8 were used throughout to study MCV subtypes I, II and III respectively. For each probe fragment, given in the left-hand column, those hybridising fragments are indicated: control blots verified the specificity of the probes and the identification of the genome terminal fragments. Those probes consisting of 2- or 3-molar fragments are indicated by (2) or (3) in the left-hand column. Refer to tables 7, 8 and 9 for sizes and molarities of hybridising fragments. In the case of MCV I digests with PstI, SmaI and XhoI, only hybridisations to Pst/A-W, Sma/A-W and Xho/A-O are documented: there were often other smaller fragments which hybridised but which were not identifiable, this being indicated by the '+' sign. For the experiments listed in (i) and (ii), fragments below Bam/M in the BamHI digest were not present on the blot. In a few cases probes consisted of pooled fragments which could not be efficiently separated due to their similar sizes, as indicated by eg. MCV III Cla/E+F. Likewise, hybridisations to one of two fragments of similar size that could not be unequivocally assigned to either one are indicated by eg. D/E in the MCV I ClaI column. n.d. = not done.

A. Hybridisation data for MCV subtype I

Probe	BamHI	ClaI	HindIII	PstI	SmaI	XhoI
(i) HindIII fragment probes						
Hind/						
A	ADFGI	AD/EFG	A	BDE/FHL MOSX +	CEGHIKO STVW +	ADEFGH LO +
B	A	A	B	CNU +	CDMOQ RTU +	AMN +
C	BLM	A	C	IP +	FJL +	+
D	DJ	A	D	KQV +	EGO	I/JO +
E	CK	B	E	AT +	A	DKN +
F	A	C	F	A +	B +	M +
G	C	C	G	A	AV +	G +
H	C	D/E	H	+	S +	+
I	H	B	I	E/F	A	D +
J (2)	BEH	ABD/E	JK	CGW	AENP +	CBFM +
K (2)	BE	AD/E	JK	CGJR	ENP +	CBM +
L	J	A	L	K +	JV +	O +
M	C	D/EH	M	+	P	+
N	C	BH	N	+	P	F
O (2)	BE	AD/E	O	CG	N	BC
(ii) ClaI fragment probes						
Cla/						
B	CHK	B	EIJN	AE/FTW +	A	DFKN +
C	AC	C	FG	A +	ABV +	GM +
F	A	F	A	H +	HT +	O +
G	F	G	A	M	HV +	LO +
(iii) BamHI fragment probes						
Bam/						
A	A	ACF	ABF			
B	BE	AE	CJKOQ			
C(2)+D(2)	CD	ABCDEH	ADEGHMN			
E	BE	AE	JKO			
F	F	DG	A			
G	G	A	A			
H	H	B	IJ			
I	I	A	A			
J	J	A	DL			
K	K	B	E			
L	L	A	C			
M	M	A	C			
N (2)	N	A	AC			
O	O	n.d.	E			

B. Hybridisation data for MCV subtype II

Probe	BamHI	ClaI	HindIII
(i) HindIII fragment probes			
Hind/			
A	BE	AEF	A
B (2)	BCEFKLM	A	B
C	EGH	AC	C
D	DIJK	A	D
E	C	A	E
F (2)	A	BFG	F
G	A	C	G
H	AD	AD	HI/J
I (3)	AD	AD	HI/J
J	A	B	I/J
K	A	DF	K
L	C	A	L
M		B	M
(ii) ClaI fragment probes			
Cla/			
B		B	FI/JM
C	AG	C	CG
D	AD	AD	HI/JK
E	B	E	A
F (2)		F	AFK
G		G	F
(iii) BamHI fragment probes			
Bam/			
C	C	A	BEL
D	AD	AD	DHI/K
E (2)	E	AF	ABC
F	F	A	B
G	G	C	C
H	H	A	C
I	I	A	D
J	J	A	D
K (2)	K	A	BD

C. Hybridisation data for MCV subtype III

Probe	BamHI	ClaI	HindIII
(i) ClaI fragment probes			
Cla/			
A	CDEFI	AE/F	ABEGIKMOP
B	B	B	CFHN
C	A	C	DFIJL
D	BGH	D	CG
E+F	ADEG	AE/F	DKMOQR
G	AE	G	A
H	A	H	AF
(ii) BamHI fragment probes			
Bam/			
D	DE	AE/F	GIKMOP
E (2)	DE	AE/FGJ	AKMOQR
F	F	A	B
G	G	DE/F	DG
H	H	D	G
I	I	A	B

D. Data obtained from cross-hybridisation of DNA from MCV subtypes I, II and III.

The following data were obtained in the same way as those in A-C above but by cross-hybridising fragments from one MCV subtype to digests of DNA from a different subtype. Controls for probe specificity were performed as above (data not included here). The isolates used were the same as those used above with the one exception that the MCV I BamHI fragments used as probes against MCV II DNA were derived from a different isolate (number 41). A dash indicates that there were no cross-hybridising fragments for a given probe.

Probe	BamHI	ClaI	HindIII
(i) Hybridisation of MCV I BamHI fragment probes to MCV II DNA			
Bam/			
A	BEGH	ACE	ABC
B	AD	AD	DHI/J
E	AD	ADF	HI/JK
F	B	EF	A
G	FM	A	B
H	A	B	F
I	C	A	B
J	CJ	A	EL
K	-	-	-
L	I	A	D
O	-	n.d.	-
(ii) Hybridisation of MCV II HindIII fragment probes to MCV I DNA			
Hind/			
A	ADF	AD/EFG	
B (2)	ADG/HI	A	
C	AC	AC	
D	BLJ	A	
E	DJ	A	
F (2)	CG/H	BHD/E	
G	C	C	
H	BE	AD/E	
I (3)	BE	AD/E	
J	C	B	
K	E	D/E	
L	J	A	
(iii) Hybridisation of MCV I BamHI fragment probes to MCV III DNA			
Bam/			
A	AGH	CDE/FH	DFGIJL
B	DE	AE/F	GIKMOP
C(2)+D(2)	ABCE	ABDGJ	ACEFG
E	DE	AE/F	KMOQR
F	A	GI	AF
G	F	A	B
H	B	B	HN
I	C	A	B
J	C	A	E
K	-	-	-
L	C	A	G
M	C	A	G
N (2)	CI	A	BG

(iv) Hybridisation of MCV III ClaI fragment probes to MCV I DNA

Cla/			
A	BDEGIJLMN	AE	ACDJKLOQ
B	CH	BEH	EHIJMN
C	A	A	AB
D	AC	C	FG
E+F	ABE	AE	BFJKO
G	DF	D	A
H	AF	F	A

(v) Hybridisation of MCV III ClaI fragment probes to MCV II DNA

Cla/			
A	ACDEF	AD	BDEI/J
B	A	BGF	FI/J
C	BE	A	AB
D	AG	C	CG
G	BE	F	A
H	B	E	A

APPENDIX 4

Nucleotide sequence data

A. Sequence of a 2.5 kb MCV II SmaI fragment

Clones mIIa and b in M13mp8 were obtained by subcloning the 2.5 kb SmaI fragment from pMBH-II, the overlap of MCV II BamHI fragment D and HindIII fragment D (chapter 6). The complete sequence of 2,571 bases is given here.

```
   1  CCCGGGCCGC CACAACAGCA GTGCGGAATC ACGCTGTTGT GCGGCGGCCC
  51  ACTACCACAA CGCAGATGCT GCGCCAGAAT ATGTACATCG CGGTCGAAAA
 101  CAAGCTACAC CGAATACTGA AGTACAAGCG TCACAGGAAA GACATCACTG
 151  TCACCTACTG AAGCCTCAAA GGAAACAAGA CCAAATGCCC GAGGGGACGC
 201  GGGCCTACGT GTACAGTTAC GCATCAGTGT GTTGTCGCTA TGCATGTGGT
 251  TTCATTGTCG AACTGCTTGG CTGACGAGAC GGAAACGAAA ACACAAAACC
 301  TCACTCACGC ACTTCACACG CACAGTGCAC ATCGCGCGGG GCACACACGT
 351  CGCGCCGCGC CGCCACGGCA GCGTCCGAGG GCGGGCGCTG CGTGGGCGCG
 401  CGGCACCCCC AGATGAACGG AGGGCGAGGC CGCGGCCCGC GCACCATCAC
 451  AAGATAGCGG TTATATTTTC AGTTTGGAAA AGACCCGCCG CGGCGGAACA
 501  GCAGCATCAA CAAACAAAAA AAACATGGTG TTTCGACTCG ACTTCGATGC
 551  GTATACAAAA AAACGGAAGG AACTGTGGCA CGTAGCCGCA GCCCGCAGCT
 601  GTGCGGCGGC AAGCCAGCGC ACAGCTGCGG ACACGCCTCG CGCCGAGTAC
 651  ACGCACCTGT GTCGCGAGCT CCCCTTGCGC ACGCGCCAGT GTCGTGGACA
 701  CCGGACTTGT CACGCCACTG CCGCGTGCAG GCGCCACAGA TGCGTGACTC
 751  CGGCACGAAG CCATGCCCGC GAGTTGTCCT CGCGTTCGAA TTGCGGATGG
```

```
 801  CTGCATCTTT GTCCATCCTT GTCGCAAAGG CGATTTTCTT GTTGTTTTTT
 851  CTTGTTGTGT TTTTTTCCTG TCATGGAGCA AAAGGTGTGG GGGAAAGATG
 901  GGCCTTCGCT GCCTTCGAAG AGAATGACTT CTCCAGAAAA GGGTGGTAGG
 951  GGGGCTATGT ACATGTCGTG TTCACATACT GCATTGAAAT ACTGGACCTT
1001  TGTGGGTCGA CACCCGGCGC CCGCCACCAT GCTTGTTTGC GAAAGCGGGC
1051  GGTTGCGAAT CACATGCGCA AGCGTGTTAT CTGTGTGTTT ATACCACGAA
1101  GCATAGAGAC GTCATATTCG CGACAAGGGA AATCCACTGG TCTGTTCGTT
1151  CACGAACGCC GCCTCATCAA AGAGACCAGG AGGTATCGTC ACGTTCAAGA
1201  CATTACTGTC ACTGGTCTGG TTGAGCCGGA CTTGTAATTC CGGCGGCATT
1251  GTGGTGTTCG CCGGCCAGTA TACTGGCGAA TAATCCGTCA GATCTTCCGC
1301  AGTCTGATTG TACACGCTCA AGGTGCACTT TTCTCCGCCG CCGTCGTCGC
1351  CGCCGTCGTC GCCGCCGCCG TCATAGTAAT AGCCTCCATC TTCTTGGCCG
1401  TCATCTCCGT CACTGCCGTC GCCTCCTCCG TAATAGAAGA ATTTAAACCA
1451  CGGAAAACGC TCGGAATCAT TTCCATGTTG GCACAGTCCT TTCAGCCCAA
1501  GAACGTTCGG GGGGCTGCCG CTCGAGGACA TTACCTCACC GATATACTTT
1551  CCCTGATGGC ATGTGCCGTT CTGATACGTA AAGTAGCCGC TGACTGTCAC
1601  GTCGGGGCCT GAAGTGCAAT TCTGCGGAGT CATGAGCGGG AACGTCTGGT
1651  TCCATAAATC CCAGATCTCT CCGCACGGTG GGAGTGGTGT CATTGCTCGT
1701  GCAATTCCAA AGGACAAGCG TCAAGTTTTC GTAAGGGAAG CCGTCGCCGT
1751  CGAGCGGCGA TCCTAGACCC AGCGAATGAG TCGCATTCGT ATTGAGCTCG
1801  ACAAAGACGT CACCTGCGCA CGAAACCATT ACCACGCACA ACGCCATCAG
1851  TGCCCCCGCG CACGCAACCG CGAGAGCGGC CGCGGCCCGA GGCCGACGGC
1901  GCGCGCTACA TGTTTGCATC GTCGTCTACC ATTGCAGTAA GCCTCTTTAT
1951  TTTCATTTTC GCGAAGAATC CGCTCCGGAG TCGAGACTCG TGTCAGAGCC
2001  GCGGCATTCT TCTCAGTTTT CGTCCCACAA GGCCGAGGGC TTTGCGTCGA
2051  CCAAGAAAAA ACAACAAGAT ACTGGCGCGC AACGCAACGA AATGGCACGG
2101  ACGGAGCGCA TCTGAAATGT GGACCTAGGC AGCGCGGAAA GGAGCAGGCG
2151  CGGGTCAGAG GTTGTTTAGC ATCATGAGCA GAGCGAGGAG AAGGCTGTGC
2201  TGCGCTCGGA CCTTGACCAA GCTCCGAGCA TCTCTCCTAG CACGCGAGCA
```

```
2251  CGCCACACCA TGCCTGACGC CAATTCCGCC CGCGCCGGAG TAGCGTAAAA
2301  ATGGTACACG TCGGGTCCGA CCCCATCTTG CAAGCACGTC GCAACTGGCG
2351  CAAGCTGCGC GTAGCTTTGG CAGTGGCCTC TTAATTCTCA ATTTTGCGGA
2401  GAAGAACGAA GTGGCGCCGA CGCGCGCTCG TGCATGCCTG CGCCGGCAAA
2451  TGCCGTGCGC AGAGCGGACA AGAGCTGCGC GGGCTGCGCA CGCAGCGCGT
2501  CGGGCACTCG CACGCGAAGG GCCACGCGCA GCTCGCCCGC CATGCCGCTG
2551  TAGTGCAGGC CGTAGCCCGG G
```

B. Sequence of a 0.45 kb MCV II SmaI/XhoI fragment

Clones mSX and mSXrev in M13mp18 and mp19 respectively were obtained by subcloning the 0.45 kb SmaI/XhoI fragment of pH/M-IIB, equivalent to the MCV II HindIII fragment L (chapter 7). The complete sequence of 423 bases is given here.

```
   1  TCGAGTCCGC GGTTGCTTTG CGCACCGCCG AGGACATGCC CGCGGCGCGC
  51  TTCTCCAGCG GCGAAATCTG CGCGGCCGTG CACTGGGGGG GCTCCAGAGC
 101  GCGCAGCGGG CGCGCGGGCC GGAAAACGGC TCGCGCTGTT CATGGCGGGC
 151  TTACACGGGC ATTCCTTTTT TTTTCAGTTC ACGGAGAGCA GCGTGGCGGG
 201  CTCCGGCGGC AGTGCCCGAG GAAGCGGCGA GCTGAAGGAC GACTGCCAGT
 251  CGCGCTCGAA GAGCGCACCG AGCGCTCGCG CGAACGTGCG TTCCGCGCAG
 301  TTGAAGCTGA CGAAGGCGTG CCGCGCGTAG TGCGTGCCGT CTATGTCGGC
 351  GTTGCTCACA TGCACGTACT CGTCGTCAAC CACGAGGAGC TTGGTGTTAT
 401  TGATGGCGTC GCCCTTGGCT CCC
```

C. Sequence of a 0.4 kb MCV I SmaI/HindIII fragment

Clones mSH18 and mSH19 in M13mp18 and mp19 respectively were obtained by subcloning the 0.4 kb SmaI/HindIII fragment of clone pH/Q-I, the MCV I HindIII fragment Q (chapter 8). The complete sequence of 392 bases is given here.

```
  1  AAGCTTAGCC ATCATCGAAG TCTTGCATGG TAGACGGCTA CTCCCATATC
 51  GTGATAGCTA GGACAGGTAT ATTGGTGTTG TAGTAGTGGT GCGCGGAAAT
101  GGTCGCAGTG ATGCCGTCGT GCTTGTCCCC GAAGTAGATG CGCCCCGCAG
151  AGGTTAAGCC CGACACCTTG CACGTAAGTA CCACGACGCA CGTGTCTTCC
201  GTCGCGGAAA AGACACTCGC AGAGACCGTT ACTCGAGGTC GCTTGACCAA
251  GACATCGAGC CGCAAAATCG TGGCGACCGC TGTTTCGTTG TGAGGCTTGA
301  CAACGAAGAC GTAGTACCGA CGTCGTCGGT TGTGACTCTG GTCAGAACGA
351  CGCAGCCCAA CTCGCAGAGT TCATTCTGTG CGTCAGCCCG GG
```